Characterising kinetoplastid pantothenate kinase, a multifunctional enzyme of the CoA biosynthesis pathway

Rebecca Barbara Roscoe

May 2023

Thesis submitted to the University of London in fulfilment of the requirements for the degree of Doctor of Philosophy of the University of London

Supervisors: Dr Martin C. Taylor and Professor John M. Kelly

Department of Infection Biology, London School of Hygiene and Tropical Medicine, University of London, United Kingdom

Funded by the Medical Research Council, UK

Declaration

I, Rebecca Barbara Roscoe, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Rebecca Barbara Roscoe

November, 2022

Abstract

Kinetoplastid parasites of the *Trypanosoma and Leishmania* genera cause fatal diseases of the tropics, subtropics and Mediterranean basin, in the form of Human African Trypanosomiasis, Chagas disease and Leishmaniasis. In addition to debilitating effects on the health of thousands of individuals, these illnesses have a crippling impact on the global economy, with an estimated annual cost of over 7 billion US dollars. Despite their clinical and economic importance, efforts to develop safe and effective treatments have made limited progress since discovery of these pathogens over 100 years ago.

Understanding parasite biochemistry is critical to the development of effective therapies. Coenzyme A (CoA) is a cofactor required for hundreds of metabolic reactions and its biosynthesis pathway may be a target for drugs against apicomplexa, fungi and bacteria. However, in kinetoplastids this pathway has been poorly characterised.

The first step in CoA synthesis requires pantothenate kinase (PanK). In this work, PanK is shown to be structurally unique in kinetoplastids compared to other eukaryotes as it contains two additional fused domains, putatively identified by homology as conferring phosphodiesterase and adenylating activity. By a combination of genetic and biochemical approaches, we investigated the role of PanK in growth and survival of the pathogens *Trypanosoma cruzi* and *T. brucei*.

Tagging of *T. cruzi* PanK by CRISPR spCas9 revealed expression of an extra-nuclear multi-domain polypeptide (~170kDa). Gene deletion using CRISPR was not achievable in *T. cruzi* epimastigotes, suggesting that at least one activity of the protein is essential to this life cycle stage. Similarly, PanK depletion in bloodstream-form *T. brucei* by RNAi was lethal, a phenotype that was partially rescued by CoA supplementation.

Add-back experiments revealed that constitutive expression of the *T. cruzi* orthologue rescues RNAi lethality in *T. brucei*, demonstrating functional homology of the two enzymes. Furthermore, using the same method we found that the additional fused domains are critical to function, as the isolated *T. cruzi* PanK domain does not rescue lethality of *T. brucei* PanK depletion. This is in spite of the fact that this domain is active, as demonstrated by kinase assays. We also found that within the PanK domain, Arg-1270 is critical to function and important for the *in vitro* proliferation of *T. brucei.* The putative phosphodiesterase appears

not to play an essential role in proliferation of this species, at least when the adenylating and PanK domains are active.

We conclude that CoA synthesis is essential for the *in vitro* growth of clinically important trypanosomes. The finding of a kinetoplastid-specific multi-functional PanK raises the possibility of therapeutic exploitability of this step of the biosynthesis pathway in kinetoplastids.

Acknowledgements

I would like to express my gratitude for all the knowledge and guidance I have been fortunate to receive from Dr Martin Taylor, who has been an exceptionally supportive and inspiring mentor and made my PhD a most stimulating and enjoyable four years. I also wish to thank Prof John Kelly for his valuable intellectual contributions throughout the PhD, and who has been another inspiration in his success and enthusiasm for science.

I would also like to thank Dr Fernanda Costa for her generosity and patience in teaching me molecular methods and whose kindness contributed greatly to my enjoyment of the project. Thank you to Dr Francisco Olmo for his microscopy and cell biology skills, particularly in the localisation. Not forgetting the rest of the lab: Dr Archie Khan, Shiromani Jayawardhana, Richard Atherton, Harry Langston, Dr Michael Lewis, Dr Amanda Francisco and Dr Alex Ward, who are exceptional scientists and were a pleasure to work and drink with. Thank you all for making this period a time of growth and friendship that I will look back on fondly.

I also wish to thank my partner Pej for the emotional support he has provided me in times of need and for being a role model as a diligent, thoughtful and passionate scientist who never takes shortcuts.

Table of Contents

List of Figures

List of Abbreviations

ACC Acetyl-CoA carboxylase

ACP Acyl Carrier Protein

ADP Adenosine Diphosphate

AMP Adenosine Monophosphate

ATF Adipose Tissue Form

ATP Adenosine Triphosphate

BLAST Basic Local Alignment Search Tool

BZ Benznidazole

BSF Bloodstream Form

CAR Carboxylic Acid Reductase

CD Cumulative Density

CNS Central Nervous System

CL Cutaneous Leishmaniasis

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats

DMSO Dimethyl sulfoxide

DNA Deoxyribose nucleic acid

DPCK Dephospho-CoA kinase

Ds Double-stranded

DSB Double-stranded break

DTU Discrete Typing Unit

EEP Endo/exonuclease Phosphatase

ELO Elongase

FA Fatty Acid

LC-FACS Long-Chain Fatty Acyl-CoA Synthetase

FBS Fetal Bovine Serum

FMN Flavin Mononucleotide

gRNA Guide RNA

GPI Glycosylphosphatidylinositol

HAT Human African Trypanosomiasis

HEK Human Embryonic Kidney cells

HI Heat-inactivated

HRP Horse radish peroxidase

IL Interleukin

iNOS Inducible Nitric Oxide Synthase

IPC Inositol phosphorylceramide

Kb Kilobase

LDH Lactate dehydrogenase

qPCR Quantitative PCR

Mbp Mega base-pairs

ML Mucocutaneous Leishmaniasis

mRNA messenger RNA

MSA Multiple Sequence Alignment

NADH Reduced nicotinamide adenine dinucleotide

NAC Nickel Affinity Chromatography

NF Nifurtimox

NG NeonGreen

ORF Open Reading Frame

PAGE Poly acrylamide Gel Electrophoresis

PBST Phosphate Buffered Saline/ Tween

PCF Procyclic Form

PCR Polymerase chain reaction

PEP Phosphoenolpyruvate

Pde Phosphodiesterase

PkD PanK Domain

PPAT Phosphopantetheine Adenylyltransferase

PV Parasitophorous vacuole

qPCR Quantitative PCR

RNA Ribonucleic acid

RNAi RNA Interference

SDS Sodium dodecyl sulfate

siRNA Small-interfering RNA

SM Sphingomyelin

Smase Sphingomyelinase

spCas9 *Streptococcus pyogenes* CRISPR- associated protein 9

TCA Tricarboxylic Acid

TcNTR Trypanosoma cruzi nitroreductase

Tet Tetracycline

TGFβ Transforming growth factor beta

Th2 T helper 2

UTR Untranslated Region

WHO World Health Organisation

VSG Variable surface glycoprotein

WT Wild-type

1. Introduction

1.1 An overview of the Kinetoplastid diseases

The order Kinetoplastida forms part of the Euglenozoa phylum, which diverged early in eukaryotic evolution. It comprises two families of flagellated protozoa, the Bodonidae and the Trypanosomatidae. The bodonids are biflagellate and can exist as parasites or as free-living organisms, the latter usually inhabiting soil and aquatic environments. One such organism is *Bodo saltans,* a free-living aquatic bodonid that feeds on bacteria*.* The clinically important kinetoplastids are trypanosomatids, which are uniflagellate and exclusively obligate parasites. They infect an exceptionally wide variety of host species, including vertebrates, invertebrates and plants.

A defining morphological feature of these protozoa, after which they are named, is the kinetoplast. This is a unique mitochondrion containing a dense network of catenated DNA circles, which is easily recognisable by microscopy. The high copy number of kinetoplast DNA facilitates molecular diagnosis of infections caused by kinetoplastid pathogens [1]. Within this network, maxicircles, which are typically 20-40 kb, encode ribosomal RNAs, proteins required for mitochondrial metabolism and some extra-mitochondrial proteins. Editing of the pre-RNAs generated by maxicircles requires guide RNA (gRNA) products encoded by smaller minicircles (0.5-10 kb in size). A 5' anchor within each gRNA anneals to a complementary sequence of pre-mRNA and the rest of the gRNA forms an imperfect duplex with non-complementary regions of the pre-mRNA. Within this duplex, un-paired uridines are removed from the pre-mRNA and un-paired purines in the gRNA specify uridine insertion into pre-mRNA at opposing positions [2].

Many trypanosomatids have multiple life cycle stages, adapting them to survival in different host environments and propagation between hosts, which is mediated by insect vectors. Within a life cycle, the parasite may visibly change in morphology, including flagellar length and location of nuclear structures, as well as undergoing changes in mitochondrial activity and other metabolic alterations. These changes adapt parasites to replication, dissemination, immune evasion or invasion of host cells, all the while making use of available nutrition sources.

Within the trypanosomatid family are pathogenic species that cause important diseases of humans and livestock. Chagas disease (American Trypanosomiasis), Human African Trypanosomiasis (HAT) and Leishmaniasis are listed by the World Health Organisation

(WHO) as Neglected Tropical Diseases and are together responsible for tens of thousands of deaths each year [3–5]. The causative trypanosomatids are the American and African trypanosomes, *Trypanosoma cruzi* and *T. brucei* respectively, and several species of *Leishmania.* They diverged hundreds of millions of years ago [6], which is reflected in their adaptation to different insect vectors, host tissue environments and use of distinct survival mechanisms. Yet, in terms of their cellular biology, biochemistry and genetics they have much in common. *T. brucei* and *Leishmania major* have 68 and 75% of their genomes, respectively, in common with that of *T. cruzi* [7]. Kinetoplastid genomes are typically 20-40 Mbp in size but can range from 14-87 Mbp. They are arranged into chromosomes: *T. cruzi* has 41, while *T. brucei* has 11 Mbp-sized chromosomes and ~100 mini chromosomes and 36 chromosomes have been described for *L. major* (per haploid genome) [8–10]. Rather than individual protein-coding genes being regulated by their own promoter, they are instead arranged in large polycistronic clusters of up to 100 genes and individual mRNAs are formed by co-transcriptional modification [11]. At the level of cell biology, kinetoplastids have much in common and this is reflected by the fact that certain drugs are effective against multiple species: fexinidazole inhibits both *T. brucei* and *T. cruzi*. Similarly, pentamidine can be used to treat *T. brucei* and *Leishmania* species.

1.2 Chagas disease

1.2.1 Transmission of *Trypanosoma cruzi*

The agent of Chagas disease, *T. cruzi*, has been endemic to South and Central America for millions of years [12]. It is currently endemic to 21 Latin American countries [13], where around 6-7 million people are thought to be infected [14], with poorer, rural communities being most at-risk. It is routinely detected in over one hundred mammalian species across the Americas, which makes eradication an impossible task [15]. Transmission to humans occurs principally via the triatomine insect vector, colloquially known as the ‘kissing bug’. Around 150 species of *Triatominae* transmit *T. cruzi*: these belong to the genera *Triatoma, Panstrongylus* and *Rhodnius* [16]. Most transmission is performed by *Triatoma infestans* and other species which thrive in domestic environments [17], such as traditional rural homes, which are easily entered and provide ample hiding places. At night, the insects feed on the blood of human occupants and defecate on their victim, depositing parasite-contaminated faeces on the skin. Infective parasites are rubbed or scratched into the bite wound by the individual or enter mucous membranes such as the conjunctiva. Other important routes of transmission are blood transfusion and organ transplants, congenital transmission and consumption of insect-

contaminated food and drink. In the 1970s, the serological prevalence of parasites in blood donors in Latin America was around 7%, but by 2006 this had decreased to below 1.3% [18]. Congenital transmission is estimated to occur in 2-10% of chronically infected pregnant women which equates to more than 14, 000 cases per year [19,20]. Hundreds of cases of oral transmission have been described within the Brazilian Amazon [21,22], with several cases resulting from consumption of juices made from açaí, sugar cane and other plants harbouring insects or their faeces.

Migration of infected individuals to non-endemic regions has lead to thousands of cases of Chagas disease being reported in Europe, North America, Australasia and Japan (Fig.1.2.1), where it can be spread by non-vectorial transmission routes [23–26]. In the USA, Chagas disease is also spread by triatomines as an enzootic disease: 11 vector species have been recognised as enzootic transmitters in the southern half of the United States, where wild mammals such as rodents and raccoons and domestic dogs act as reservoirs [25].

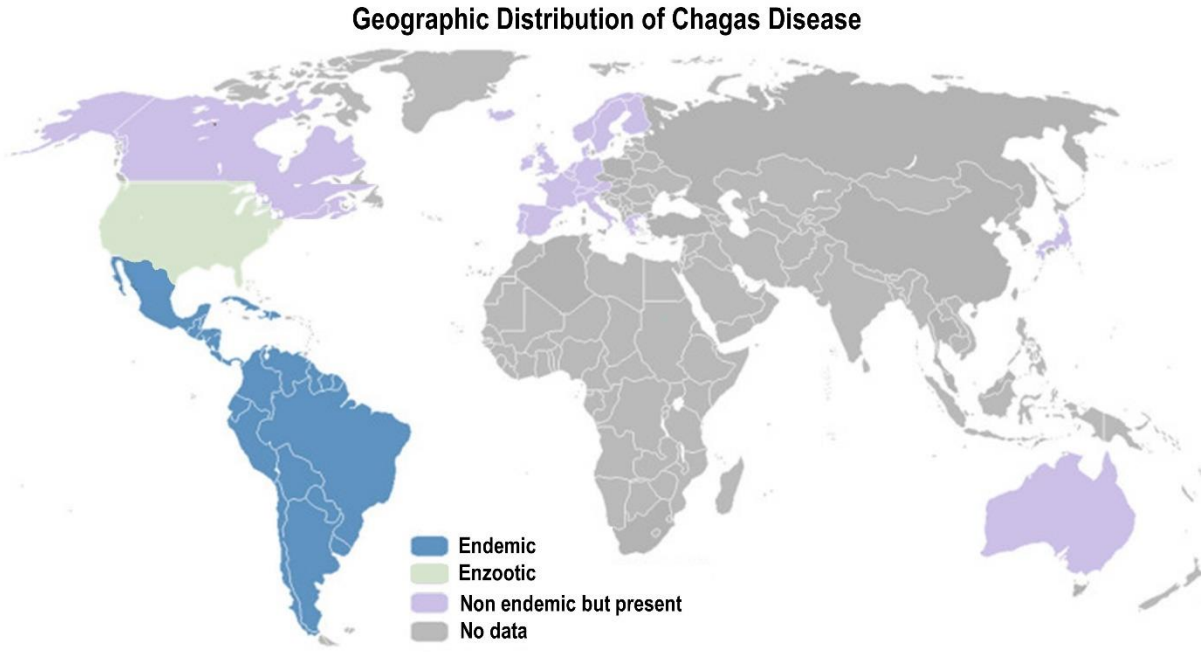


Fig. 1.2.1 Geographic distribution of Chagas disease: *Migration of people from endemic countries of Latin America is driving the global spread of infection to non-endemic countries. Figure adapted from Sangenito et al. 2020* [27].

1.2.2 Clinical manifestations of Chagas disease

T. cruzi establishes a life-long infection which carries the risk of fatal organ dysfunction (Fig.1.2.2). Initial acute infection is characterised by high parasitaemia, which can be easily

detected by microscopy. Typically, this phase is asymptomatic or involves mild flu-like symptoms, such as fever, headache, painful muscles and joints and swelling of lymph glands[14]. Entry of parasites into the eye can cause local swelling due to infiltration of monocytes - this is referred to as Romaña's sign. Similarly, 'Chagoma' occurs when swelling takes place at the insect bite wound. In some cases, and more commonly in children, life-threatening complications such as encephalitis and myocarditis can develop [28]. Oral transmission is associated with a higher risk of severe acute disease [21,22,29], which is probably due to higher inoculative doses of *T. cruzi* [30]. Typically, acute symptoms subside after around 4-8 weeks.

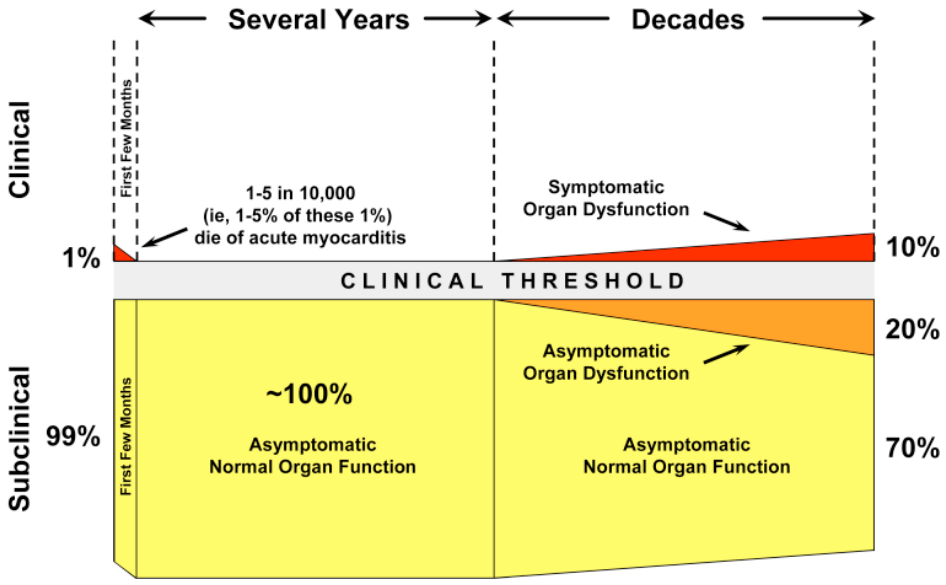


Fig. 1.2.2 Clinical profile of Chagas disease: *Approximately 99% of T. cruzi infections are asymptomatic or mild over the first few months,1% experience severe organ failure such as myocarditis and a small fraction of these people die. The proportion of cardiac disease cases increases in the decades following infection, however most infections (70%) remain asymptomatic for life. Figure from Bonney and Engman, 2015.*

The next phase is characterised by lower parasitaemia, which becomes undetectable by bloodstream microscopy and variably detectable by PCR [28]. In the absence of clinically recognisable pathology this phase is referred to as the indeterminate stage. In contrast, approximately 30-40% of individuals experience chronic disease [31]. In this case, symptoms typically occur decades after acute infection. Progressive and long-lasting inflammatory tissue damage, potentially involving autoimmunity, underlies the development of chronic pathologies, which most commonly involve the heart (~30% of cases) [32]. Cardiac Chagas disease is a

severe cardiomyopathy, characterised by arrythmias and infarctions and carrying the risk of sudden death from dilated cardiomyopathy or congestive heart failure and embolic stroke [33].

Around 10-15% of individuals with Chagas disease develop severe gastrointestinal complications, again carrying a risk of mortality and typically involving the oesophagus and/or colon [34]. Digestive Chagas disease may occur in conjunction with cardiac pathology. In severe cases, it involves enlargement of digestive organs, referred to as 'mega-syndromes' [35]. Digestive tract hypertrophy results in peristaltic and ischaemic disturbances and an elevated risk of carcinoma [36]. The specific cause is not well understood but it has been suggested that inflammatory damage, in combination with age-related enteric nerve loss, plays some involvement [35].

Evidence suggests that the anti-parasitic drug treatments in current use offer little benefit to established pathologies [37]. For this reason, heart transplants are usually required to deal with cardiac Chagas disease. Whilst they can be successful, the risk of reactivation of latent infection resulting from immune suppression is an important barrier to recovery [38]. WHO recommends that all patients with positive serology receive curative treatment [28], however only around 1% of cases in endemic areas are treated [39]. This is partly due to insufficient diagnostic testing and partly because of unpleasant side-effects of treatments, which for some are not worth risking when there is a high chance of remaining asymptomatic.

Due to the chronic and progressive nature of disease, extent of surgical requirements and extensive treatment periods, the estimated global economic cost of Chagas disease is extremely high, at 7 billion US dollars per year [40].

1.2.3 Evolution and classification of *Trypanosoma cruzi*

Carlos Chagas discovered the agent of disease, *T. cruzi*, over 100 years ago. It is an evolutionarily ancient eukaryote, which is believed to have evolved several million years ago from a chiropteran (bat)- restricted ancestor [12]. This theory is supported by the fact that the closest relative is *T. marinkellei*, a parasite of South American bats. *T. marinkellei* diverged from *T. cruzi* lineages approximately 6.5-8.5 million year ago [41]. While still classed as a single species, sequencing has revealed extensive genetic diversity amongst *T. cruzi* lineages, which are classified into discrete typing units (DTUs) based on phylogeny and sequence analysis of several gene loci [42,43]. This includes single nucleotide polymorphisms (SNPs) in 3 housekeeping genes; rDNA, Hsp60 and glucose phosphate isomerase**.** Originally, six DTUs

TcI-TcVI were assigned, but more recently, a seventh has come to be recognised, to encompass the chiropteran genotype TcBat [44]. Whereas TcI-IV are homozygous, TcV and TcVI are heterozygous genetic hybrids, originating from TcII and TcIII. TcVI includes the CL Brener strain which was the first to have its genome sequenced and is commonly used in laboratory research [45].

Certain strains or DTUs have been associated with particular epidemiological and ecological profiles. Genotypic associations with disease symptoms and drug responsiveness have been made, but with limited evidence [46]. There is also a genotype-host relationship: while most DTUs are found in both sylvatic (wild) and domestic hosts (humans and farm animals), certain types are found more frequently in sylvatic stages, including opossums, monkeys and coyotes, while others are primarily associated with domestic hosts [42].

1.2.4 The life cycle of *Trypanosoma cruzi*

T. cruzi has a complex life-cycle, undergoing multiple transformations to complete its journey from insect vector to mammalian host and back again (Fig. 1.2.4). As described, inoculation of the mammalian host typically occurs when the vector takes a blood meal and excretes parasites in its faeces. This infectious life-cycle stage is the metacyclic trypomastigote, an extracellular, flagellated and slender cell form. Unable to replicate, this stage is instead adapted for evasion of the complement system and for host cell invasion - it has the remarkable ability to infect any mammalian nucleated cell [47]. The processes underlying host cell invasion have been extensively studied [48,49]. To provide a brief overview, parasite surface molecules interact with the host glycocalyx, providing anchorage and then triggering host signalling receptors that mediate a rise in intracellular calcium levels in both host and parasite cells [48]. This in turn leads to disruption of the cytoskeleton and fusion of lysosomes with the plasma membrane. Mobilised lysosomes act as donor membranes for internalisation of the trypomastigotes by poorly characterised pathways, which envelope the parasite within a parasitophorous vacuole (PV) in the host cell cytoplasm. Inside the PV, differentiation into the amastigote stage begins, until parasite pore-forming toxin release, facilitated by increased H^+ concentration, triggers escape from the PV. Differentiation is then completed within the cytosol. Amastigotes are smaller, ovoid cells which lack a prominent flagellum and are therefore non-motile. As a replicative life-stage, they use their cytoplasmic location to access host nutrients, fuelling multiple rounds of binary fission. During chronic infection, over a thousand parasites can be present within a single host cell [50]. After several divisions, progeny differentiate into bloodstream trypomastigotes which are similar in appearance to the

metacyclic bloodstream form and are also replication-arrested and complement-resistant. Release of trypomastigotes from the host cell typically involves its rupture but can be non-lytic. Trypomastigotes disseminate throughout blood, lymph and other tissue fluids and go on to invade new cells and initiate subsequent replication cycles. Some circulating parasites will be ingested by triatomines during feeding, which continues the life cycle in the insect. Inside the insect midgut, ingested trypomastigotes differentiate into epimastigotes, another flagellated slender form. Epimastigotes multiply by binary fission, amplifying the population prior to differentiation of progeny into metacyclic trypomastigotes.

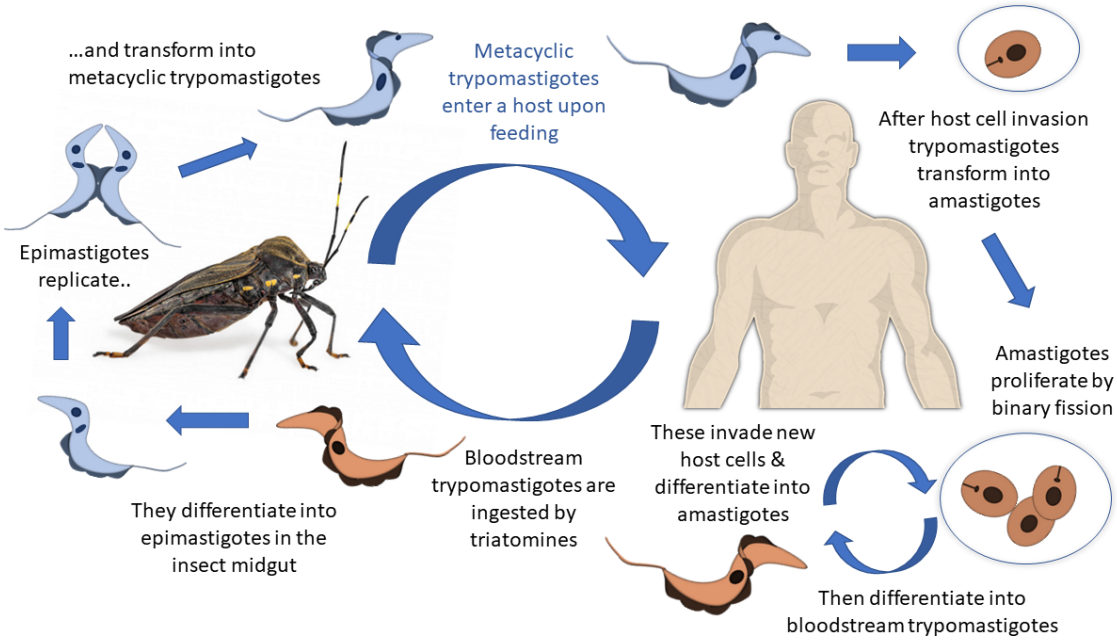


Fig. 1.2.4 Life Cycle of Trypanosoma cruzi: *Metacyclic trypomastigotes are inoculated into the mammalian host where they invade nucleated host cells. First, parasites are encapsulated within a parasitophorous vacuole, where they begin to differentiate, a process which is completed in the host cell cytosol. The differentiated amastigote undergoes several rounds of cell division. Progeny then differentiate into motile trypomastigotes, leading to rupturing of the plasma membrane and dissemination of infectious stages. Bloodstream trypomastigotes enter new host cells and initiate subsequent replication cycles. Triatomines ingest circulating parasites which differentiate into the epimastigote form in the insect midgut. Epimastigotes undergo many cell division cycles and progeny differentiate into metacyclic trypomastigotes for faecal infection of the next host. Figure generated using PowerPoint.*

1.3 Human African Trypanosomiasis

In sub-Saharan Africa, kinetoplastids of the *Trypanosoma* genus cause Human African Trypanosomiasis (HAT), colloquially known as 'sleeping sickness', and Nagana, a disease of cattle and other non-human vertebrates. HAT has caused three deadly epidemics, most recently in 1970-1990s [51]. Its prevalence is highest in the Democratic Republic of the Congo. Two subspecies of *T. brucei* cause HAT: *T. brucei gambiense* causes a more prevalent, slow-progressing variation of the disease and *T. brucei rhodesiense* causes a fast-progressing, more deadly form. The parasites are transmitted by tsetse flies of the Glossina genus.

The hematophagous flies release infective salivary trypomastigotes into the bloodstream of a human host during blood-feeding. Being extracellular parasites, trypomastigotes persist in the host bloodstream and lymphatic fluids. The two mammalian forms are long-slender proliferative and non-dividing short, stumpy forms. If ingested by another tsetse fly, the stumpy parasites complete their life-cycle by differentiating into the insect procyclic form in the insect's midgut. This stage proliferates and migrates to the salivary gland. Here, differentiation occurs, forming the non-dividing, infectious metacyclic form, which participates in subsequent infection cycles.

HAT has two clinical stages: a haemolymphatic stage and a central nervous system (CNS) stage. In the first, proliferation of bloodstream form parasites throughout the peripheral circulation and lymphatic system causes symptoms such as anaemia, fever, weight loss, swelling and painful joints. After a period of weeks to several months, depending on the sub-species of *T. brucei*, progression to the CNS stage occurs. This is characterised by the invasion of CNS tissue by trypomastigotes and inflammatory infiltrate, with the choroid plexus and circumventricular organs being primary sites of infiltration due to their relatively higher blood-brain barrier (BBB) permeability [52–54]. Histopathology reveals meningitis, perivascular inflammation and in later stages, encephalitis [55,56]. Such damage to vital processing centres causes psychiatric and other neurological symptoms; the afflicted individual may exhibit strange behaviour and/ or experience painful sensations, muscular abnormalities and sleep disturbances [57]. Characteristically at this stage, one may experience an uncontrollable desire to sleep at any time of day, often accompanied by insomnia at night, hence the name 'sleeping sickness'. In the final stages of illness, protracted brain damage causes violent seizures, progressing to coma and ultimately death.

A remarkable adaptation of *T. brucei* to survival in its host is the dynamic expression of dense glycoprotein coats that are highly immunogenic and continuously replaced to present the host

immune system with everchanging antigens. Meanwhile, the glycoproteins shield the underlying membrane from innate immunological attacks. Within the mammalian host, between 1×10^{-3} and 1×10^{-6} coat switches are estimated to occur per population doubling [58,59].

HAT has been quite successfully controlled over the past 20 years by vector-targeted and diagnostic surveillance programs and the estimated number of actual cases is now below 20 000, with less than 700 being reported in 2021 [51]. However, huge epidemic outbreaks remain a possibility, and the drugs used for treatment carry the risk of severe adverse effects (discussed in Chapter 1.5).

1.4 Leishmanial diseases

The leishmaniases are the most widespread of all the kinetoplastid diseases, endemic to around 100 countries throughout the tropics, subtropics and Mediterranean basin, where ~1 million cases are thought to occur each year [3,60]. Disease can present as cutaneous, mucosal or visceral leishmaniasis, depending on the species of *Leishmania,* amongst other factors. 31 species are known to cause disease in mammals, at least 20 of which are pathogenic to humans [61,62]. Other mammals which are susceptible to infection include dogs, which are clinically important reservoir hosts [63]. The parasites are transmitted by around 100 species of female sand flies of the *Phlebotomus* and *Lutzomyia* genera [64].

The *Leishmania* life cycle consists of two main cell forms: the long-slender promastigote and the ovoid amastigote. In a similar fashion to HAT, the infectious parasite stage, which is the metacyclic promastigote, is delivered to a mammalian host from the hematophagous vector during blood feeding. Inside the mammal, motile promastigotes are phagocytosed by macrophages, dendritic cells and other immune cells, via complement receptor-1 activation [65,66]. This results in encapsulation of the parasite within a phagolysosome or PV, within which differentiation into the intracellular amastigote stage occurs. Unlike, *T. cruzi*, the *Leishmania* amastigote stays within the PV, undergoing cell division and scavenging nutrition until it eventually lyses the host cell, releasing progeny to infect nearby cells. The life cycle continues when a sand fly ingests infected phagocytes or free amastigotes during blood feeding. In the insect midgut, amastigotes differentiate into procyclic promastigotes, which are a replicative stage. Following replication, differentiation into the metacyclic promastigote occurs and these forms migrate to the proboscis for infection of a new host.

A number of mechanisms protect *Leishmania* from immune destruction during its life cycle. These ultimately suppress the activation of macrophages and consequential parasite destruction by reactive oxygen and nitrogen species. Th2 cytokines have been shown to play a role in this process by increasing arginase-1 expression, resulting in reduced generation of nitric oxide by iNOS [67]. Indirect methods of macrophage suppression by *Leishmania* include inhibition of antigen-presentation by host cells [68] and modulation of T cell functions e.g. causing the expansion of counter-protective T helper cells and inhibiting the expansion of memory and effector T cells [69]. Amongst other cytokines, IL-10 and TGFβ are important mediators of immunosuppression in leishmaniasis [70].

The most common clinical manifestation of disease is cutaneous leishmaniasis (CL), whereby the presence of parasites and infected macrophages in tissues in the bite region causes ulcers or nodular skin lesions to develop. These are usually self-healing pathologies but their greater impact, it seems, is the social stigma associated with their appearance [71]. The metastasis of parasites and infected macrophages throughout the skin and visceral organs can lead to diffuse skin lesions, and more serious mucocutaneous and visceral disease. Mucocutaneous leishmaniasis (ML) involves the ulceration and erosion of mucosal tissue of the mouth and nose, initially at the bite site and then spreading to other mucosal tissues as parasites metastasise. Progressive disease can involve pharyngeal mucosa, facial cartilage and upper airway structures, resulting in disfiguration, secondary infection and airway obstruction [72]. Most cases of ML occur in Bolivia, Brazil and Peru and are caused by strains of the Viannia subgenus (mainly *L. braziliensis and L. panamensis*) [73]. This form of disease can appear concomitantly with or several years after an initial cutaneous infection and even in patients without any history of CL [74]. The severity of disease seems to depend in part on the ability of the parasite to elicit a Th2-type immune response, characterized by IL-4 production [70], which illustrates the role of host immunology in disease outcome. ML lesions are not self-healing and are more resistant to treatment than the primary lesions, with frequent relapses [74]. The most severe form of disease is visceral leishmaniasis (VL), predominantly caused by *L. donovani* and *L. infantum* [75]. This form is fatal if left untreated in over 95% of cases [60]. Symptoms include fever, weight loss, anaemia, and hepatosplenomegaly (enlargement of the liver and spleen). Most cases occur in Brazil, East Africa and India [75]. Between 50 and 90 thousand new cases of VL are estimated to occur each year, making it a high-risk disease in terms of outbreak and mortality potential, compared with other parasite diseases [60].

1.5 Treatments

1.5.1 A need for new anti-trypanosomatid drugs

While HAT has largely been brought under control in affected areas, less toxic and more broadly effective treatments will still be of great benefit to overcoming this disease. For instance, melarsoprol, the first-line drug for CNS-stage *T. brucei rhodesiense* causes a post-treatment reactive encephalopathy (PTRE) in about 10% of cases [76]. Furthermore, drug resistance has been widely reported for melarsoprol [77] and *in vitro* selection experiments indicate potential for resistance against eflornithine [78,79], another first-line treatment in many endemic regions. The latest treatment to be recommended by WHO is fexinidazole, an orally active compound which is effective against *T.b. gambiense* late-stage disease [80]. Its usefulness is however limited by the requirement for an initial disease severity assessment followed by 10 days of treatment. Promisingly, a single-dose treatment, acoziborole (a benzoxaborole -6-carboxamide) was proven to be effective in Phase II/III clinical trials [81]. This compound was shown to be orally active against *T. b. gambiense* regardless of disease stage, having 95% efficacy in late-stage and 100% efficacy in early- intermediate stage disease at 18 months, based on group sizes of 167 and 41 patients respectively.

In contrast to HAT, both Chagas disease and leishmaniasis are a long way from being eliminated. Despite this, effective vaccines are lacking and therefore treatment again depends on chemotherapeutic agents. Chagas disease can be treated with two nitroheterocyclic compounds, benznidazole (BZ) and nifurtimox (NF) (discussed in more detail in 1.5.2), and leishmaniasis is treated mainly with pentavalent antimonial and amphotericin B [61]. The pentavalent antimonial sodium stibogluconate, which has been in use for a century, is still the frontline drug for leishmaniasis in several countries. This is despite its known toxicity, high production costs and cumbersome regimen of intravenous or intramuscular administration [61]. Its mechanism is unclear but potentially involves binding to thiol groups in the parasite and inhibition of adenosine and guanosine triphosphate synthesis [82]. Amphotericin B, an anti-fungal agent synthesised by certain *Streptomyces* species is the first-line treatment for VL in many parts of the world [83]. It acts by disrupting plasma membranes upon binding to constituent ergosterols [84]. When delivered as a liposomal formulation (Ambisome) or in the form of amphotericin B desoxycholate (Fungizone), its toxicity is limited. However, it must be injected, again is expensive and furthermore requires cold storage [83]. The first-line treatment for CL caused by *L. guyanensis* and *L. panamensis* is pentamidine isethionate, an aromatic diamidine, which is also administered by intramuscular injection or intravenous infusion [85]. The drug mainly targets the mitochondrion and nucleus of trypanosomatids, causing reduced mitochondrial membrane potential and disrupting activities

of mitochondrial topoisomerase, S-adenosyl-L-methionine decarboxylase and nucleoside triphosphate diphosphohydrolases [86,87]. Other less widely used compounds include miltefosine, a repurposed anti-cancer drug used orally for VL treatment [88] and paromomycin, an aminoglycoside antibiotic used topically for CL or administered by intra-muscular injection to treat VL. Both carry toxicity risks: miltefosine is a teratogen with a long half-life [89] which also predisposes it to resistance [90], whilst paromomycin can cause hepatotoxicity [91] amongst other effects. Resistance to paromomycin has also been demonstrated in laboratory-based studies [92]. In light of side-effects, expense, administration difficulties and emerging resistance, there is a clear need for new anti-leishmanial drugs.

1.5.2 Treatment limitations for Chagas disease

Nifurtimox (NF), a synthetic nitrofuran was the first drug licensed for Chagas disease, but has now largely been replaced with benznidazole (BZ), a nitroimidazole-derivative (Fig. 1.5.2.A). They are both pro-drugs, which are activated by the parasite-specific mitochondrial type 1 nitroreductase (*Tc*NTR-1) to exert their cytotoxic effects [93,94]. *Tc*NTR-1 catalyses two-electron reduction of the nitro (NO2) group on both compounds using flavin mononucleotide (FMN) as a cofactor. This reduction pathway produces hydroxylamine derivatives, which can further react to generate highly electrophilic intermediates, which are cytotoxic, causing damage to DNA and other macromolecules. The trypanocidal activity of NF has been specifically linked to saturated open-chain nitrile derivatives produced by nitroreduction (Fig. 1.5.2.B) [93]. A major drawback of both compounds requiring *Tc*NTR-1 for activity is the risk of cross-resistance that this presents, which can readily be induced *in vitro* by diminished *Tc*NTR-1 expression and/or inactivity of the enzyme [95–97]. Cultivated trypanosomes maintained under BZ selection readily delete a copy of the *Tc*NTR gene [97,98], or even an entire chromosome in which it is located [96]. Inactivating mutations in the FMN-binding region of the *Tc*NTR have also been linked to BZ resistance of cultivated strains [96]. If NTR function is lost in clinically relevant strains of *T. cruzi*, there will be a shortage of viable treatments.

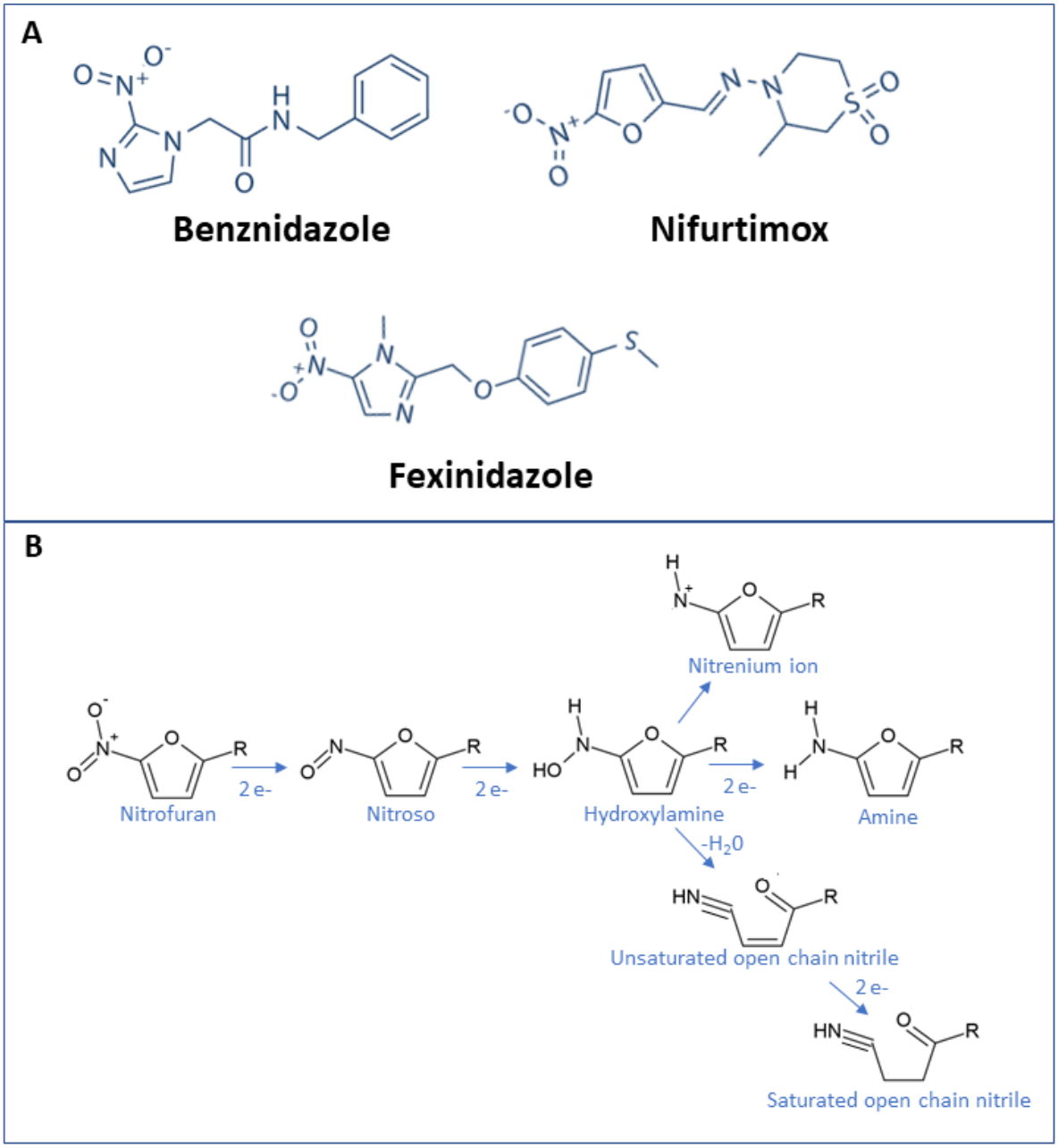


Fig. 1.5.2 Anti-chagasic nitro-compounds and their nitro-metabolites: (A) *Structures of frontline drugs benznidazole and nifurtimox and anti-chagasic agent fexinidazole (in clinical trials).* ***(B)*** *Reduction of nitrofurans such as nifurtimox by type I nitroreductases. Type I nitroreductases reduce the conserved nitro group of nitrofurans to generate a nitroso intermediate followed by a hydroxylamine. The hydroxylamine can be further metabolised to form either a nitrenium ion, an amine, or unsaturated and then saturated open-chain nitriles. Figure generated using chemsketch.*

Additional drawbacks of these compounds include the potential for treatment failure, particularly for more established infections [99,100]. One study on the efficacy of BZ treatment in chronic disease found that treatment failure occurred in 7-45% of adults, based on haemoculture, PCR and clinical assessment [101]. Furthermore, these drugs require long treatment regimens (typically 30-120 days), which is inconvenient and expensive - particularly for NF, which requires around 10 mg per kg of body weight in three daily doses [102,103]. On top

of this, the main limitation of these drugs is their toxicity, which can prevent their use in patients with established pathology. For example, NF frequently causes gastrointestinal effects, headache and sleeping disorders [103]. In comparison, BZ is better tolerated, yet adverse effects are still common, particularly cutaneous reactions such as dermatitis, which occur in at least half of patients [104]. Other common adversities are paraesthesia, peripheral neuropathy, anorexia, and weight loss. In severe cases, decreased bone marrow, thrombocytopenic purpura, and agranulocytosis can occur. As well as posing additional health risks, side effects, along with the invasiveness of the administration routes, limit the efficacy of these compounds by discouraging completion of the treatment regimen. Treatment interruption is reported to occur in at least 12% of patients, with some studies estimating as many as 41.5% [37,104,105]. As well as limiting the effectiveness of treatment this may permit the development of resistance.

A newer compound which initially showed promise as a Chagas disease treatment is the anti-fungal ergosterol inhibitor posaconazole. This drug significantly diminished parasite load in mouse models, but not to the extent of sterile cure [106] and human trials have been disappointing [107]. Fexinidazole, another nitroaromatic compound, dependent on *Tc*NTR-1 (Fig.1.5.2), showed pre-clinical efficacy against Chagas disease [108–110] and recently underwent phase II clinical trials [80], but again carries the risk of cross-resistance. More promisingly, the benzoxaborale compound, AN15368, recently proved to be highly effective and safe for the treatment of *T. cruzi* infection in non-human primates [111]. Its potency, which stems from inhibition of mRNA processing, extends to multiple genetically distinct *T. cruzi* lineages. The results of *in vitro* and *in vivo* assessment indicate that this is the most promising clinical candidate in last 50 years *T. cruzi* research.

Despite the promise of recent discoveries, to be well-equipped to combat Chagas disease, there is a considerable need for new well-tolerated drugs that are effective at all disease stages, amongst all clinically important lineages and which do not depend on nitroreductases for their activity.

1.6 Coenzyme A synthesis as a target for anti-parasitic drugs (literature review)

1.6.1 Coenzyme A and its biosynthesis

Coenzyme A (CoA) is an essential co-factor found in all organisms. Like other co-factors it is broadly used in metabolism, being required by at least 80 known enzymes [112]. Its primary role

is to activate acyl groups, which occurs by reactivity of the terminal sulfhydryl (—SH) group. The acyl group RCO, containing C=0, comes in many forms. Important acyl derivatives of free CoA, which are formed via thioester linkage of a carboxylic acid, are acetyl-CoA, malonyl-CoA and succinyl-CoA (Fig.1.6.1.1). Since thioester bonds are energy rich, the chemical structure of acyl-CoAs facilitates the transfer of the acyl moiety to a variety of acceptor molecules [113]. Thus, they act as metabolic intermediates and precursors that are essential to ubiquitous energy metabolism and biosynthesis pathways.

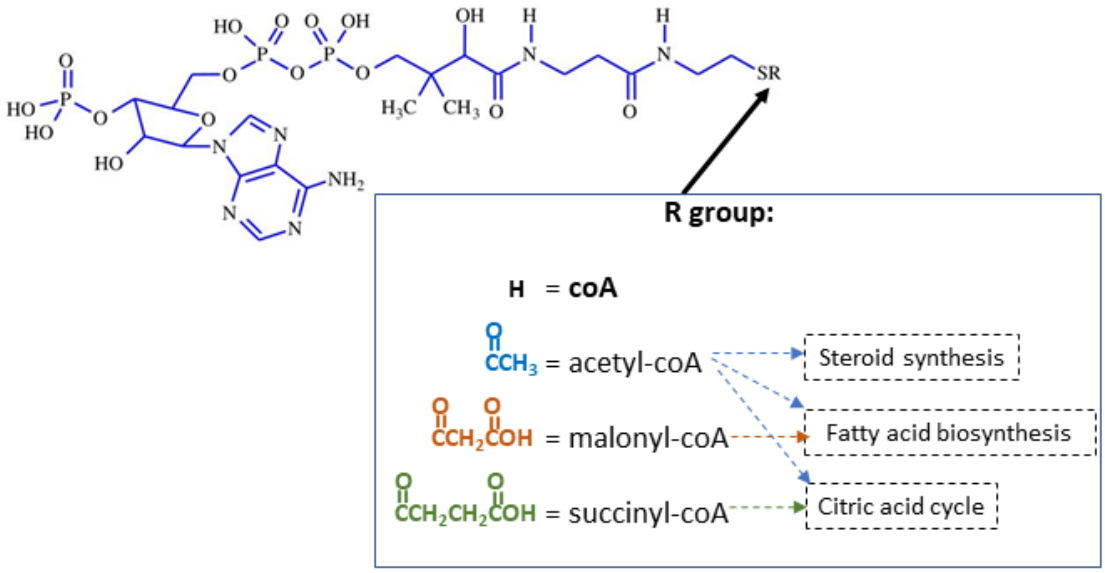


Fig. 1.6.1.1 Structure of Coenzyme A (CoA) and its acyl derivatives*: Illustration of the chemical groups of metabolically important acyl-CoAs: acetyl-CoA, malonyl-CoA and succinyl-CoA, which are required for ubiquitous pathways such as steroid synthesis, fatty acid synthesis and the citric acid cycle. Figure generated with chemsketch.*

The CoA precursor is pantothenate (vitamin B5), which in some organisms is synthesised *de novo* while others acquire it from exogenous sources [113]. Most bacteria, plants and fungi perform de *novo* synthesis, occurring in 3 steps (Fig.1.6.1.2 A):

1. hydroxymethyl transfer of α-ketoisovalerate, a metabolite of valine, to form α-ketopantoate;
2. reduction of α-ketopantoate to pantoate;
3. ligation of pantoate with β-alanine to form pantothenate.

These reactions are performed by the respective enzymes ketopantoate hydroxymethyl transferase (KPHMT or panB), ketopantoate reductase (KPR or panE) and pantoate-β-alanine ligase, also called pantothenate synthase (PBAL or panC). A key source of β-alanine for this third step is L-aspartate, which is decarboxylated by the enzyme panD.

Animals and most protozoa, including kinetoplastids lack these biosynthesis enzymes and instead acquire pantothenate from dietary sources or host tissues.

Generated or salvaged pantothenate is then converted to CoA in 5 steps (Fig. 1.6.1.2B):

1. Pantothenate kinase (PanK or CoaA) catalyses the ATP-dependent phosphorylation of pantothenate, producing 4'-phosphopantothenate (PPan).
2. Phospho-pantothenoylcysteine synthetase (PPCS) catalyses the transfer of cysteine to PPan, using water, generating 4'-phospho-pantothenoylcysteine
3. 4'-phospho-pantothenoylcysteine decarboxylase (PPCDC) forms 4'-phosphopantetheine (PPanSH). The function of PPCDC can alternatively be bypassed by the promiscuous activity of some PanKs, which can catalyze the phosphorylation of pantetheine (PanSH).
4. Phosphopantetheine adenylate transferase (PPAT) generates dephospho-CoA using AMP.
5. CoA is formed by dephospho-CoA kinase (DPCK), which contributes a third phosphate group.

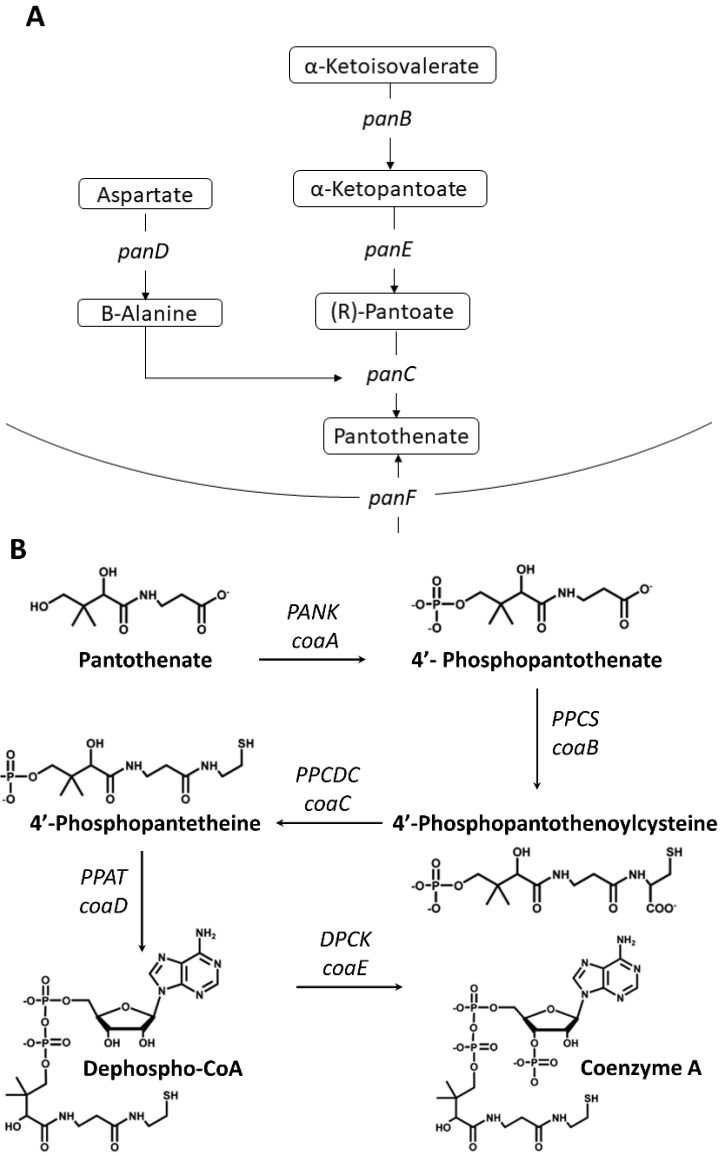


Fig. 1.6.1.2. CoA Biosynthesis: *(**A**) In some organisms, such as bacteria, de novo pantothenate synthesis occurs from α-ketoisovalerate, catalysed by ketopantoate hydroxymethyl transferase (panB), ketopantoate reductase (panE), and pantoate-β-alanine ligase (panC or pantothenate synthase). The final reaction uses β-alanine, acquired from L-aspartate by aspartate decarboxylase (panD). Pantothenate uptake by the transporter panF can bypass de novo synthesis. (**B**) To form CoA from pantothenate, pantothenate is phosphorylated to 4'-phosphopantothenate by pantothenate kinase (PANK), then condensed with cysteine by 4'-phosphopantetheneoylcysteine synthase (PPCS) and then decarboxylated to form 4'phosphopantetheine by 4'-phosphopantothenoylcysteine decarboxylase (PPCDC). 4'phosphopantetheine is AMPylated by phosphopantetheine adenylyltransferase (PPAT) and the resulting dephospho-CoA phosphorylated by dephospho-CoA kinase (DPCK). Figure generated with chemsketch.*

1.6.2 The metabolic importance of acyl-CoAs

The most widely used CoA derivative is acetyl-CoA, which links several biosynthetic and catabolic pathways, as well as having important non-metabolic roles [114]. In trypanosomes, synthesis of this molecule requires acetate from amino acids, fatty acids or glucose [115]. The preferred source for *T. brucei* PCF and BSF stages is threonine, salvaged from vector or host

tissues [115,116]. L-threonine dehydrogenase produces amino-oxobutanoic acid which is then converted, using free CoA, to acetyl CoA and glycine by the enzyme glycine C-acetyltransferase. As well as threonine, alanine, isoleucine and leucine can all be converted to acetyl-CoA, with leucine being an important source for promastigotes of *L. mexicana* [117].

The preferred source of acetyl-CoA in amastigotes of *Leishmania* spp. and in *T. brucei* adipose-tissue forms (ATFs) is fatty acid oxidation [118]. This occurs in the mitochondrion and in glycosomes (organelles specialised for glycolysis). The process of β-oxidation (so called because the β-carbon of the fatty acid is oxidised to a carbonyl group) begins with the dehydrogenation of acyl-CoA. Host fatty acid sources such as myristate have been demonstrated to fuel the pathway in *T. brucei* ATFs [118]. In this initial step, trans-enoyl-CoA is formed by acyl-CoA dehydrogenase. Enoyl-CoA hydratase then converts this to L-β-hydroxyacyl-CoA. In the third step, the latter is dehydrogenated by β-hydroxyacyl-CoA dehydrogenase. Finally, β-ketoacyl-CoA and free CoA are converted to acetyl-CoA and an acyl-CoA by β-ketoacyl-CoA thiolase (Fig.1.6.2.1).

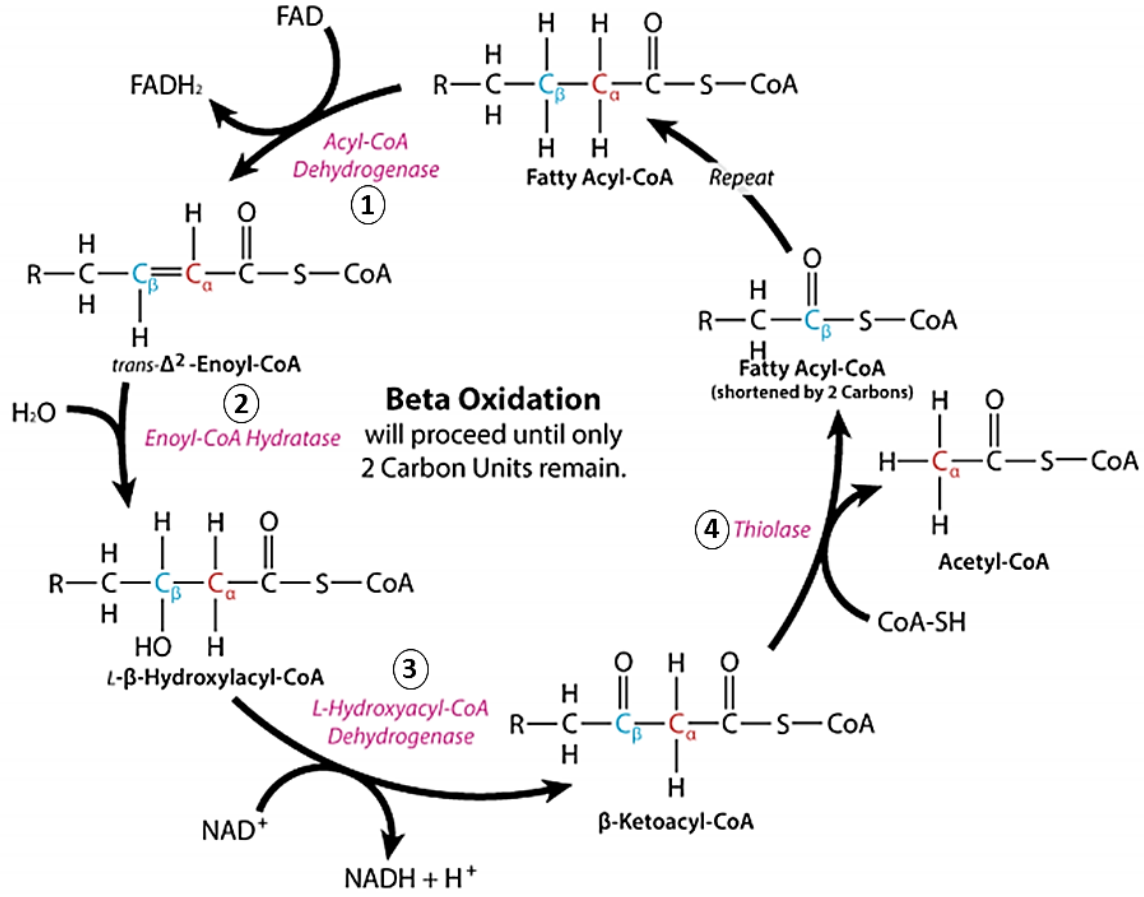


Fig. 1.6.2.1 Fatty acid oxidation. *Step 1: Dehydrogenation of fatty acyl-CoA derived from myristate and other host or parasite fatty acids. Step 2: Saturation of trans-enoyl-CoA by enoyl-CoA hydratase. Step 3: Dehydrogenation of L-β-Hydroxyacyl-CoA. Step 4: -ketoacyl-CoA thiolase converts β-Ketoacyl-CoA and free CoA to acetyl-CoA and an acyl-CoA. The latter undergoes subsequent cycles of oxidation until 2 carbon units remain. Adapted from Ahern and Rajagopal, 2014* [119].

Glucose can also be a source of acetyl-CoA, via oxidation to pyruvate: this is the glycolytic pathway. Trypanosomatids perform most of this pathway in glycosomes, with the final steps performed in the cytosol [120]. Pyruvate is oxidised to acetyl-CoA by the pyruvate dehydrogenase complex [121], using CoA as a co-substrate [122].

Sodium stibogluconate (SSG), a pentavalent antimonial, is the first-line drug for all forms of Leishmaniasis and has been since the 1940s [75]. It is thought to act via inhibition of fatty acid β-oxidation and glycolysis [82,123]. *In vitro* exposure of *Leishmania amazonensis* to the drug and development of resistant strains was associated with upregulation of a putative acetyl-CoA synthetase and of acetyl-CoA carboxylase [124], which highlights the link between these pathways and acetyl-CoA homeostasis and the importance of this molecule to these parasites.

One of the vital pathways fuelled by acetyl-CoA is the tricarboxylic acid (TCA) cycle, which occurs in mitochondria (or the cytosol in prokaryotes). It is the final oxidative catabolism pathway, which releases energy from carbon skeletons derived from carbohydrates, amino acids, and fatty acids and produces glutamate, which is used in protein synthesis and the generation of glutathione (and trypanothione in kinetoplastids). Each oxidative step reduces a coenzyme including nicotinamide adenine dinucleotide (NADH) and flavin adenine dinucleotide (FADH2). NADH and FADH2 reducing agents are used in ATP generation via the electron transport chain. The role of acetyl-CoA in this cycle is provision of the acetyl group for transfer to oxaloacetate in a reaction which forms citrate. Free CoA is also required in the cycle for the oxidative decarboxylation of α-ketoglutarate, which generates NADH and succinyl-CoA. Succinyl-CoA is then used for synthesis of succinate which donates electrons to the ATP-generating electron transport chain when converted to fumarate.

Most organisms, including trypanosomatids, require this pathway. *T. brucei* BSFs are an exception, as the bloodstream of its host provides sufficient glucose for glycolysis to satisfy its energy requirements [125]. PCFs and adipose tissue forms (ATFs) of *T. brucei,* on the other hand, do use the citric acid cycle. In PCFs, certain stages of the pathway are used rather than a full cycle [126]. Important cycle products for this parasite stage include succinate (for electron provision) and malate, which is used for gluconeogenesis in the absence of glucose [126]. In ATFs, the TCA cycle is the main source of energy and is fuelled mainly by β-oxidation of host lipids, particularly myristate [118]. The TCA cycle was also proven to be essential to *Leishmania* amastigotes. 13C-labelling and chemical inhibition studies suggested that *L. mexicana* uses the TCA cycle for ATP and glutamine synthesis. Glutamine provides an essential amino group for synthesis of pyrimidines and amino sugars in promastigotes and amastigotes [127,128]. Sodium fluoroacetate and methionine sulfoximine which inhibit the TCA cycle enzyme

mitochondrial aconitase and TCA-fuelled glutamine synthetase respectively, strongly inhibited growth and virulence of amastigotes [128]. In *T. cruzi,* the TCA cycle seems to be active, based on radioisotope labelling experiments showing incorporation of carbon from 14C-labelled glucose and acetate into TCA cycle intermediates [129].

As well as energy production, a number of *de novo* synthesis pathways required by trypanosomatids use acetyl-CoA and other acyl-CoAs. An important example is fatty acid (FA) biosynthesis (Fig.1.6.2.2). *De novo* synthesis of myristate (C14) is important to BSF *T. brucei* for the synthesis of glycosylphosphatidylinositol (GPI) anchors of their variant surface glycoprotein (VSG) coats [130]. VSG cell surface proteins mediate antigenic variation and thus evasion of host immune systems. On the other hand, procyclic forms of *T. brucei* and life-stages of *Leishmania* and *T. cruzi* mainly produce the fatty acid stearate (C18) [117,131]. Like other FA-synthesis systems, malonyl-coA is the precursor for fatty acid synthesis in trypanosomes [115].

In most organisms, malonyl-CoA is made from acetyl-CoA by acetyl-CoA carboxylase. The acyl group of malonyl-CoA is then 'activated' or transferred to Acyl Carrier Protein (ACP) by ACP transacylases. ACP itself contains a 4'-phosphopantetheine moiety derived from CoA using holo-ACP synthase (AcpS) [132]. Malonyl-ACP is the building block for carbon chain extension by the reaction of ACP-bound acyl groups in reactions catalysed by 3'ketoacyl-ACP synthase. After chain extension, the carbon 3 (C3) ketone is converted to a methyl group in three steps, involving reduction of C=O to C-OH_2, elimination of OH_2, which forms a C2-3 double bond, and then reduction of this double bond. In animals and fungi, the enzymes which catalyse these reactions are combined as one or two large multi-domain cytosolic proteins, described as type I synthases [125]. Plants and bacteria instead use many discrete cytosolic or plastidic enzymes, described as a type II FA synthase system. The genomes of trypanosomes encode components of a conventional type II FA synthase system that localise to the single mitochondrion, however, their main route of fatty acid synthesis appears to use a CoA-dependent elongase (ELO) pathway (Fig.1.6.2.3), which occurs in the endoplasmic reticulum (ER) [125]. Other eukaryotes, in contrast, use elongase pathways to extend pre-existing long fatty-acyl coAs(>C14) but do not carry out *de novo* synthesis.

In the trypanosomatid ELO pathway, malonyl-CoA, synthesised from acetyl-CoA is utilised to extend an acyl-CoA primer, the smallest being made of 4-carbons (butyryl-CoA). This forms β-ketoacyl-CoA, catalysed by β-ketoacyl-CoA synthase, also named elongase (ELO). In the next step, β-ketoacyl-CoA reductase reduces β-ketoacyl-CoA to β-hydroxyacyl-CoA. This is then dehydrated by β-hydroxyacyl-CoA dehydrase. The resulting acyl-enoyl-CoA is reduced

by *trans*-2-enoyl-CoA reductase to produce a saturated acyl-CoA chain. The main difference between these reactions and typical fatty acid synthesis is the direct use of acyl-CoAs, without conversion to acyl-ACPs. In *T. brucei,* three ELOs produce saturated fatty acids from different starting carbon-chain lengths: ELO1 converts C4 to C10, ELO2 extends the chain length from C10 to myristate (C14), and ELO3 extends myristate to stearate (C18). *T. cruzi* and *L. major* genomes encode additional ELOs, including *T. cruzi* ELO5 which synthesizes very long-chain FAs by extending C18-C26 substrates [115]. The major product of ELO5 is C24, which is a component of the GPI-anchors of infective trypomastigotes. In *L. major* promastigotes, synthesis of lipophosphoglycan (LPG), required for attachment of parasites to the sand fly mid-gut, is thought to require ELO3 and 4 isoforms [115].

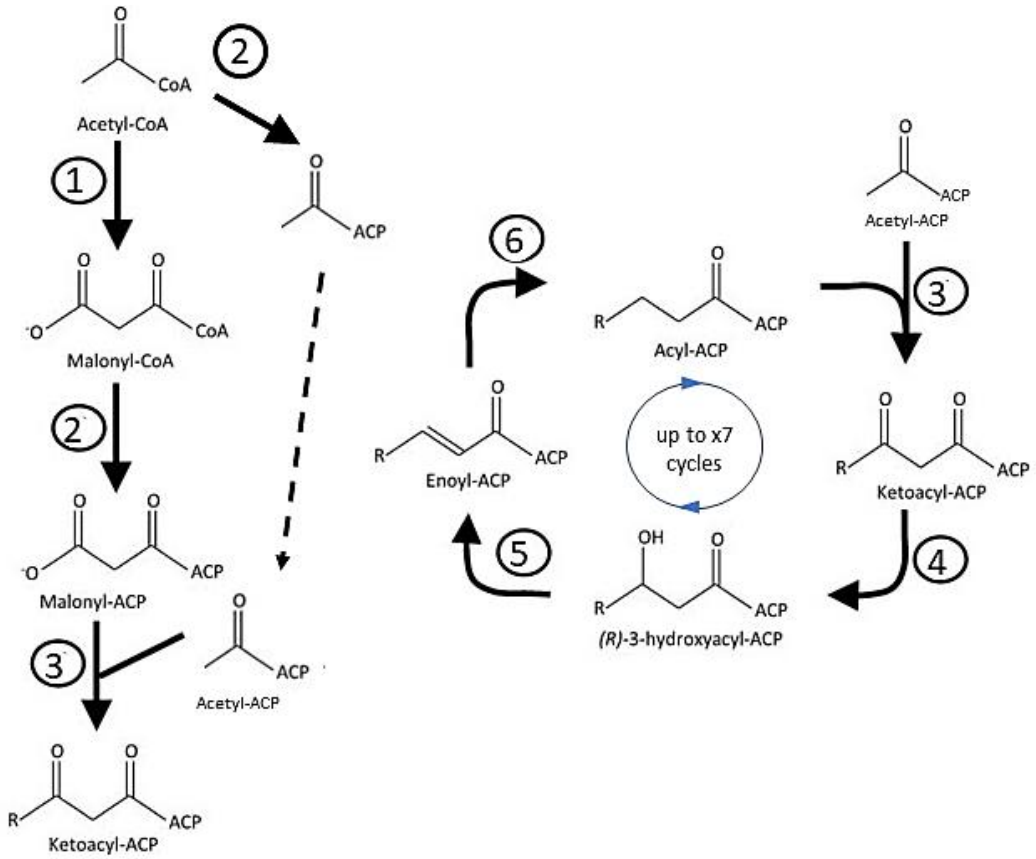


Fig. 1.6.2.2 Fatty acid biosynthesis: *(1) Acetyl-CoA carboxylase converts acetyl-coA to malonyl-CoA (2) The acyl-CoAs are activated by ACP transacylases. (3) β-ketoacyl-ACP is synthesised from acetyl- and malonyl-ACP by β-ketoacyl-ACP synthase (4) β -ketoacyl-ACP reductase reduces the (C3) ketone forming β -hydroxyacyl-ACP (5) this is dehydrated by β-hydroxyacyl-ACP dehydratase forming α,β -trans enoyl-ACP(6) enoyl-ACP reductase reduces the C2-C3 double bond forming butyryl-ACP. This cycle can occur up to x7 (Figure adapted from Riedel et al., 2013)* [133].

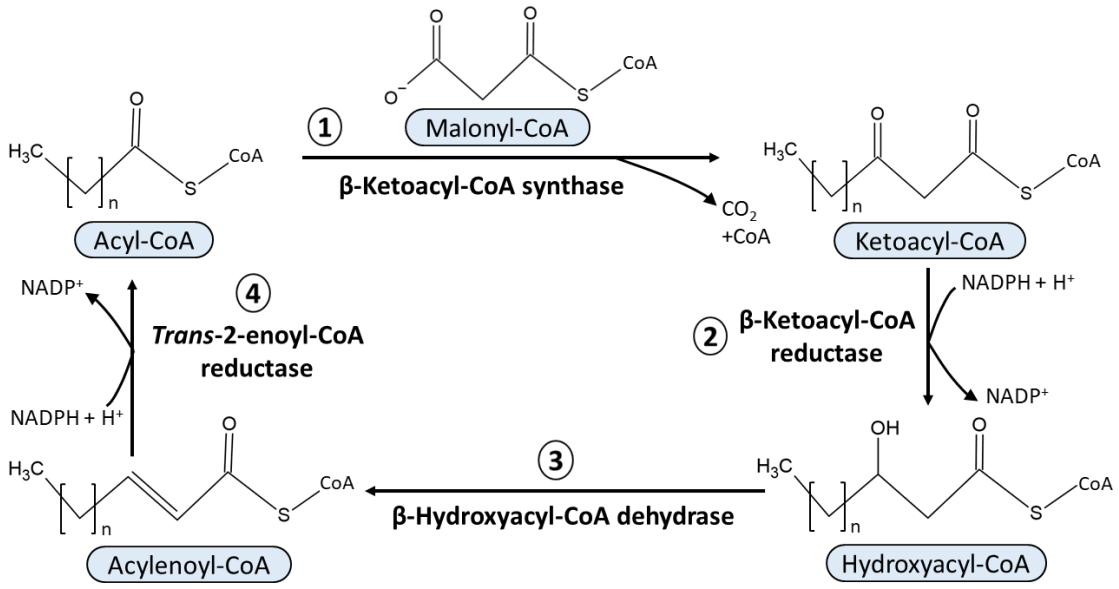


Fig. 1.6.2.3 Elongase pathway of trypanosomatids*: 1: β-ketoacyl-coenzyme A (CoA) synthase (ELO) condenses a long-chain acyl-CoA (where n is an even number) with two carbons from malonyl-CoA to form β-ketoacyl-CoA. 2: β-ketoacyl-CoA reductase reduces β-ketoacyl-CoA to β-hydroxyacyl-CoA using NADPH 3: β-hydroxyacyl-CoA is dehydrated by β-hydroxyacyl-CoA dehydrase. 4: acylenoyl-CoA is reduced by trans-2-enoyl-CoA reductase using NADPH to form a saturated acyl-CoA chain. Figure based on Lee et al., 2007.*

The importance of fatty acid synthesis to trypanosomes is demonstrated by inhibition studies. Two inhibitors have been shown to target keto-acyl synthase of *T. brucei*, thiolactomycin and cerulenin and exert toxic effects on cultured BSF *T. brucei* [130]. Thiolactomycin has also been demonstrated to inhibit mycobacteria and blood stages of the malaria parasite *Plasmodium falciparum*, while cerulenin restricts growth of the intestinal protozoan parasite *Cryptosporidium parvum*, responsible for cryptosporidiosis [134]. Haloxyfop, an aryloxyphenoxypropionate herbicide, inhibits *T. brucei* acetyl-CoA carboxylase (ACC) activity, as well as plastid ACCs in many plants and that of *Toxoplasma gondii* [135].

Another important anabolic pathway requiring acetyl-CoA is the mevalonate pathway, or isoprenoid pathway. This produces isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) which are the five-carbon building blocks used to make isoprenoids such as sterols. Sterols are essential structural components of plasma membranes, which regulate membrane fluidity and are essential for growth of trypanosomatids [117,136,137]. Whereas mammals use cholesterol as their main membrane sterol, most trypanosomatid life-stages (with the exception of BSF *T. brucei*) instead utilise ergosterol-based compounds, as do fungi. The initial synthesis of the isoprenoid skeleton requires acetyl-CoA as the substrate. It is converted to acetoacetyl-CoA and then hydroxymethylglutaryl-CoA (HMG-CoA). HMG-CoA reductase reduces this to malonate, which is twice phosphorylated to form IPP and DMAPP

isomers. Statins such as lovastatin (mevinolin) and synvinolin inhibit this pathway by competitive inhibition of HMG-CoA reductase. When tested on *Leishmania* and *T. cruzi*, lovastatin potentiated anti-proliferative effects of other sterol-based inhibitors, demonstrating the importance of CoA to sterol biosynthesis [122]. *In vitro* growth of *T. brucei* PCFs was also inhibited by statin treatment [138]. Synvinolin as well as being anti-proliferative, induced morphological modifications, including an arrest at cytokinesis and swelling of the kinetoplast-mitochondrion system. The effectiveness of sterol biosynthesis inhibitors is limited by the ability of some kinetoplastid life-stages to obtain and metabolise host cholesterol. To sustain extensive membrane remodelling, BSF *T. brucei* acquire cholesterol from low-density lipoprotein (LDL) by receptor-mediated endocytosis [138]. Other kinetoplastid life-stages which do not ordinarily use cholesterol as a major membrane component can also use it as a substitute when ergosterol biosynthesis is pharmacologically inhibited: *Leishmania* promastigotes have been shown to acquire exogenous cholesterol by endocytosis [139] and this increases during azole treatment [140].

In addition to providing a source of metabolic intermediates for major catabolic or anabolic pathways, CoA and its acetylated form are important for gene expression (reviewed by Pietrocola et al) [141]. Firstly, acetyl-CoA is required for co-translational modification of the majority of proteins. Following the removal of initiator methionine residues from proteins undergoing translation, N-acetyltransferases (NATs) transfer an acetyl group from acetyl-CoA to the α-amino group of the N-terminal residue of the protein. This influences their stability, localization, and function [142]. Acetylation is also important to post-translational modification, this time involving the N^{ε} amino group of lysine and catalysed by lysine acetyltransferases (KATs), or sometimes occurring non-enzymatically [143,144]. The effect is neutralisation of the lysine's positive charge and increased sterical hindrance. In turn this influences catalytic activity, stability, subcellular localization and interactions with other molecules [143]. Finally, acetyl-CoA is the obligate cofactor for histone acetyltransferases (HATs), thereby participating in epigenetic regulation of gene expression. High levels of histone acetylation are required to support global transcription and to preferentially transactivate genes involved in cell growth and replication [114,145–147], glycolysis [148] and resistance to oxidative stress [149].

1.6.3 Targeting CoA metabolism for drug development

1.6.3.1 Anti-bacterial drug development targeting CoA metabolism

The multitude of metabolic roles of CoA, which are essential to many pathogenic parasites and other microbes, raises the possibility of it being a highly-impactful target for inhibition by antimicrobial drugs, and one that would likely be difficult to overcome by resistance mechanisms. Investigation of the pathway as a therapeutic target began in bacteria and has been extensively studied in these organisms.

In *E. coli*, the CoA pathway has been characterised by analysis of mutants derived by methods such as chemical selection [150] and random transposon mutagenesis [151]. It has been demonstrated that synthesis of pantothenate from α-ketoisovalerate and β-alanine is non-essential, due to acquisition of exogenous pantothenate (Fig. 1.6.3.1.1) [151,152]. Uptake is mediated by a sodium-dependent symporter, panF [152,153]. PanF shares homology with a mammalian pantothenate transporter [154]. *E. coli* cells lacking both panF and panC activities are non-viable [152]. Studies in *Mycobacterium tuberculosis* revealed that a strain carrying mutations in both panC and panD genes, which consequently lacked capacity for *de novo* pantothenate synthesis, was associated with attenuated infection in a tuberculosis mouse model [155]. This led to a search for pantothenate synthase (panC) inhibitors in this pathogen and vaccine development using strains defective in this pathway [155–157].

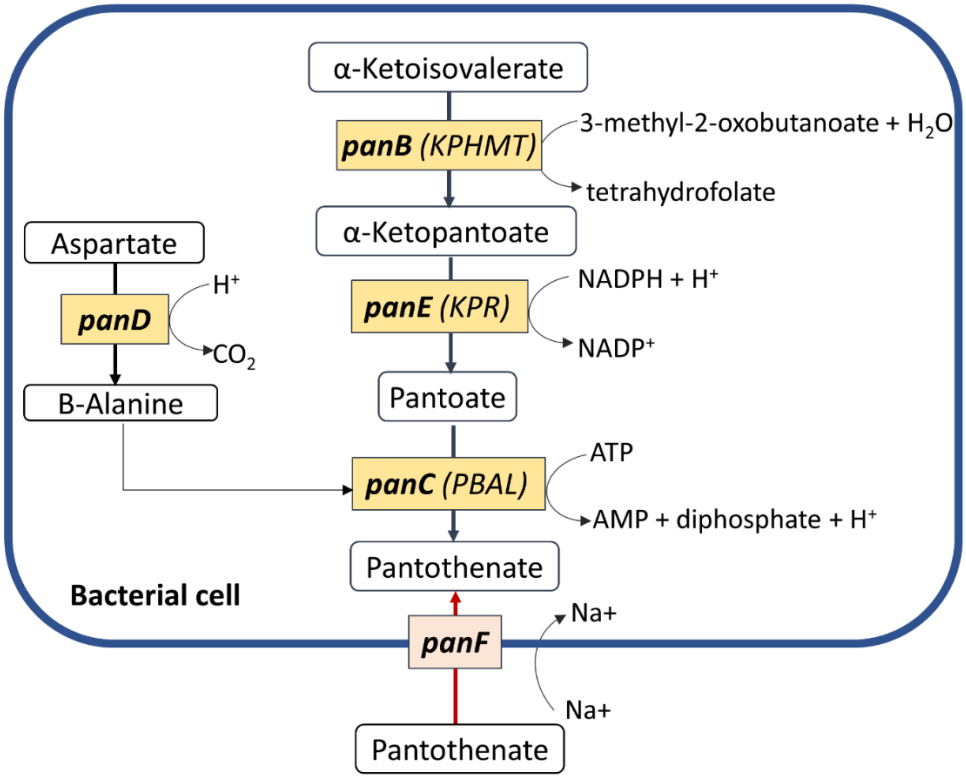


Fig. 1.6.3.1.1. Pantothenate synthesis in bacteria: *De novo pantothenate synthesis occurs from α-ketoisovalerate, catalysed by ketopantoate hydroxymethyl transferase (panB/KPHMT), ketopantoate reductase (panE/KPR), and pantoate-β-alanine ligase (panC/PBAL or pantothenate synthase). The final reaction uses β-alanine, acquired from L-aspartate by aspartate decarboxylase (panD). Pantothenate uptake by the Sodium-dependent transporter panF can bypass de novo synthesis.*

All enzymes required for converting pantothenate to CoA (coaA, coaBC, coaD, and coaE) are essential in *E. coli* [150,151,158], which is partly due to the fact that the phosphorylated intermediates are not membrane permeable (Fig. 1.6.3.1.2, refer to section 1.6.1) [151]. Other bacteria for which PanKs have been shown to be essential include *Salmonella typhimurium,* where strains carrying a mutation causing heat-induced PanK inactivation failed to grow at 40°C [159]. Similarly, *Bacillus anthracis* requires PanK activity for growth - null mutants could not be generated, but conditional, temperature-sensitive gene disruption resulted in growth arrest at the non-permissive temperature [160].

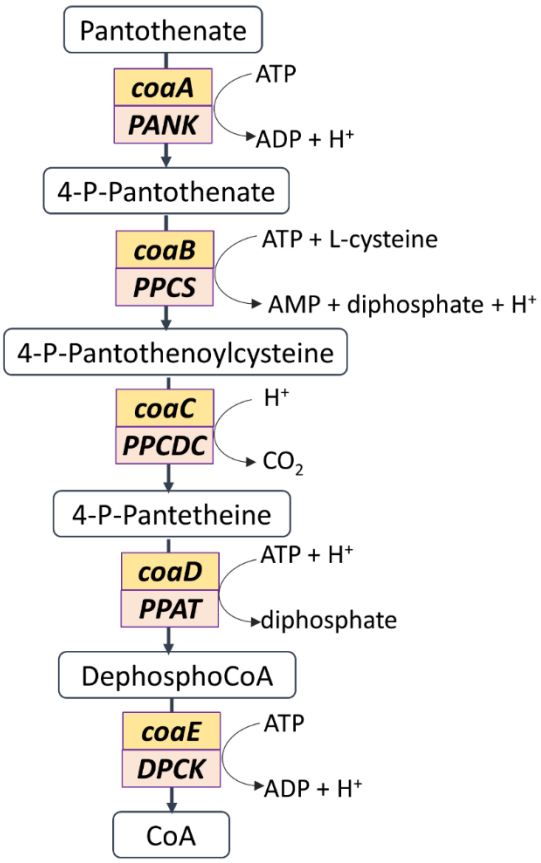


Fig. 1.6.3.1.2. CoA Biosynthesis: *To form CoA from pantothenate, pantothenate is phosphorylated to 4'-phosphopantothenate by pantothenate kinase (coaA/PANK), then condensed with cysteine by 4'-phosphopantethenoylcysteine synthase (coaB/PPCS), subsequent decarboxylation by 4'-phosphopantothenoylcysteine decarboxylase (coaC/PPCDC) forms 4'phosphopantetheine. 4'phosphopantetheine is AMPylated by phosphopantetheine adenylyltransferase (coaD/PPAT) and the resulting dephospho-CoA phosphorylated by dephospho-CoA kinase (coaE/DPCK). Bacterial enzyme nomenclature is shown in red and eukaryotic names in yellow.*

Consistent with genetic evidence for the importance of bacterial PanKs, several classes of pantothenate analogues have been found to possess anti-bacterial activity. N-substituted pantothenamides, which are structurally similar to pantothenate, are known to be active against *E. coli, Lactobacilli* and *Staphylococci* [161]*.* Some competitively inhibit PanKs and others act as antimetabolites that are converted to a non-usable form of CoA substitute that cannot form acyl-CoA esters [162–164].

Pantothenol and other pantothenyl alcohols are another class of CoA pathway inhibitors which were demonstrated to inhibit growth of the lactic acid bacteria *Leuconostoc mesenteroides* and *Lactobacillus arabinosus* [165,166] . Similarly, growth of several strains of *Staphylococci* was suppressed by more than 70% [167]. It was later shown in studies of *M. tuberculosis* and *E. coli* that bacteriostatic or bactericidal effects of pantothenols are exerted by inhibition of CoaBC, i.e. synthesis of 4'-phosphopantetheine from 4'-phosphopantothenate (Fig. 1.6.3.1.2). PanK converts the pantothenol into a 4'-phosphopantothenol which is unable to be processed by

downstream CoaBC and competitively inhibits its activity [168]. Since the inhibitory effects of pantothenol can be rescued by provision of pantothenate, its therapeutic potential is unfortunately limited by the potential for scavenging of host pantothenate [166,167].

Inhibitors of CoaD (or phosphopantotheine adenyltransferase, PPAT) have also been investigated as antibacterial agents. Cyclo-alkyl-pyrimidines designed to target *S. aureus* PPAT, were effective *in vitro* and in animal infection models, however their effect on bacterial growth was delayed and toxic side-effects were observed, therefore they were not considered for clinical development [169]. Triazolopyrimidinone and azabenzimidazole displayed nanomolar IC50 values but were not sufficiently effective *in vitro* against *E. coli*. Furthermore, L-ascorbic acid (Vitamin C) and food additive tri-sodium citrate were reported to inhibit bacterial PPATs and possess bactericidal activity, based on *in vitro* growth of *Acinetobacter baumannii*, *S. aureus,* and *Klebsiella pneumonia,* but are yet to be tested *in vivo* [170]. CoA analogues have also had limited success as anti-bacterial compounds, which is attributed to limited permeability [171].

1.6.3.2 Anti-fungal drugs targeting CoA synthesis

Like bacteria, some species of fungi can synthesise pantothenic acid from β-alanine. Homologues of the *E. coli* genes coding for panB, panE and panC (Fig. 1.6.3.1.1) have been identified in the genomes of *Schizosaccharomyces pombe. Saccharomyces cerevisiae* and *Neurospora crassa* [172] and activities of some of these enzymes have been confirmed in *S. cerevisiae* and *Aspergillus nidulans* [173,174]. Unlike bacteria, the precursor β-alanine is synthesised from spermine or sometimes uracil, rather than from L-aspartate [175–178]. Despite possessing a *de novo* synthesis pathway, uptake of exogenous pantothenate occurs in *S. cerevisiae* via Fen2p, a high-affinity proton-dependent pantothenate transporter [179] and via the transporter liz1 in *S. pombe* [178,180]. Mutation of Fen2p has been associated with resistance to fenpropimorph, an anti-fungal compound [181] and mutation in liz1 was shown to impede fatty acid synthesis [178,180]. The transporter shares limited homology with pantothenate transporters from *E. coli* and mammals [153,154,182]. In *S. cerevisiae* exogenous β-alanine can also be acquired using the amino acid permease Gap1p [179].

S. cerevisiae possesses all enzymes required for synthesis of CoA from pantothenate and a large-scale deletion study demonstrated that all are essential for viability [183]. Crystal structures of these fungal enzymes are lacking, which has impeded drug targeting of the pathway. Pantothenate analogues which have been documented to inhibit growth of *S. cerevisiae* include pantoyltaurine, a sulphonic acid analogue of pantothenic acid and keto- analogues

of pantothenic acid: D-methylpantothenone and D-phenylpantothenone, the latter being the most potent inhibitor [184].

1.6.3.3 CoA synthesis in protozoa

Most research on the CoA pathway in protozoa has focused on apicomplexan pathogens or related organisms, particularly *P. falciparum* and other clinically important malaria-causing species, and *Toxoplasma gondii*, which causes toxoplasmosis. CoA synthesis from pantothenate is typically a conserved and essential pathway in protozoa and has been considered as a target for anti-parasitic chemotherapy [172,185–187].

Pantothenic acid was discovered to be important to the growth of erythrocytic stages of *Plasmodium* during *in vitro* studies on the avian malaria parasite, *Plasmodium lophurae,* where Trager showed that parasite survival was enhanced by addition of calcium pantothenate to the culture medium [188]. Subsequently other strains including *Plasmodium gallinaceum* and *P. falciparum* were found to require pantothenate for intraerythrocytic growth *in vitro* [187,189]. In fact it is the only water-soluble vitamin for which intracellular *P. falciparum* stages have a definitive requirement [187]. Uninfected erythrocytes appear to lack a pantothenate transporter and take up low amounts of the vitamin under normal conditions [190]. However, when infected with *P. falciparum,* pantothenate uptake occurs readily as a result of 'new permeability pathways' (NPP) in the erythrocyte membrane, which are induced by the developing parasite [190]. Within the infected erythrocyte, pantothenate is thought to diffuse across the parasitophorous vacuole membrane [191], which encapsulates the intracellular parasite and is actively transported into its cytosol via a plasma membrane H^+-coupled transporter. A candidate transporter, *Pf*PAT, was identified in *P. falciparum*, and shown to be essential for intraerythrocytic growth of this species [192,193]. It is a low-affinity transporter, unrelated biochemically to the mammalian and bacterial sodium-coupled pantothenate symporter [193]. A PAT homologue was later identified in the murine parasite *Plasmodium yoelli* [194]. Blood-stage parasites of this species did not require the pantothenate transporter for development in mouse erythrocytes, however it was required for sporozoite and oocyst formation and hence for transmission of to the mosquito.

P. falciparum possesses all genes encoding biosynthesis of CoA from pantothenate (see Fig. 1.6.3.1.2). PanK activity has been biochemically characterised in *P. falciparum* lysates and by recombinant bacterial expression [190,193,195]. Furthermore, inhibitors of *Pf*PanK were active against *P. falciparum* erythrocytic stages *in vitro*, demonstrating the importance of the pathway [195]. PanK expression has also been confirmed in asexual blood stages of the murine parasite

Plasmodium yoelii [196]. Consistent with pantothenate transport studies of this species, transgenic *P. yoelii* strains lacking PanK activity were able to undergo normal asexual development and sexual differentiation in mouse erythrocytes. However, these knockout parasites were severely deficient in ookinete development and failed to produce sporozoites in Anopheles mosquitoes [196].

The downstream predicted activities PPAT and DPCK are encoded as separate proteins. The putative *P. falciparum* DPCK localises to the apicoplast [197], a plastid-type organelle which is the site of fatty acid synthesis in these organisms [198]. This places CoA biosynthesis in proximity to ACP and other important CoA-utilizing enzymes. In humans, fatty acid synthesis occurs in the cytoplasm and the bifunctional PPAT/DPCK enzyme localizes to the outer mitochondrial membrane [199], placing CoA synthesis in proximity to fatty acid synthesis, the TCA cycle and other mitochondrial CoA-utilising pathways. Interestingly, not all malaria parasites possess these enzymes. Intracellular blood-stages of *P. lophurae* scavenge CoA from duck erythrocytes and lack the required enzymes to synthesise CoA from pantothenate [200]. Consistent with this, a pantothenic acid analogue pantoyltauramido-4- chlorobenzene had no effect on the growth of extracellular parasite stages [201,202].

A vast number of pantothenate analogues have been investigated as anti-malarial compounds. *In vitro* and *in vivo* experiments using avian and mammalian malaria species have identified several classes of analogues, including pantoyltauramides and phenylpantothenones, that outperformed quinine with regards to potency, being effective at micromolar concentrations, with some also displaying superior tolerability [184,201,203–205]. For example, methoxyquinolyl-substituted pantoyltauramide suppressed parasitaemia in *Plasmodium cynomolgi*-infected monkeys when orally administered (10 mg/kg) for 7 days, eliciting a cure in one monkey [204]. Pantothenol is another inhibitory analogue with high tolerability *in vitro* and *in vivo, based on P. falciparum* studies [187]. However, its activity is antagonized by pantothenic acid as the two compounds compete for the active site of the parasite's PanK. Its potency against *P. vinckei vinckei* - infected mice was limited [187], probably due to its oxidation to pantothenic acid, catalysed by mammalian alcohol dehydrogenase [165]. Pantothenamides, which contain an amide in place of the carboxyl group of pantothenate, act as anti-metabolites of the CoA pathway, being processed by PanK and downstream enzymes into an analogue of CoA that fails to convert to acetyl-CoA [206]. Their susceptibility to hydrolysis by serum pantetheinases [207] requires that they are structurally optimised to resist degradation [208–210]. One such compound, MMV693183, displayed nanomolar potency against blood stages of *P. falciparum* and *P. vivax*, and against *P. falciparum* gametocytes [211].

Another clinically important apicomplexan parasite is the coccidian parasite *T. gondii.* Coccidia are able to synthesise pantothenate *de novo* from α-ketopantoate using a bifunctional ketopantoate hydroxymethyl transferase- ketopantoate reductase (KPHMT-KPR), and pantoate-β-alanine ligase (PBAL) (see Fig. 1.6.3.1.1 for bacterial homologues) [197]. In *T. gondii,* the former is located in the mitochondrion and the latter in the nucleus [212]. Targeted knock-out of these enzymes indicated that they are dispensable for *in vitro* proliferation of *T. gondii* tachyzoites, due to effective pantothenate salvage mechanisms [212]. However, both enzymes were found to be important to virulence during chronic infection. Bradyzoites with targeted disruption of either KPHMT-KPR or PBAL genes produced fewer cysts during chronic mouse infection.

CoA synthesis from pantothenate was found to be essential in *T. gondii* [212]. The pathway appears to be cytoplasmic, with the final enzyme, DPCK, being membrane-bound within the cytoplasm, but unlike *Plasmodium* is not associated with the apicoplast. The importance of confirmed activities of PanK, PPCDC and DPCK in *T. gondii* was demonstrated by downregulation of each activity. Depletion of each enzyme clearly altered cell morphology and prevented formation of lytic plaques in human cell monolayers [212].

The importance of pantothenate synthesis in chronic infection suggests that KPHMT-KPR and PBAL could be plausible targets for intervention against chronic stages of *T. gondii.* Several inhibitors of *the M. tuberculosis* PBAL were shown to inhibit growth of *T. gondii* tachyzoites, some with nanomolar IC50 values [213]. *T. gondii* PBAL was not confirmed as the target of these compounds, however their inhibitory effects were rescued with pantothenate indicating that the CoA pathway was targeted.

There have been few studies on CoA synthesis in trypanosomes. Analyses of nutritional requirements of cultivated trypanosomatids indicated that pantothenic acid is required by several species [214–216]. However, the vitamin was not required for growth of symbiont-harbouring trypanosomatids such as the insect parasites *Strigomonas oncopelti* and *Angomonas deanei* [216,217]*.* This is due to the fact that between their resident endosymbiotic bacteria and their own metabolic machinery they possess the enzymes required for pantothenate synthesis [218]. Genomic analysis of trypanosomatids revealed that symbiont-harbouring parasites have a KPR gene and their endosymbionts possess the missing KPHMT and PBAL. Genes encoding the five enzymes required to convert pantothenate to CoA appear to be present in trypanosomatid genomes, but have yet to be characterised [218]. A more recent study investigated the effect of CoA pathway inhibitors on trypanosomes. *In vitro* screening assays indicated that *T. brucei* and *T. cruzi* trypomastigotes were susceptible to growth

inhibition by three cyclohexane- based compounds which displayed IC_{50} values in the nanomolar range: these were STK 740987 (a cyclohexane sulfonate), Amb3377585 (3-acetyl-6H-anthra[1,9-cd]isoxazol-6-one) and Amb4317088 (2-acetyl-3-phenylnaphthoquinone) [219]. CoA and dephospho-CoA supplementation rescued inhibition of *T. brucei* in this experiment, suggesting that the compounds either inhibited PPAT or DPCK. The feasibility of CoA metabolism, more broadly, as an anti-trypanocidal target is supported by evidence for its targeting by the anti-leishmanial drug miltefosine. This compound, an alkylphosphocholine, has been shown to inhibit the glycosomal enzyme alkyl-specific acyl-CoA acyltransferase, which likely contributes to disruption of fatty acid and sterol metabolism [220,221]. However, other cellular targets surely contribute to the cytotoxicity of miltefosine, which is thought to involve apoptosis and disruption of lipid-dependent signalling pathways [221,222].

1.6.4 Characteristics of pantothenate kinases

The phosphorylation of pantothenate, catalysed by pantothenate kinase (PanK) is the rate-determining step of CoA biosynthesis in many organisms [223–226] and for this reason is of particular interest as a drug target. Consistent with this, several experiments in bacteria have demonstrated that PanK expression and pantothenate availability are the key determinants of cellular CoA levels [227–230]. Three types of PanK have been described (type I, II and III). Most bacteria possess PanK-I which is also referred to as CoaA. The first to be characterised was *E. coli Ec*CoaA [150]. PanK-II is mainly found in eukaryotes, but has also been identified in a small subset of bacteria, in which it is named CoaW. The genera in which PanK-II predominates are *Staphylococci* and *Bacillus*, which typically possess both PanK-I and PanK-II enzymes [158]. PanK-III, also known as CoaX, was more recently discovered in bacteria, being first described in *Helicobacter pylori and Bacillus subtilis* [231] and subsequently identified in thousands of bacterial species. All three types have been shown to be essential, except where multiple isomers are encoded in the same organism, in which case functional redundancy tends to occur. For example *M. tuberculosis* contains an active Pank-I and an inactive PanK-III [232] . In contrast, *Bacillus anthracis* contains an active PanK-III and an inactive Pank-II [160]. In both cases the active protein is essential to growth. Eukaryotes may express several isoforms of PanK-II. This is the case for mammals, which possess four genes, PANK1-4. While human PANK4 encodes a non-functional PanK [233], the other three isoforms are active, metabolically important enzymes. Mutation in the mitochondrial isoform PanK2 causes pantothenate kinase-associated neurodegeneration (PKAN) - an often fatal disease characterised by dystonia and spasticity [234].

All three types of PanK have been shown to act as dimers [228,235–237]. The dimerization interface is formed by a long alpha helix at the C-terminal end, which stabilises the complex by van der Waals forces and hydrogen bonds. Another common structural element in all PanKs is the P-loop, a glycine-rich loop which performs ATP binding [113]. The monomer of PanK-II is considered as two domains: A and B, where the A domain performs substrate binding and the B domain forms the dimerization interface [236]. In domain B, the long alpha helix forms a hydrophobic core with several smaller helices to provide further stability. Bacterial and human PanKs were shown to assemble as homodimers, however the apicomplexan parasites *P. falciparum* and *T. gondii* use heteromeric PanKs composed of two separately encoded isoforms [238]. One of the oligomers of *P. falciparum* (*Pf*PanK2) does not possess PanK activity, therefore may only have structural importance, but it is an essential component of the heterodimer complex [195,238]. The oligomers of *T. gondii* PanKs (*Tg*PanK1 and 2) are both active and both individually essential, i.e. neither isoform is active without the other [212,238].

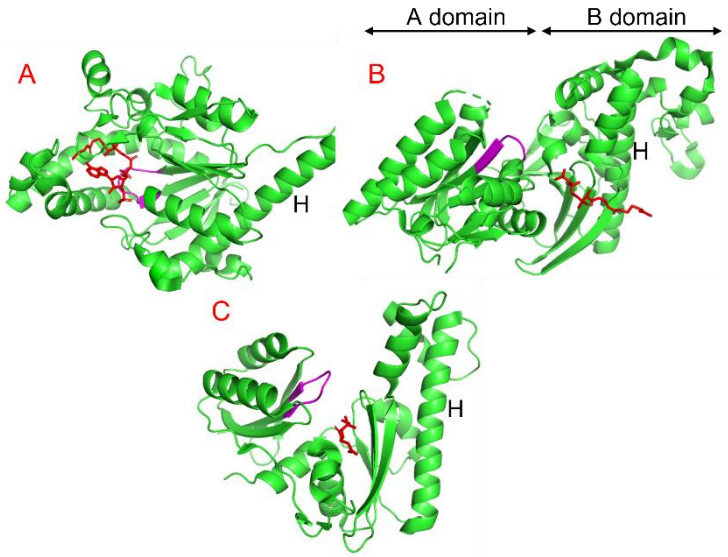


Fig. 1.6.4 Structures of PanK types I-III*: (**A**) E. coli PanK-I (CoaA) complexed with Coenzyme A (in red) (1ESM), (**B**) Human PanK-II (PanK3 isoform) with acetyl-CoA (217P) (**C**) Pseudomonas aeruginosa PanK-III (CoaX) with pantothenate (2F9W). ATP-binding P-loop motifs are indicated in magenta and the dimer-interface-forming central helix (H) is labelled.*

Consistent with PanK being the rate-determining enzyme of CoA biosynthesis, it is typically regulated by feedback inhibition by CoA and/or its thioesters. Most Type I and II PanKs are inhibited by CoA, acetyl-CoA and malonyl-CoA by their association with the ATP binding site [113]. The mammalian pantothenate kinases are known to exist in two distinct conformational states: an open conformation that is catalytically inactive and is stabilized by the binding of acetyl-CoA, and a closed, active conformation, which is stabilized by ATP·Mg^{2+} [239]. A rotational hinge motion of the helix α4 mediates the switch between the two states. It places a catalytic

glutamate residue (Glu-138 in PanK3), which is situated at the end of the rotating helix, outside of the catalytic pocket in the inactive conformation, or into the active site in the active conformation. PanK-IIIs on the other hand, seem not to be regulated by CoA or its derivatives [237]. The need for tight regulation of the first step of the CoA pathway in most organisms highlights the suitability of PanK as a target for chemotherapy.

1.7 Challenges of genetic characterisation of *Trypanosoma cruzi*

The development of new treatments to combat Chagas disease has been restricted by limited means to fully understand the biology of the parasite and the pathogenic mechanisms involved in disease. Crucially, scientists have been lacking flexible and efficient tools to undertake genetic manipulation and thereby derive information on gene function.

T. cruzi has a diploid genome, which displays considerable variation across the species. For example, strain CL Brener has 32.5 mega base-pairs per haploid genome and strain TCC has as many as 87 mega base pairs (tritrypdb.org).The nuclear genome is organised into approximately 41 chromosomes [9,240,241], which encode ~11, 000 genes (tritrypdb.org). This large genome size can be attributed to chromosomal polymorphisms and expanded surface antigen gene families [7,241]. Approximately 50% of the genome is composed of repetitive sequences – mainly multigene families and retrotransposons [7]. Duplicated gene families can consist of more than 2000 copies. The largest of these are the trans-sialidases and mucin-family proteins, which are surface-proteins shown to have roles in immune evasion and physical interactions used for host cell invasion [7]. Their importance to survival and life-cycle progression may have resulted in selection for their genetic expansion.

Gene expression is also unusual in kinetoplastids. Rather than individual protein-coding genes being regulated by individual promoters, they are instead arranged in long polycistronic clusters of up to 100 genes [11] which are constitutively transcribed. Classical RNA polymerase II (pol II) promoter sites [242] and transcription factors are lacking in trypanosomatids [243], and although pol II drives transcription of most genes, transcription mediated by RNA polymerase I (pol I) has been described for variant surface protein and procyclin genes of *T. brucei* [244]. Individual mRNAs for protein-coding genes are processed through trans-splicing, involving the addition of a short spliced-leader RNA sequence to the 5'-end, which occurs alongside polyadenylation reactions [245]. Post-transcriptional regulation of mature mRNAs then determines the levels of their protein product according to cellular demand [242]. Regulatory *cis*-element sequences, mainly located in the 3' UTRs, act as protein-binding sites [246], an example being the AU-rich elements (ARE) [247]. *Trans*-acting RNA binding proteins also control gene

expression [248]; these include poly(A)-binding protein [249], *Tc*UPB1 [250] and zinc finger proteins [251].

Compared to its close relative *T. brucei*, genetic manipulation of *T. cruzi* has made few advances, being hindered by two major issues. Firstly, it takes much longer to generate clonal populations from transfection of *T. cruzi* epimastigotes (4-6 weeks), compared with transfection of *T. brucei* bloodstream forms (~6 days). Secondly, *T. cruzi* lacks the ribonuclease components of the RNA interference (RNAi) pathway [252]. This is a highly effective gene silencing system, triggered by expression or introduction of double-stranded RNA (dsRNA). A Dicer endonuclease cleaves the dsRNA into small interfering (si) RNA (20–30 nucleotides). These siRNAs are then unwound and one strand (the guide strand) incorporated into a protein complex known as the RNA-induced silencing complex (RISC). The RISC searches the transcriptome for messenger RNA (mRNA) sequences that form complementary base pairs to the guide strand. Once identified, target mRNAs are cleaved by a RISC component endonuclease known as argonaute [253].

As a research tool, RNAi-mediated depletion of target mRNAs can be initiated by transfection of compatible cells with dsRNA homologous to the target sequence. *T. brucei,* possesses the classical RNAi proteins *Tb*AGO1, a member of the Argonaute ribonuclease family and two Dicer-like homologues, *Tb*DCL1 and *Tb*DCL2 [252]. Furthermore, an inducible RNAi system, based on tetracycline regulation has been developed for this parasite to facilitate the study of essential genes [254].

Until recently, modification of genes in *T. cruzi* has relied on the ubiquitous mechanism of homologous recombination, which in comparison is laborious and inefficient. Homologous recombination is a natural DNA repair mechanism, which maintains genome stability and produces genetic diversity during meiosis [255]. Amongst other types of DNA damage, it can repair double-stranded DNA breaks (DSB) by using an undamaged chromosome as a template. Fig 1.7.1 shows how this occurs. To introduce specific mutations with this method, a plasmid containing the desired DNA insert can be integrated if it is homologous with the target locus and can therefore act as a repair template. To produce the DNA insert requires cloning in *E. coli* followed by two rounds of transfection to produce null *T. cruzi* mutants, which can take several months overall.

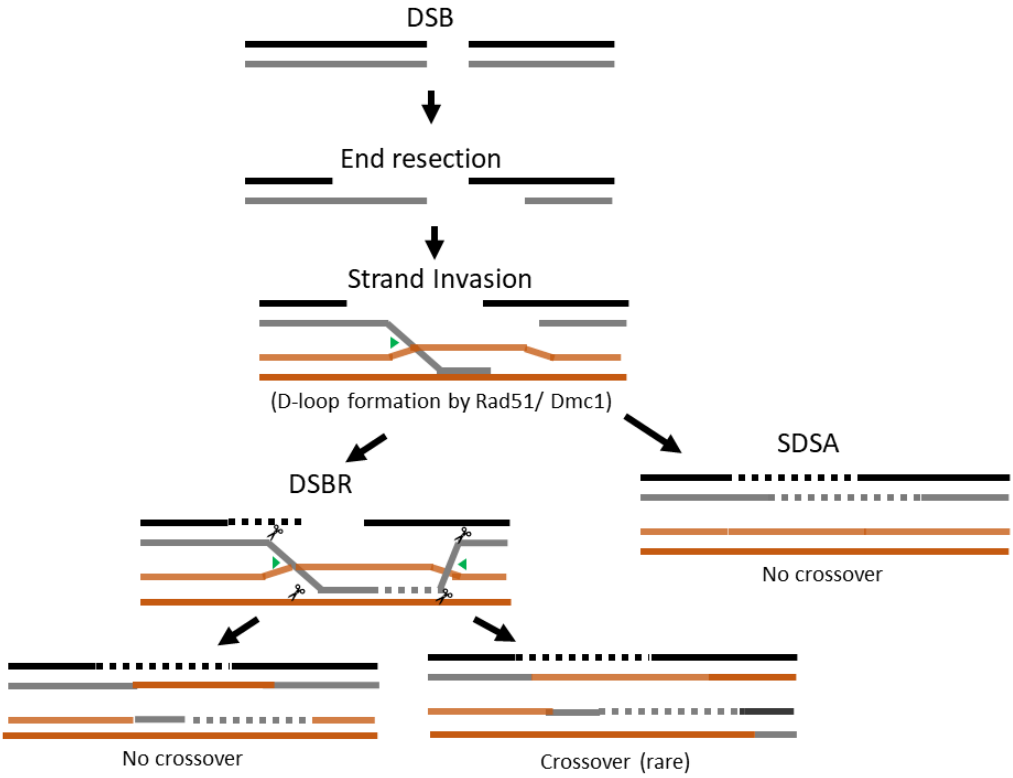


Fig. 1.7.1. Homologous recombination: *Repair of a double stranded break (DSB) begins by resection of the ends, forming 3' single-stranded DNA (ssDNA) overhangs. One of these ssDNA overhangs invades a complementary repair sequence (i.e a sister chromatid), forming a displacement loop (D-loop), which requires a recombinase (Rad51 and/or Dmc1 in most eukaryotes). DNA lost at the DSB site is replaced by a DNA polymerase. The remaining overhang may also invade and undergo DNA extension in what is called the DSB repair (DSBR) pathway. The strands are cleaved at junction sites (green arrows) by nicking endonucleases (represented by scissors) and recombined. Alternatively, synthesis-dependent strand annealing (SDSA) can occur, whereby the extended overhang is released and anneals to the other 3' overhang in the damaged chromosome through complementary base pairing and no crossover products are formed.*

Recent advances in the genetic manipulation of *T. cruzi* have allowed researchers to overcome some of these barriers. To replace sequential homologous recombination, parasite lines are now available which are compatible with CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 (CRISPR-associated protein 9) [256]. CRISPR-Cas9 is a bacterial gene-editing tool, which evolved to provide adaptive immunity against viruses. Cas9 is an endonuclease which when complexed with a single guide RNA (sgRNA) creates a DSB in a target sequence. The sgRNA contains two parts: a specific CRISPR RNA (crRNA) of approximately 20 nucleotides and a *trans*-activating RNA (tracr RNA). The crRNA can be synthesised to target a gene of interest; this should be complementary to a region of the target gene, referred to as the protospacer sequence, so that it can be located by base-pairing. The tracr RNA acts as a binding scaffold for the Cas9 nuclease. The specificity of cutting is further increased by the requirement for Cas nuclease to recognise a short but specific sequence of nucleotides at the target site, called the protospacer adjacent motif (PAM). The nuclease cuts a few nucleotides upstream of the PAM. The breaks are subsequently mended by homology-

directed repair, at which point insertions can be made at the target locus. To improve the chances of a DNA insert being incorporated into the desired locus, it should contain flanking sequences of at least ~30 base-pairs that correspond to the target site, which are referred to as homology arms. With this technology, it is simpler and faster to generate null mutants and parasites with modified proteins. The required targeting constructs can be generated by PCR, removing the need for cloning in *E. coli,* saving additional time.

1.8 Aims of the PhD Project

There is an evident need for new treatments to combat the widespread impact of kinetoplastid diseases. Drug targets can be components of essential metabolic pathways, such as vitamin salvage. One pathway that has demonstrated essentiality for growth and virulence of apicomplexan, fungal and bacterial pathogens is CoA biosynthesis from the vitamin pantothenate. This project will explore CoA biosynthesis in kinetoplastids, with a focus on the pathogens *T. cruzi* and *T. brucei.*

Broad project aim:

- Perform the first genetic and biochemical characterisation of kinetoplastid PanKs

Specific aims:

- Using bioinformatics, identify CoA biosynthesis enzymes in *T. cruzi* and other clinically important kinetoplastids
- Use CRISPR-spCas9 to tag the identified *T. cruzi* PanK alleles and determine subcellular location
- Generate trypanosome strains deficient in PanK by targeted gene replacement and inducible knockdown using RNA interference
- Use the engineered strains to determine the importance of PanK to *in vitro* growth and viability of the parasite
- Confirm activity of the identified PanK using biochemical assays

2. Materials and Methods

2.1 Bioinformatic analysis

For identification of orthologues of CoA biosynthetic enzymes in trypanosomatid genomes, the databases at TriTrypDB (http://tritrypdb.org) [257] were searched using the Basic Local Alignment and Search Tool (BLAST). Query sequences for homology searches (characterised enzymes from humans, bacteria, fungi and other specified organisms) were obtained from UniProt (https://www.uniprot.org) [258] and are identified by their UniProtKB code. Target organisms of interest for CoA pathway homologue searches (Chapter 3.2) were *Trypanosoma cruzi (T. cruzi), Trypanosoma brucei* and *Leishmania major*. The Expectation value was adjusted to 1 in these searches to ignore hits of low significance. Hits from each search belonging to the genome of *T. cruzi* strain CL Brener, were then aligned to orthologues from specified organisms in Jalview (https://www.jalview.org) [259], using the ClustalO program to visualise sequence-wide conservation. Protein structure models were downloaded from AlphaFold (https://alphafold.com) [260,261] and aligned using pymol (https://pymol.org/2) [262].

Pantothenate kinase sequence analysis was performed using *T. cruzi* CL Brener sequence: NCBI gene ID: Tc00.1047053511153.120 (non-Esmeraldo-like allele). The NCBI Blastp tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [263] was used to infer homology from amino acid sequences using the standard NCBI databases. The conserved domains (CD) search within the database CDD v3.19-58235 PSSMs was also used for domain analysis (https://www.ncbi.nlm.nih.gov/cdd) [264] . To assess conservation of PanK domains, multiple-sequence alignments were performed using Clustal Omega with default settings applied, or Jalview using default Clustal alignment settings and Clustalx or percentage identity colour assignment. The generated alignments were manually annotated to show metal-interacting or other substrate-binding residues, based on published literature [265–267]. *Tc*PanK structure models and associated data were downloaded from AlphaFold and can be identified by UniProt codes Q4DHB2 and Q4DME2. To clearly delineate predicted domains, colouring and surface representation was applied in Pymol. To predict domain functions from structure, the AlphaFold model was submitted to the CO-FACTOR and COACH servers [268,269]. This derived similar protein structures and predicted ligands and binding sites. The most similar proteins were structurally aligned with their respective *Tc*PanK domains in Pymol.

For gene tagging experiments, PSORT (https://wolfpsort.hgc.jp) [270] was used to predict localisation signals in *T. cruzi* PanK (*Tc*PanK) sequences TcCLB.506779.80 and TcCLB.511153.120.

2.2 Cell culture

2.2.1 *Trypanosoma cruzi* culture

The *T. cruzi* strain (Cl Brener) used for CRISPR experiments has been engineered to constitutively express T7 RNA polymerase and spCas9 nuclease [256]. Epimastigotes were grown at 27°C in RPMI-1640 medium (Gibco) supplemented with 10% (v/v) heat-inactivated foetal calf serum, 0.5% (w/v) trypticase, 0.03 M haemin, 2.5 U/ml penicillin, 2.5 µg/ml streptomycin, 2 mM sodium glutamate and 2 mM sodium pyruvate. Episomal spCas9 expression was maintained under G418- selection (100 µg/ml). Exponential phase parasites were cryopreserved in liquid nitrogen in 5% DMSO (v/v) and 20% FBS in culture medium.

2.2.2 *Trypanosoma brucei* culture

For RNA interference experiments, the *T. brucei* strain 2T1 (427/MITat1.2) was used, which constitutively expresses the tetracycline-responsive repressor protein and a partial Hygromycin-resistance marker (Chapter 4, Fig. 4.4.1) [254,271]. Bloodstream-form trypomastigotes were grown at 37°C under a 5% CO_2 atmosphere in HMI-9 medium [272]. They were maintained under selection with puromycin (0.2 µg/ml) and phleomycin (2 µg/ml). Exponentially growing parasites were cryopreserved in liquid nitrogen in 10% glycerol (v/v) and 20% FBS (v/v) in culture medium.

2.2.3 Bacterial strains and cultures

DNA cloning was performed using *E. coli* DH5α, which carries inactivating mutations in recombinase (*rec*A1) and endonuclease (*end*A1) genes, promoting stability of plasmid inserts and high DNA yields. Recombinant protein expression was performed in *E. coli* BL21 (DE3), which is deficient in Lon and OmpT proteases, promoting high protein yields and expresses T7 RNA polymerase for specific expression of genes under the control of T7 promoters.

E. coli were grown at 37°C in NZCYM medium with agitation, or on solid NZCYM/agar plates, with Ampicillin (100 µg/ml). Bacterial stocks were prepared in NZCYM containing 20% glycerol (v/v) at -80°C.

2.2.4 Mammalian strains and cultures

HEK293T cells used for recombinant protein expression were grown in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) containing 5% FBS (Sigma), at 37°C in 5% CO_2. They were cryopreserved in liquid nitrogen in DMEM containing 20% DMSO and 10% FBS.

2.3 Preparation of DNA constructs

2.3.1 Phusion master mix PCR conditions

For all PCRs used in cloning and construct generation, the high-fidelity Phusion master mix (NEB) was used. 200 nM primers were combined with 10-30 ng cloned DNA or 50-200 ng genomic DNA template, 1x master mix and adjusted to 50 µl with distilled water. Unless otherwise specified, PCR cycling conditions were as follows: Initial denaturation of 30 seconds at 98°C, followed by 30 cycles of denaturation (10 seconds at 98°C), annealing (30 seconds at the lowest primer T_m + 3°C) and extension (20-30 seconds/kb at 72°C), and then a final extension of 5 minutes at 72°C.

2.3.2 Ligation and transformation conditions

15 ng of linear vector was combined with 10-50 ng insert according to a 1:3 molar ratio and incubated at room temperature with 5 U T4 DNA ligase and 1x rapid ligation buffer (ThermoFisher) for up to 40 minutes. 10 µl ligation reaction was combined with 100 µl of chemically competent *E. coli*, cooled on ice for 10 minutes, heated for 2 minutes at 42°C, cooled on ice for another 15 minutes and then streaked onto an NZCYM/agar plate using a sterile spreader.

2.3.3 Cloning of pPOT-PTP for CRISPR tagging

The plasmid pPOT-PTP (Appendix SF2) was constructed by amplification of the PTP tag from pCptpNEO (10 ng) (Appendix SF1 [273]) using 200 nM primers which incorporated Hind-III (5') and Bsm-I (3') restriction sites. The 500 bp product consisting of the PTP tag was digested with Hind-III, Bsm-I and Dpn-I and column-purified using Qiaquick PCR purification kit (Qiagen). pPOTv6 puro-puro mNG was also digested with Hind-III and Bsm-I to release 600 bp of the mNeonGreen tag. After gel purification, the 6.5 kbp linear plasmid was ligated with the digested PTP fragment. Sequencing confirmed successful ligation of the pPOT plasmid with an intact PTP ORF. Primer sequences can be found in Appendix Table A.1.

2.3.4 CRISPR construct preparation by PCR

30 nt primer annealing sites were designed to identical or near-identical regions (up to 2 mismatches) of the aligned sequences. To design the single guide RNAs (sgRNAs), the online tool EuPaGDT (http://grna.ctegd.uga.edu) [274] was used to locate a protospacer sequence within ~300 nucleotides of the desired integration sites, which also lacked secondary targets in the genome.

To amplify homology donor constructs, 30 ng pPOT plasmid DNA (pPOTv6 puro-puro mNG or pPOTv7 blast-blast mScarlet, a gift from Samuel Dean [256,275]) was amplified using 2 µM each of gene-specific forward and reverse primers in a 40 µl reaction containing Phusion master mix (ThermoScientific) and 3% DMSO (For plasmid maps: SF3, SF4). The pair of primers 'N-F Tag' and 'N-R Tag' amplified ORFs of puromycin N-acetyl-transferase and mNeonGreen and intergenic Ty repeats from pPOTv6 puro-puro mNG, forming the N-terminal tagging construct (Chapter 4, Fig. 4.2.1). Two C-terminal tagging constructs were synthesised using primers 'C-F Tag' and 'C-R Tag'. Amplification using pPOTv7 blast-blast mScarlet produced a construct containing Scarlet fluorescent protein and blasticidin deaminase, whereas the pPOT-PTP-amplified construct contained Ty repeats, a PTP tag and puromycin deaminase. All tagging constructs also contained the required RNA processing signals and intervening linker peptides. Knockout or gene-replacement constructs were synthesised using primers 'F KO' and 'R KO' which targeted the 5' UTR and a 3' PanK fragment (>154 nt upstream of the STOP codon) respectively. One construct contained puromycin N-acetyl-transferase, amplified from pPOTv6 puro-puro mNG and the other contained blasticidin deaminase, amplified from pPOTv7 blast-blast mScarlet. Both constructs also contained the necessary RNA processing signals. PCR cycling conditions for all CRISPR constructs were as follows: 5 minutes at 98°C, followed by 40 cycles of denaturation (30 seconds at 98°C), annealing (30 seconds at 65°C) and extension (2 minutes and 15 seconds at 72°C), and then a final elongation of 7 minutes at 72°C.

sgRNA templates were amplified using a forward primer containing a gene-specific 20 nt sequence and a reverse primer containing the sgRNA scaffold (named 'R-gRNA scaffold') [276]. Forward primers 'N F gRNA Tag' and 'C-F gRNA Tag' were used to make one sgRNA for use in N- and C-terminal tagging transfections respectively. Two sgRNAs were synthesised for each knockout transfection, made using 'N-gRNA KO' or 'C-gRNA KO' forward primers. 2 µM of each primer was combined in a 20 µl reaction with Phusion™ master mix and deionised water. Incubation conditions were as follows: 30 seconds at 98°C, followed by 35 cycles of denaturation (10 seconds at 98°C), annealing (30 seconds at 60°C) and extension (15 seconds at 72°C).

All PCR products were visualised on agarose gels and purified using the Qiaquick PCR purification kit (Qiagen) using Milli-Q water for elution. A list of primers used is shown in Appendix Table A.2.

2.3.5 Preparation of RNAi constructs

To make the *Tb*PanK mRNA depletion construct, two complementary RNAi-trigger fragments corresponding to nucleotides 1028-1436 of *Tb*PanK were cloned in opposite orientations into plasmid pRPai [271] (SF5) in a 2-step cloning process. Step 1: *Tb*PanK fragment amplification from *T. brucei* (421) gDNA using primers (RNAi F and R) containing two restriction sites each in 4 x 50 µl PCR reactions. For cloning in a sense orientation into MCS1, half of the PCR product was digested with Kpn-I and BamH-I and purified, then ligated with $pRPa^{SLi}$ (cut with the same restriction enzymes and gel purified) (Chapter 4, Fig. 4.4.2). To insert the antisense fragment into MCS2, the second half of the PCR product was digested with Apa-I and Xba-I and cloned into $pRPa^{SLi}$ containing the sense insert (digested with Apa-I and Xba-I and gel purified). The resulting construct was linearised with Asc-I for transfection.

The WT *Tc*PanK complementation vector was constructed by amplification of the complete 4443 bp open reading frame of *TcPanK* from *T. cruzi* genomic DNA using primers (TcPK F and R) which incorporated flanking Sbf-I and Asc-I restriction sites. The digested product was cloned into plasmid pTubEX (Appendix SF6) [277], containing targeting sequences for the tubulin array. Comparison of *Tc*PanK with the homologues in *T. brucei* strain 427 revealed that the two genes were 59% identical at the nucleotide level and lacked conserved stretches ≥20 bp, indicating that RNAi targeting of the *T. cruzi* transcript was unlikely to occur.

pTub*Tc*PanK-mNG (SF8) containing wild-type *Tc*PanK tagged at the C-terminus with mNeonGreen (with a glycine-serine linker peptide separating the two ORFs) was generated by a triple insert cloning method. pTub$TcPanK^{WT}$ (Appendix SF7) was digested with Afl-II and Asc-I to remove a 3' 2 kbp fragment of PanK. The digest was gel purified to isolate the remaining 7 kbp fragment (insert 1). A 2 kbp *Tc*PanK fragment was amplified from an intact pTub$TcPanK^{WT}$ template using primers (NG F1 and R1) that amplified from upstream of the Afl-II site and replaced the stop codon with an Sbf-I restriction site. The product of this PCR was digested with Dpn-I, Afl-II and Sbf-I and then PCR purified (insert 2). Finally, the mNeonGreen ORF was amplified from pPOTv6 puro-puro mNG (SF3) using primers (NG F2 and R2) which inserted an Sbf-I restriction site and linker sequence (30 nt) at the 5' end and an Asc-I site at 3' terminus. The product was digested with Dpn-I, Sbf-I and Asc-I and PCR purified (insert 3) and the three inserts were ligated (insert 1 contained the plasmid backbone).

For complementation with S1254A and R1270A mutants, highly conserved PanK domain residues were chosen based on ClustalO sequence alignments using characterised PanKs from distantly related organisms (*Homo sapiens*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Arabidopsis thaliana)*. The mutations were incorporated into *Tc*PanK by site-directed mutagenesis using a two-step PCR protocol. In the first step, cloned *Tc*PanK template was amplified with primers (1254 F or 1270 F with TcPK R) which incorporated an alanine codon in place of Ser-1254 or Arg-1270 codons and a 3' Asc-I site. A partial (3') *Tc*PanK fragment was generated. This fragment was gel-purified and then used as a reverse primer in the second amplification step: forward (TcPK F) and reverse (product from PCR 1) primers were used at a concentration of 100 ng to extend across the remaining gene sequence using 30 ng cloned template. The complete *Tc*PanK ORF product with incorporated alanine codon substitutions and flanking restriction sites (Sbf-I and Asc-I) was digested with Dpn-I to remove contaminating template DNA and cloned into pTubEX.

To express the PanK domain alone in *T. brucei,* the PanK domain (nucleotides 3130-4443) was amplified from cloned *Tc*PanK template using a forward primer (PkD F) which inserted Sbf-I and a start codon (5') and reverse primer which inserted a 3' Asc-I site (PkD R). The Sbf-I, Asc-I and Dpn-I-digested product was ligated with appropriately digested and gel purified pTubEX.

The SDH motif within the predicted phosphodiesterase domain (nucleotides) was selected for mutagenesis based on 100% conservation of these residues in *Tc*PanK, human neutral sphingomyelinase 2 (nSMase 2, Q9NY59*)*, *staphylococcus aureus* Phospholipase C (PLC, P09978), *bacillus cereus* sphingomyelinase C (P11889) and *saccharomyces cerevisiae* Inositol phosphosphingolipid phospholipase C (ISC1, P40015), assessed by ClustalO multiple-sequence alignment. D331A and H332A substitutions were incorporated into the WT *Tc*PanK pTubEX construct by PCR. The pTub-*Tc*PanK^{WT} construct (10 ng) was used as a template with 500 µM primers (SDH F and R) that amplified the entire plasmid and incorporated alanine codon substitutions in a single reaction of 25 µl volume. The following cycling conditions were used: initial denaturation of 30 seconds at 98°C, followed by 29 cycles of denaturation (10 seconds at 98°C), annealing (30 seconds at the lowest primer T_m + 3°C) and extension (20-30 seconds/kb at 72°C), and then a final extension of 2 minutes at 72°C. The PCR product was treated with KLD enzyme mix (NEB), containing a kinase, ligase and Dpn-I in a 10 µl reaction containing 1 µl un-purified product and 1x KLD reaction buffer (NEB). After incubation for 10 minutes at room temperature, the 10 µl mixture was directly used for transformation.

To generate hexa-histidine-tagged *Tc*PanK constructs (Appendix SF9,10), 10 ng pTub-*Tc*PanK^{WT} (Appendix SF7) was used as the template and primer pairs 'N His F and R' or 'C His F and R' (500 µM each) amplified the full-length plasmid and incorporated the hexa-histidine tag in frame with *Tc*PanK, separated by a glycine-serine linker (30 bp). A 25 µl reaction was set up using the following conditions: initial denaturation of 30 seconds at 98°C, followed by 29 cycles of denaturation (10 seconds at 98°C), annealing (30 seconds at the lowest primer T_m + 3°C) and extension (20-30 seconds/kb at 72°C), and then a final extension of 2 minutes at 72°C. 1 µl of PCR product was treated with 1x KLD enzyme mix (NEB) in a 10 µl reaction, incubated for 10 minutes at room temperature and then directly used for transformation (10 µl reaction:100 µl cells).

All pTubEX complementation constructs were linearized with Not-I and Xho-I, column-purified and heat-sterilised prior to transfection. Primers are listed in Appendix Table A.3.

2.4 Transfections

2.4.1 Transfection of *T. cruzi*

2 x 10^7 exponentially growing epimastigotes were electroporated with heat-sterilised PCR-generated constructs derived from 1 x 40 µl and 1 x 20 µl PCR reactions per targeting construct and sgRNA respectively. Electroporation was carried out in a 50 mM HEPES (pH 7.3) transfection buffer containing 90 mM sodium phosphate, 5 mM KCl and 0.15 mM $CaCl_2$ [278] using the Amaxa Nucleofector II system, set to programme X-001. Cells were allowed to recover in 5 ml fresh culture medium for 24 hours, after which time, selection was performed. This involved addition of the appropriate drugs (10 µg/ml blasticidin or 5 µg/ml puromycin) to each culture, followed by serial dilution across a 48-well plate to facilitate selection of clones. Plates were then incubated at 27°C.

2.4.2 Transfection of *T. brucei*

2 x 10^7 mid-log phase parasites were electroporated with 4-6 µg of heat-sterilised (72°C, 20 mins) linear construct in 50mM HEPES transfection buffer (see 2.4.1) [278]. Electroporated parasites were left to recover in 50 ml fresh HMI-9/FBS for 4-6 hours before adding the selection drugs hygromycin (2.5 µg/ml), phleomycin (2 µg/ml) and, in the case of complementation cell lines, blasticidin (10 µg/ml). The culture was plated in 48-well plates, which were incubated at 37°C under a 5% CO_2 atmosphere.

2.4.3 Transfection of HEK293T

Cells from a 60% confluent plate (5 x 10^6 cells per 9 cm plate) were incubated in 2% serum for 2 hours and then transfected by incubation with Polyethyleneimine (PEI, Aldrich). The incubation solution contained a 5:1 w/w ratio of PEI:DNA, using a minimum of 30 µg PEI and 5 µg DNA in 520 µl DMEM (serum and antibiotic-free). After mixing and incubating the freshly prepared solution for 10 minutes at room temperature it was applied to the cells, which were incubated in the mixture for 3-4 hours at 37°C (5% CO_2). After this time the solution was replaced with DMEM containing 10% FBS. Cells were incubated at 37°C (5% CO_2) for a minimum of 24 hours until harvest, during which time additional media changes were performed as required.

2.5 Preparation of parasite material

2.5.1 Parasite Genomic DNA isolation

Genomic DNA was isolated from parasites using the proteinase K/phenol/chloroform method according to Kelly (1993) [279]. Briefly, parasites were pelleted at 2000 g for 10 minutes, then washed once with PBS, followed by gentle resuspension in TENS buffer (50 mM Tris HCl, 50 mM EDTA, 50 mM NaCl, 1% SDS, pH 8). Proteinase K (100 µg ml^{-1}) was added and the mixture was incubated overnight at 37°C or for 2 hours at 55°C. DNA was extracted by application of an equal volume (v/v) of phenol/chloroform followed by centrifugation at 17000 g for 15 minutes to enable extraction of the upper aqueous phase and its precipitation with 2 volumes of isopropanol. The purified DNA was re-suspended in TE buffer (1 mM Tris, 100 µM EDTA, pH 7.2).

2.5.2 Parasite protein isolation

Proteins were isolated by pelleting of parasites at 2000 g for 10 minutes, followed by washing with PBS, and resuspension in Laemmli buffer (60 mM Tris-HCl pH 6.8, 10 % (v/v) Glycerol, 2% (w/v) SDS, 0.01 % (w/v) Bromophenol Blue and 5% β-Mercaptoethanol) using 10 µl per 1 x 10^7 parasites. Residual genomic DNA was sheared by brief (2 x 15 seconds) sonication.

2.5.3 Isolation of total RNA from parasites

T. brucei RNA was extracted using the RNeasy mini kit (Qiagen) using manufacturer's protocol. Briefly, ~1 x 10^7 cells were pelleted and re-suspended in Buffer RLT containing 1% β-Mercaptoethanol. Ethanol was added and the lysate was immediately applied to the RNeasy mini spin column for binding to the column membrane. On-column DNase digestion was performed and then contaminants were washed from the column using Buffers RW1 and RPE. Purified RNA was eluted in 40 µl RNase-free water.

2.6 Construct integration PCRs

To determine the success of construct integration mediated by *Sp*Cas9 in transfected *T. cruzi* cell lines, PCRs were performed on extracted gDNA (Chapter 4, Fig. 4.2.3). For N-(mNeonGreen) and C-(Scarlet or PTP) tagged cell lines, one primer within each pair targeted the unaltered part of the PanK gene and the second primer targeted the ORF of the fluorescence or PTP tag (Appendix Table A.4). Three cell lines with interrupted PANK ORFs were analysed: one containing the puromycin resistance gene (Puro) and two containing the blasticidin resistance marker (Bla 1 and 2) inserted into the PANK ORF. For these cell lines, one primer pair targeted the 5' UTR of *Tc*PanK and the 3' UTR of the inserted drug resistance marker (KO F1 and KO R1) and another pair amplified the drug resistance marker exclusively (KO F2 Bla and KO R2 Bla or KO F2 Puro and KO R2 Puro). gDNA from the parental *T. cruzi* Cl Brener T7-Cas9 cell line was amplified with each primer pair to assess specificity of PCR products. 200 nM primer was combined with 1x Phusion master mix and ~100 ng gDNA. PCR cycling conditions were as follows: Initial denaturation of 30 seconds at 98°C, followed by 30 cycles of denaturation (10 seconds at 98°C), annealing (30 seconds at the lowest primer T_m + 3°C) and extension (30 seconds/kb at 72°C), and then a final extension of 5 minutes at 72°C. Primer sequences are in Appendix Table A.5.

2.7 Western Blotting

Whole protein lysates collected in Laemmli buffer were boiled for 4 minutes at 98°C and centrifuged (2000 g for 5 mins) to pellet debris. Approximately 20 µg of each lysate was resolved on SDS-PAGE, using a 4-20% polyacrylamide gradient or stain-free TGX pre-cast gel (BioRad). Protein samples run on stain-free gels were visualised by UV-induced fluorescence using a ChemiDoc imager (Bio-Rad) to confirm equal loading. Gels were

incubated in 20% ethanol for 5 mins and then protein transfer to nitrocellulose membranes was performed using a high ionic strength buffer (336 mM Tris, 260 mM glycine, 140 mM tricine and 2.5 mM EDTA pH 8.8) [280] and the Trans-Blot Turbo Transfer System (Bio-Rad) at 25 V for 10 minutes. Blocking was performed for 30 minutes using 5% skimmed milk in PBST and primary antibodies (mouse α-mNeonGreen (1:2000, Chromotek), mouse α-Ty1 (1:1000, Invitrogen), rabbit α-mCherry (1:2500, Abcam), mouse α-His (1:1000, Invitrogen)) were incubated at room-temperature for 3 hours or overnight at 4°C. After washing with 0.1% PBST, secondary antibody incubation was performed using horseradish peroxidase-conjugated α-mouse or α-rabbit secondary antibodies (1:5000, ThermoFisher) for 1.5 hours at room temperature. After 4 x 15 min PBST washes, reactive proteins were detected using an enhanced chemiluminescence substrate kit (Abcam). For Coomassie staining, gels were stained overnight at 4°C in 0.1% Coomassie Blue R250 (Sigma), 10% acetic acid and 50% methanol and de-stained in water.

2.8 Southern Blotting

Prior to blotting, gDNA was treated with RNAse A and removal of RNA contamination was verified by gel visualisation. gDNA was then digested with Nsi-I at 37ºC overnight. Digested products were separated on a 0.8% agarose/TBE (Tris-Borate-EDTA) gel. Gels were then incubated in the following solutions in series: 0.25 M HCl (15 minutes), 0.5 M NaOH /1.5 M NaCl (60 minutes), 0.5 M Tris-HCl, /1.5 M NaCl pH 7.6 (30 minutes) and then capillary-transferred onto nylon membrane (Zeta probe, Bio-Rad) using 10x SSC (150 mM Sodium Citrate, 1.5 M NaCl). DNA was cross-linked to the membrane by UV cross linking using 120 mJ/cm^2. For hybridisation, DNA probes complementary to a 390 bp (Blast) or 404 bp (PanK) target sequence on the blasticidin resistance ORF and PanK ORF, respectively, were made by chemical labelling of the fragment using digoxigenin-11-dUTP (DIG, Roche), which was achieved by PCR. The probe reaction was set up with the following reagents: DIG-11-dUTP (dATP, dCTP, dGTP, all at 2 mM; dTTP at 1.8 mM; and DIG-11-dUTP at 0.2 mM), 1x reaction buffer, 1.5 mM $MgCl_2$, 1 µM each primer (Appendix Table A.6), deionised water, cloned template DNA (10 ng) and Taq polymerase. A control PCR reaction was also set up, containing standard dNTPs (dATP, dCTP, dGTP, dTTP, all at 2 mM) in place of the DIG-11-dUTP mixture. The following PCR conditions were used: Initial denaturation (2 minutes at 95°C), followed by 30 cycles of denaturation (30 seconds at 95°C), annealing (30 seconds at 58°C), extension (27 seconds at 72°C), then a final extension of 5 minutes at 72°C. Success of the labelling reaction was indicated by slower migration of the DIG-labelled PCR product on an agarose gel relative to the control. Prior to addition of the probe, the membrane was

blocked using a solution of 5x SSC, 1x Denhardt's solution, 0.5% SDS and 0.1 mg/ml herring sperm DNA (heat-denatured, sheared and sonicated), incubated at 65°C for 1 hour. 25 ng/ml of probe was denatured (95°C for 5 minutes), cooled on ice (5 minutes) and then added to the blocking solution for overnight incubation at 65°C. The probed membrane was washed in 0.2x SSC containing 0.05% SDS (2 x 1 hour at 65°C) and then equilibrated for 5 minutes in Washing buffer (0.1 M maleic acid, 0.15 M NaCl (pH 7.5), 0.3% Tween-20), followed by blocking for 30 minutes in Blocking Solution (1% blocking reagent (Roche), 0.1 M maleic acid and 0.15 M NaCl (pH 7.5)). Detection was performed by a 30 minute incubation with anti-DIG conjugated to alkaline phosphatase (Roche) diluted 10,000:1 in Blocking Solution, followed by rinsing in Washing Buffer, equilibration in Detection Buffer (0.1 M Tris-HCl, 0.1 M NaCl (pH 9.5)) and application of the chemiluminescent substrate CSPD (Roche) diluted 100:1 in Detection Buffer (15 mins at 37°C). Chemiluminescent imaging was performed using a BioRad ChemiDoc imager.

2.9 Immunofluorescence imaging

For immunofluorescence, exponentially growing parasites were fixed in 2% (v/v) paraformaldehyde in PBS for 30 minutes, then washed twice in PBS and air-dried onto glass slides. Parasites were then permeabilized by incubation with 0.1% TritonX-100 on ice for 2 minutes, followed by 3 washes in PBS for 5 min each. Afterwards, slides were blocked with 10% donkey serum in PBST (PBS + 0.1% Tween 20) for 30 min. Primary antibody incubation was performed overnight in a humid chamber at 4ºC with mouse anti-Ty (1:500), followed by 3 washes with PBS and incubation with which donkey anti-mouse AlexaFluor 488 (1:1000) for 2 hours at room temperature. Finally, slides were washed 3 times with PBS before being mounted with Vectashield Antifade Mounting Medium with DAPI (Vector Laboratories, Inc.), sealed with nail varnish and imaged under an inverted Nikon Eclipse Ti epifluorescence microscope. Images were acquired with Nikon Ti2 Flash 4.0 software and using 60 and 100x objectives. To assess antibody specificity, a control sample containing wild-type parasites was prepared according to the same protocol.

2.10 *T. brucei* growth curves

Proliferation of *T. brucei* 2T1 BSF cell lines was measured by seeding an equal starting number (~5 x 10^4) of mid-log phase cells in a 24-well plate, with 3 wells seeded per experimental group. For each cell line an uninduced group was included as well as an induced group which received tetracycline (1 µg/ml) at the time of seeding (T=0 hours). The parental RNAi cell line was monitored alongside complementation clones to confirm that the RNAi

response was intact. In CoA rescue experiments, an additional experimental group containing tetracycline (1 µg/ml) and 1 mM CoA was included in analyses. CoA was delivered at the time of tetracycline induction or at 20 hours-post-induction as specified. Every 24 hours following initial CoA supplementation, the compound was re-administered to the culture medium to a final concentration of 1 mM. Growth was monitored at 2-5 consecutive time-points every 24 hours (24, 48, 72, 96 and 120 hours) post-induction. Each well was thoroughly mixed by pipetting and then counted using a haemocytometer (3 samples per well were counted in 3 chambers per sample). Blind counting was performed on each occasion. After counting, each well was diluted to a density of 1 x 10^5 parasites to maintain logarithmic growth. The counts were compiled into a cumulative growth curve using GraphPad Prism software.

2.11 RNA quantification by qPCR

To prepare RNA from *T. brucei* BSF cell lines (RNAi *Tb*PanK and *Tc*PanK complementation clones), parasites were grown to a density of 5 x10^5 per ml in 100 ml of HMI-9 medium, at which point another 100 ml media was added and the culture was split into 2 flasks. To one flask tetracycline (1 µg/ml) was added. 16 hours after induction, cells were pelleted and RNA was harvested using the RNeasy mini kit (Qiagen). The quality of RNA was assessed using a Qubit Fluorimeter (ThermoFisher) and by visualisation on an agarose gel.

cDNA synthesis reactions were performed in triplicate per RNA sample using 1 µg of total RNA, 1x SuperScript enzyme (ThermoFisher) and 1x VILO reaction mix (ThermoFisher) in 20 µl. Reverse-transcriptase free control reactions were also set up where Superscript was substituted by RNase-free water. Reactions were incubated according to the program: 25°C for 10 minutes, 42°C for 90 minutes and 85°C for 5 minutes.

Each cDNA sample (3 per experimental group) was diluted 1:5 in RNase-free water and 1.5 µl of diluted cDNA was amplified in triplicate per primer pair in a 25 µl qPCR reaction containing 1x SYBR Green master mix (Qiagen) and 400 nM each of forward and reverse primers. Pooled reverse-transcriptase free samples were also diluted 1:5 and amplified with each primer pair. Primers targeted 108-112 bp fragments of *Tb*PanK, *Tc*PanK or *T. brucei* Telomerase reverse transcriptase (TeRT). Sequences can be found in Appendix Table A.7. Reactions were performed using Applied Biosystems 7500 fast RT-PCR machine (ThermoFisher) as per the following cycling conditions: 1 cycle of 95°C for 15 min, followed by 40 cycles of 94°C for 15 s, 59°C for 20 s and 72°C for 30 s.

The specificity of amplification was confirmed by melting curve analysis: a single melting curve peak was confirmed for each reaction to establish that any differences in Ct values are valid and not due to amplification of non-specific products or primer-dimer formation.

To quantify mRNA abundance for each experimental group, the mean cycle threshold (Ct) was calculated for three cDNA samples (derived from one RNA sample), each one amplified in triplicate (see Fig. 2.11). These values were used to derive $2^{-\Delta\Delta Ct}$, a measure of fold change in gene expression, calculated according to the ΔΔCt method [281]. For quantification of *Tb*PanK mRNA relative to the control TeRT:

1. Δ**Ct *Tb*PanK** = mean **Ct *Tb*PanK-** mean **Ct TeRT**
2. ΔΔ**Ct *Tb*PanK** = Δ**Ct *Tb*PanK** - mean Δ**Ct *Tb*PanK** of the uninduced clones
3. 2^-ΔΔ**Ct *Tb*PanK**

To quantify ***Tc*PanK** relative to ***Tb*PanK** in each RNA sample:

1. Δ**Ct *Tc*PanK** = mean **Ct *Tc*PanK-** mean **Ct TeRT**
2. ΔΔ**Ct *Tc*PanK** = Δ**Ct *Tc*PanK-** mean Δ**Ct *Tb*PanK** of the uninduced clones
3. 2^-ΔΔ**Ct *Tc*PanK**

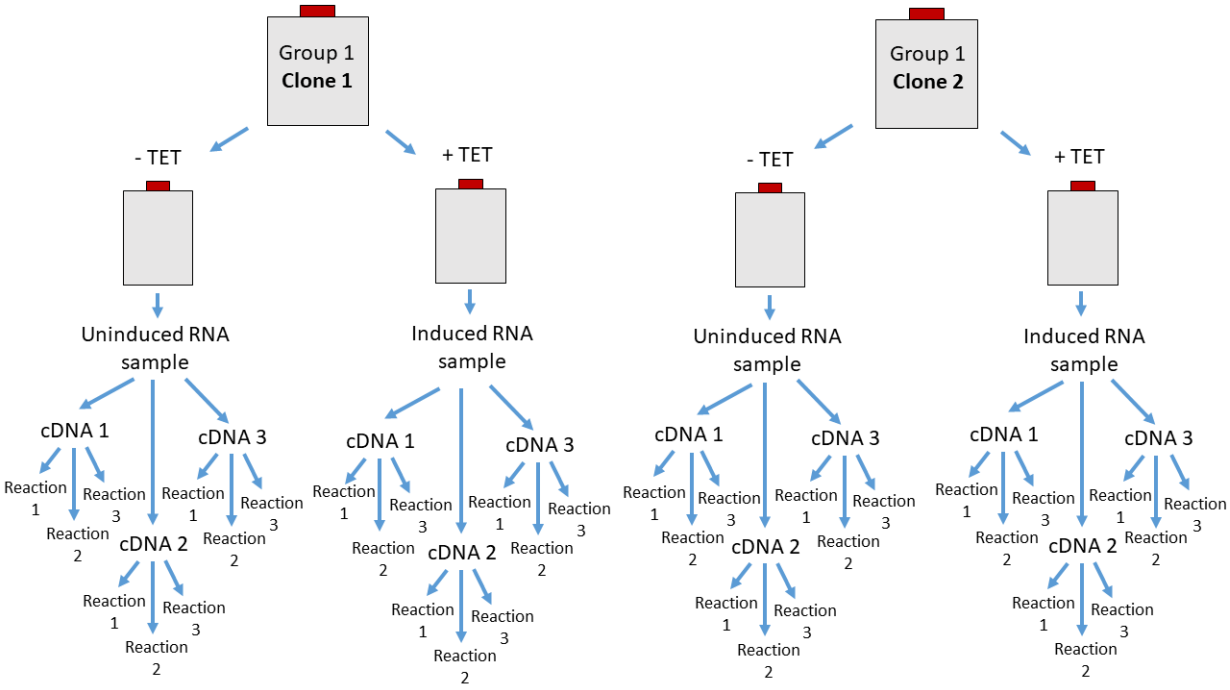


Fig. 2.11: Overview of samples collected for qPCR analysis of each RNAi experimental group: *For each experimental group, samples were obtained from two parasite clones. One flask of parasites was divided to create the uninduced and induced groups. Each extracted RNA sample was amplified by RT-PCR to derive three cDNA samples. Each cDNA sample was used in triplicate qPCR reactions.*

2.12 Recombinant protein expression

All herein described constructs were sequenced to verify successful cloning before being used in protein expression analyses.

2.12.1 Cloning and expression of recombinant *Tc*PANK using NEBExpress cell-free system

The entire ORF of *Tc*PanK (4443 bp, TcPK primers) or the PanK catalytic domain (recTcPK-domain, nucleotides 3130-4443, PK-D primers) was amplified from cloned DNA using 'Nt' or 'Ct' primers which incorporated a hexa-histidine tag at the N or C terminus, respectively (Appendix Table A.8). An Nde-I restriction site containing an in-frame Start codon was inserted via the forward primer and a BamH-I site was incorporated by the reverse primer. A 10 amino acid (Glycine-Serine) linker sequence was incorporated between the tag and gene ORF. PCR products were cloned into a NEBExpress plasmid containing a T7 promoter (provided in the NEBExpress cell-free synthesis kit, SF11) which required excision of a DHFR gene by complete digestion followed by gel purification. For expression, the following components were combined in a final volume of 50 µl in a 1.5 ml Eppendorf: plasmid template (150 ng/µl), 12 µl NEBExpress S30 Synthesis Extract, 1x Protein Synthesis Buffer, 50 U T7 RNA

Polymerase (NEB), and 40 U Murine RNase Inhibitor (NEB), milli-q water. The reaction was incubated at 37°C for 4 hours with vigorous shaking and generated products were analysed by SDS-PAGE.

2.12.2 Cloning and expression of recombinant *Tc*PANK in HEK293-T

*Tc*PanK inserts (full-length ORF or PanK domain nucleotides 3130-4443) were amplified from cloned DNA using a forward primer which incorporated an Acc-65I site, Kozak sequence (ACCATGG), hexa-histidine tag, 10 amino acid linker sequence and a reverse primer containing a Not-I site (Appendix table A.9). Inserts were cloned into pcDNA3.1+ (SF12). For HEK293T expression, plasmid constructs were sterilised by heating at 72°C for 20 minutes and a minimum of 3 µg DNA was transfected per 5 x 10^6 cells by Polyethyleneimine (Aldrich) incubation, using cells at approximately 60% confluency. Protein was collected from cells at 24-, 48- or 72-hours post-transfection to determine the optimal duration. A duration of 48 hours was used in subsequent experiments based on it providing the optimal yield. Cells were washed with PBS and collected using a cell scraper. Cells were pelleted at 2500 g for 5 mins, washed with PBS containing 1x protease inhibitor (EDTA-free, Roche) and re-suspended in RIPA buffer (150 mM sodium chloride, 1.0% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris, 1x protease inhibitor, pH 8.0). 3 cycles of freeze/thawing were performed followed by sonication (30 seconds x 3), and then a 15-minute incubation on ice with vigorous agitation. Finally, lysates were pelleted at 14000 g for 15 minutes (4°C) and analysed by SDS-PAGE.

2.12.3 Cloning and expression of recombinant *Tc*PANK in *E. coli*

Wild-type PanK inserts (full-length *Tc*PanK ORF or nucleotides 3130-4443) were amplified from cloned DNA using primers which inserted BamH-I (5') and Xho-I (3') sites, for cloning into vectors pTrcHisC (Invitrogen, SF13), or pET43.1a (Novagen, SF14). (Appendix Table A.10). The BamH-I site in pTrcHisC is in frame '+2' therefore additional nucleotides 'GA' were incorporated upstream of PanK nucleotides 4-4443 in the full-length construct, while cytosine at position 3130 was removed from the PanK domain construct. In pET43.1a the BamH-I site is '+1' so this was not required. Both vectors contained a start codon and N-terminal 6x Histidine tag. pTrcHisC, a 4.4 kbp vector contains a hybrid Trp/lac promoter (Trc) and pET43.1a contains a T7 promoter. The following constructs were generated: *Tc*PK^{WT}pTrcHisC and *Tc*PK^{WT}pET43.1a, containing the full *Tc*PanK ORF, and *Tc*PK-Domain^{WT}pTrcHisC and *Tc*PK-Domain^{WT}pET43.1a containing residues 1044-1480.

To create the construct *Tc*PK-Domain^{MT}pTrcHisC, a PanK domain fragment (nucleotides 3130-44430) carrying the R1270A substitution was generated using *Tc*PK^{WT}pTrcHis as a template for PCR mutagenesis using the Q5 Site-Directed Mutagenesis Kit (NEB). In a 25 µl reaction, 10 ng of template (5.8 kbp) was amplified using 1x Q5 High-Fidelity or Phusion Master Mix, and 0.5 µM of primers 'Trc Pk-D-Mt F' and 'Trc Pk-D-Mt R' (Appendix A.10), which replaced the codon CGC encoding arginine 1270 with GCC (encoding alanine). The cycling conditions were as follows: 30 seconds at 98°C; 29 cycles of 10 seconds at 98°C, 30 seconds at 69°C and 2 minutes at 72°C; final 2 minutes at 72°C. 1 µl of PCR product was combined with 1x KLD enzyme mix (NEB) containing kinase, ligase and Dpn-I enzymes in a 10 µl reaction. After a 10 minute incubation the ligated, product was used directly in transformation.

For expression, *E. coli* BL21(DE3), were transformed with pTrcHisC-based constructs and empty vectors (as controls) and resulting colonies were grown overnight at 37°C and 210 rpm in NZCYM containing 1% glucose. The overnight stationary phase culture was diluted 1:20 in 300 ml of fresh medium and grown in a baffled flask. Once the culture reached an optical density (OD_{600}) of 0.4 units, expression was induced with 50 µM isopropyl-β-D-thio-galactopyranoside (IPTG). Incubation was continued for 3-6 hours at 37°C or overnight at 16°C-23°C. Alternatively, overnight inoculates were diluted 1:1000 in ZYM-5052 auto-induction media and incubated at 37°C/210 rpm until turbid (5-6 hours). The temperature was then lowered to 20°C for expression to occur overnight. Cells were harvested by centrifugation at 4000 g for 15 minutes.

Lysis was performed by re-suspending pellets in 10-15 ml of ice-cold lysis buffer (20 mM Tris, 10% Glycerol, 700 mM NaCl, 0.02% tween, 5 mM $MgCl_2$, 5 mM imidazole, 0.5 mg/ml lysozyme pH 7.4) containing freshly prepared 1x EDTA-free complete protease inhibitor cocktail (Roche) and freeze-thawing (x3). DNA was sheared by sonication (3 x 15 second bursts). Fractionation of the soluble lysate was carried out at 10,000 g for 20 minutes at 4°C.

2.12.4 Purification of recombinant *Tc*PANK

For purification, 900 µl HisPur nickel resin (Thermofisher) was washed in distilled water and equilibrated in lysis buffer before being combined with the soluble lysate. Overnight binding was performed by rotation at 4°C. The resin-protein mixture was then applied to a poly-prep column and washed continuously with approximately 80 ml of sodium phosphate buffer (5% glycerol, 10 mM imidazole, 50 mM sodium phosphate, 750 mM sodium chloride; pH 7.4) by gravity-flow, until the flow-through no longer contained detectable protein when applied to Bradford reagent (Thermo). Elution was achieved using an imidazole gradient (40-100 mM): the resin was rinsed with 1 ml of elution buffer (5% glycerol, 50 mM sodium phosphate, 300 mM sodium chloride; pH 7.4) containing 40 mM, 50 mM and then 100 mM Imidazole. Each

elution fraction was analysed by SDS-PAGE using Stain-Free TGX gels (BioRad). The purest fractions were concentrated and exchanged to storage buffer (50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 300 mM NaCl, 10% glycerol) using a 10 kDa Amicon Ultra-4 centrifugal filter (Merck-Millipore). The purified and concentrated protein samples were quantified using the Bradford micro assay protocol published by BioRad. Purified samples were stored at -70°C with 40% glycerol, 1 mM DTT and 0.5 mM EGTA in aliquots.

2.13 Measurement of recombinant *Tc*PANK activity

BL21-expressed recombinant *Tc*PK-Domain (rec*Tc*PK-domain) activity was measured indirectly using a pyruvate kinase-lactate dehydrogenase coupled assay in which reduction of NADH is followed spectrophotometrically at 340 nm (ε =6270/M/cm). See Fig. 2.13. Each assay was set up in triplicate in a 96-well plate. Reactions contained 1 mM PEP, 33 U PK/LDH, 200 µM NADH, 10 mM $MgCl_2$, 100 mM Tris and 8 µg purified rec*Tc*PK-domain in a volume of 150 µl. Kinetic parameters for substrates ATP and pantothenate (Pan) were determined for rec*Tc*PK-domain^{WT} using 0.1-5 mM ATP and 4 mM Pan or 0.05-4 mM Pan and 3.5 mM ATP. Inhibitory assays were performed using serial dilutions of Coenzyme A (CoA) or Acetyl-CoA (Sigma), within the range of 0-24 µM and saturating concentrations of pantothenate (4 mM) and ATP (2 mM). Each inhibitor concentration was assayed in duplicate (CoA) or triplicate (acetyl-CoA). Initial velocities of the mutant rec*Tc*-PK-domain were calculated for 0.5, 1, 2 and 3 mM ATP using 4 mM pantothenate and 8 µg purified protein.

Activity was measured continuously for a total of ~30 minutes at 24°C. 340 nm absorbance readings were taken every 30 seconds before and after adding ATP to initiate reactions. Before ATP addition, readings were taken for approximately 5 minutes to obtain a rate for background activity. After adding ATP, absorbance was measured for >20 minutes or until absorbance readings reached a plateau.

Graphical analyses were performed using (GraphPad Software Inc., San Diego, USA). Absorbance values were plotted against time (minutes) and reaction rates before and after ATP addition were calculated using initial regions of constant slope. Background rates were subtracted from post-ATP reaction rates to obtain initial velocities (Vo) for each substrate concentration (n=3). To calculate kinetic parameters, mean initial velocities (n=3) were plotted (in units of nmol NADH/minute) against ATP or pantothenate concentrations (mM) and non-linear curve-fitting was performed. The Michaelis-Menten model was used to determine V_{max} and K_m. K_{cat} was calculated using a constrained Et (enzyme concentration) value of 0.156 nmol (based on 8 µg enzyme and 51 kDa mass). Lineweaver-Burk plots (1/Vo against

1/[substrate]) were generated for kinetics datasets for further visual confirmation of Michaelis-Menten behaviour of reactions. Inhibition of *Tc*PK-Domain^{WT} activity was calculated using at least 8 concentrations of inhibitor. Initial reaction rates at each inhibitor concentration were plotted as a percentage of measured rates in the absence of inhibitor (% maximal activity). Non-linear curves fitted to this data were used to derive inhibition parameters IC50 and K_i. For the K_i model the following constraints were applied: E_t = 1.04 µM (both substrates); S = 2000 µM (ATP), 4000 µM (Pan); and Km = 990 µM (ATP), 240 µM (Pan). All data are provided as best-fit values with a 95% confidence interval range.

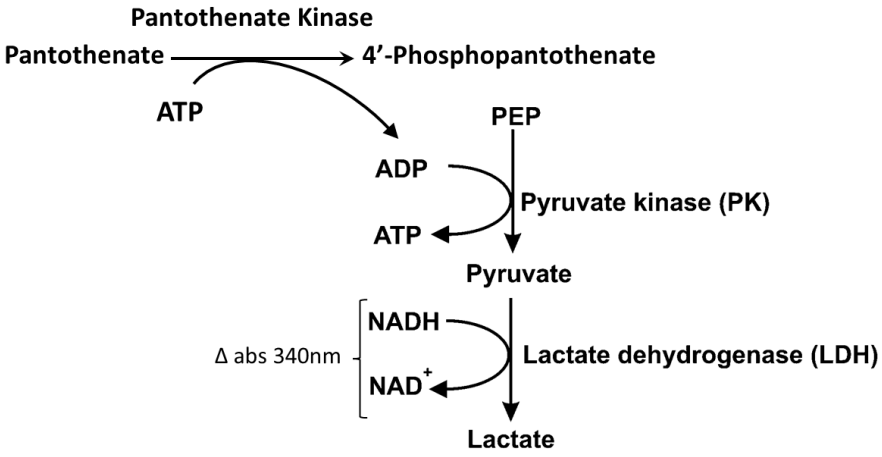


Fig. 2.13 Basis of 4'phosphopantothenate synthesis assay

2.14 Construct sequencing

Sanger sequencing of plasmid constructs was performed by Eurofins genomics (Germany). Chromatogram files were manually inspected using Chromas and sequence integrity was inferred using BLAST.

3. Bioinformatic analysis predicts a multifunctional kinetoplastid pantothenate kinase

3.1 Introduction

Biosynthesis of the metabolically crucial cofactor CoA shows promise as a target for antimicrobial and antiparasitic drugs, yet the potential exploitability of this pathway in kinetoplastids has largely been ignored. The purpose of this doctoral research was to characterise kinetoplastid PanK, the first enzyme in the CoA synthesis pathway. The initial step, which is discussed in this chapter, was a comprehensive bioinformatic analysis. Searches of trypanosomatid genome databases were performed to identify candidate enzymes of the CoA pathway and the extent of their homology was determined by sequence alignments - between kinetoplastids themselves and distantly related organisms. Basic local and domain-enhanced alignments revealed additional domains fused to kinetoplastid PanKs, which have not been previously described. Conservation within amino acid sequences and predicted 3D structures was assessed to infer the functions of these unique PanK-associated domains.

3.2 Conservation of the CoA pathway in kinetoplastids

To determine the completeness of the CoA biosynthesis pathway in the clinically important trypanosomatids *T. cruzi, T. brucei* and *L. major*, I searched for genes encoding the enzymes required for each catalytic step and transporters for uptake of extracellular pantothenate. Blastp searches of the TriTryp database were performed using functionally characterised protein sequences to identify putative homologues in the specified genomes. The query sequences used in these searches can be identified from the NCBI Gene IDs provided in Table 3.2.

The enzymes which synthesise pantothenate from L-valine: branched-chain-aminotransferase (BCAT), ketopantoate hydroxymethyl transferase (KPHMT), ketopantoate reductase (KPR), and pantoate-β-alanine ligase (PBAL) were absent from the genomes of interest, with the exception of BCAT which is found in *T. brucei* and *Leishmania* species (Table 3.2). However, homologues were identified for all activities downstream in the pathway.

Alignment of the bacterial sodium/pantothenate symporter (PanF) identified dispersed gene family 1 (DGF-1) proteins as having significant similarity (<22%, E=0.057), but no homologous transporters were found. This was also the case for the mammalian sodium-dependent multivitamin transporter (SLC5A64), which most closely aligns to putative fructose-bisphosphate aldolases**,** and shares 23-25% sequence identity. On the other hand, a putative major facilitator superfamily (MSF) transporter was identified in all three trypanosomatids. These sequences share up to 21.5%, 20% and 22% percentage identity with Fen2 (*S. cerevisae*), Liz1 (*S. pombe*) and PfPAT (*P. falciparum*), respectively. Homologues for the enzymes that catalyse the conversion of pantothenate to CoA (see Fig.1.6.1.2) were identified in all three genomes. Compared with human enzymes, the percentage identities were 30% for pantothenate kinase (PANK), 31% for phosphopantothenoylcysteine synthetase (PPCS) and 36% for phosphopantothenoylcysteine decarboxylase (PPCDC), phosphopantetheine adenylyltransferase (PPAT) and dephospho-CoA kinase (DPCK).

Table 3.2. Genes encoding Pan and CoA biosynthesis and salvage proteins:

NCBI gene IDs for putative trypanosomatid proteins and query sequences used in BLAST. Where a partial query sequence was used, residues for that sequence are indicated in brackets. H. sap: Homo sapiens, S. cer: Saccharomyces cerevisiae, E. col: E. coli. NF, not found.

	Gene	Query ID	Query Organism	*T. cruzi* gene IDs	*T. brucei* gene IDs	*L. major* gene IDs
Pan synthesis	BCAT	586	H. sap	NF	3655760, 3655762	12982795
	KPHMT	852474	S. cer	NF	NF	NF
	KPR	856460	S. cer	NF	NF	NF
	PBAL	854661	S. cer	NF	NF	NF
Pan	panF	947752	E. col	NF	NF	NF
	Fen2	850394	S. cer	3550707	3661036	12980433, 12980436
	SLC5A6	8884	H. sap	NF	NF	NF
CoA synthesis	PANK	53354	H. sap	3547284, 3545218	3665554	5653219
	PPCS	79717	H. sap	3542047	3656136	5652671
	PPCDC	60490	H. sap	3550770, 3542588	3657929	5653684
	PPAT	80347 (1-350)	H. sap	3540384, 3547163	23866925	5656326
	DPCK	80347 (351-564)	H. sap	3548396, 3540932	3657706	5651926

For clarity, Fig. 3.2.1 provides an overview of CoA biosynthesis pathway conservation within the investigated kinetoplastid genomes.

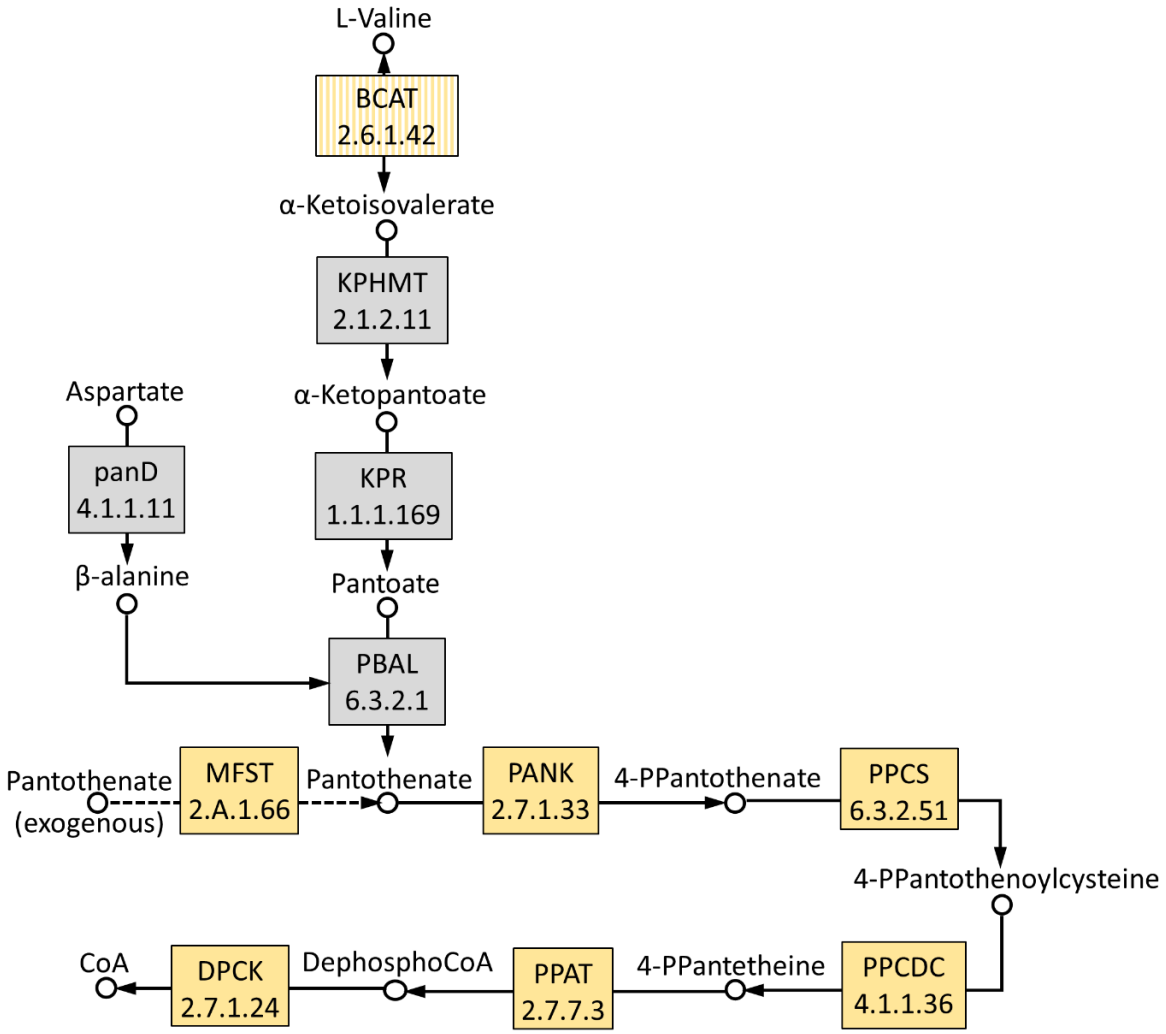


Fig. 3.2.1. Schematic of the predicted CoA Biosynthesis pathway in Kinetoplastids: *EC/TC numbers for enzymes and transporters in the CoA biosynthesis pathway. Presence in genomes of T. cruzi, T. brucei and L. major is indicated by orange colouring, grey denotes absence from all 3, striped orange denotes presence in T. brucei and L. major only.*

Each of the *T. cruzi* sequences identified from Blastp searches was then aligned with putative homologues from distantly related organisms to visualise the extent of conservation (Figs. 3.2.2-7). Alignment of Q4DVR4 (Fig. 3.2.2 A) with plasmodium and yeast MSF transporters revealed conserved sites throughout the protein length, but identical sites were discontinuous (groups of ≥2 identical residues) and overall percentage identity was below 20% in both cases. Due to the limited conservation, homology was further assessed by structural alignment using models generated by the artificial intelligence (AI) system AlphaFold. This software predicts 3D protein structure based on the amino acid sequence, with atomic accuracy competitive with experimental methods [260,261]. Structure models predicted for the putative *T. cruzi* homologue and *Pf*PAT consisted of α-helices arranged in parallel around a central channel,

consistent with the characterised yeast Fen2 (Fig. 3.2.2 B-D). A Pymol-generated alignment of the predicted parasite transporters strongly suggested structural homology between the two proteins (Fig. 3.2.2 E). Alignments of the predicted CoA biosynthesis enzymes PanK (Q4DHB2), PPCS (Q4D9S8), PPCDC (Q4DB42), PPAT (Q4D5T1) and DPCK (Q4DQ34) revealed many conserved sites shared with distantly related organisms (Figs. 3.2.3- 3.2.7). Q4DHB2 possesses motifs which are highly conserved amongst eukaryotic PANKs (Fig. 3.2.3), including the DIGGT(S)XXK motif required for phosphate binding of eukaryotic PanK catalytic domains, which is 100% conserved [282]. The downstream enzymes are clearly highly conserved in sequence, containing sites which are identical in eukaryotes and bacteria (Figs. 3.2.4- 3.2.7).

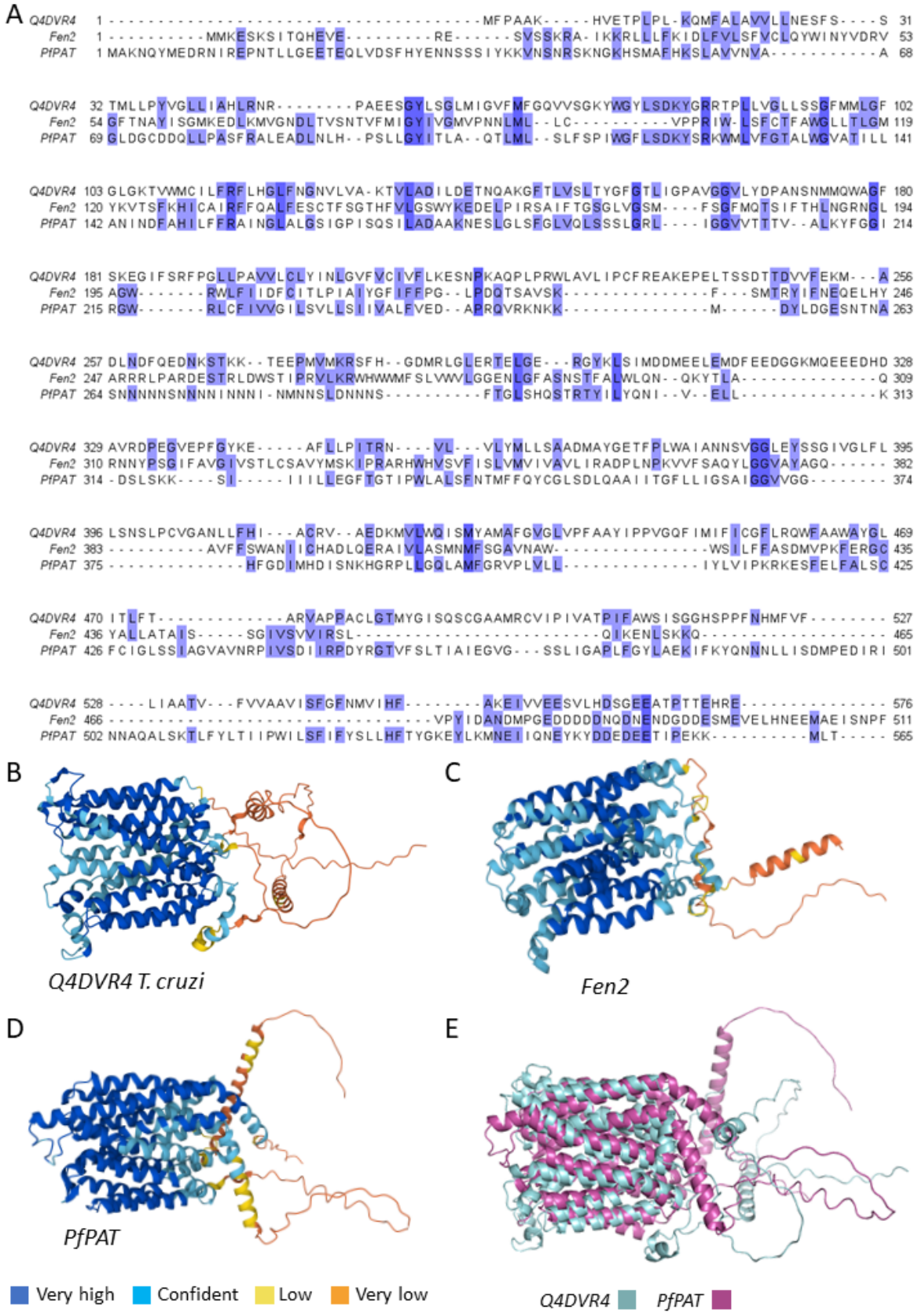


Fig. 3.2.2. T. cruzi protein Q4DVR4 aligned with major facilitator superfamily transporters: ***A****: ClustalO aligned protein sequences: T. cruzi (Q4DVR4), Fen2 of S. cerevisiae (P25621) and putative P. falciparum pantothenate transporter PfPAT (O96156). Light/dark shading indicates 70/100% residue identity.* ***B-E:*** *AlphaFold predicted structures for each protein, coloured by residue confidence scores (****B-D****) and a pymol alignment (****D****) of structures Q4DVR4 (blue) and PfPAT (purple).*

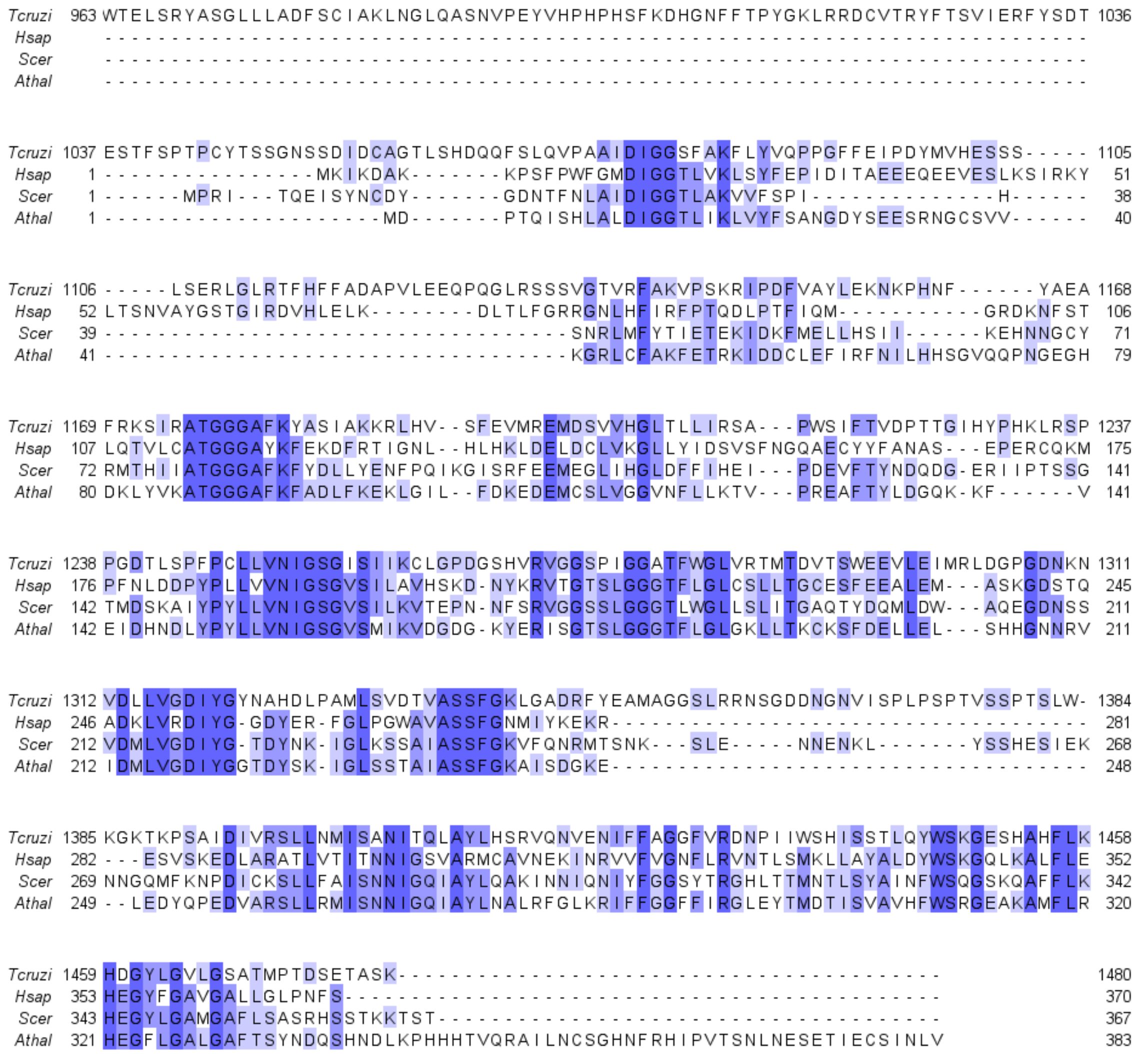


Fig. 3.2.3. PanK alignment: *ClustalO-aligned PanK sequences from T. cruzi (Q4DHB2, only residues 963-1480 displayed), S. cerevisiae (Q04430), Homo sapiens isoform PanK3 (Q9H999) and Arabidopsis thaliana (O80765). Shading indicates 50, 75 or 100% identity in order of intensity.*

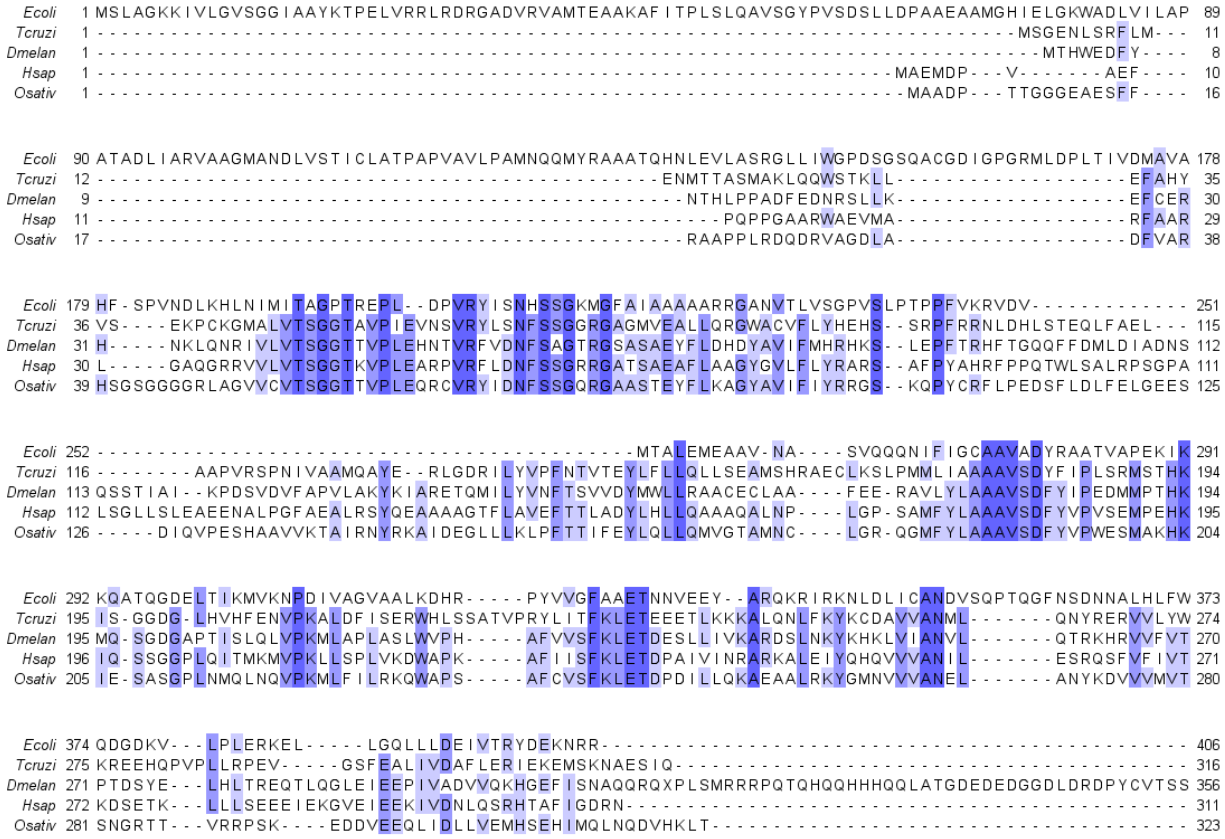


Fig. 3.2.4. PPCS Alignment: *ClustalO alignment of a putative PPCS of T. cruzi CL Brener (Q4D9S8) with E. coli bi-functional CoaBC (P0ABQ0) and PPCS of D. melanogaster (E2QD26), H. sapiens (Q9HAB8) and Oryza sativa Japonica (Q0J7N5). Shading as in Fig. 3.2.3.*

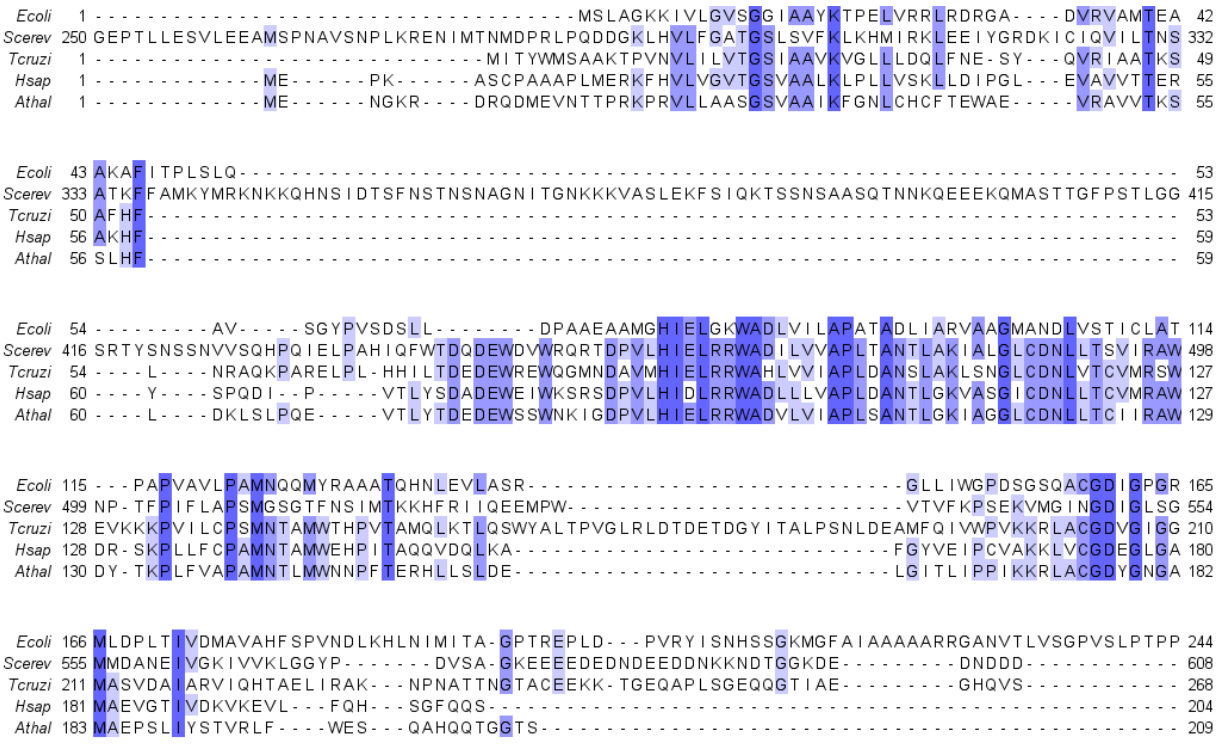


Fig. 3.2.5. PPCDC alignment: *ClustalO alignment of putative PPCDC of T. cruzi (Q4DB42) with sequences from E. coli (P0ABQ0), S. cerevisiae (Q08438), H. sapiens (Q96CD2) and A. thaliana (Q9SWE5). Shading as in Fig. 3.2.3.*

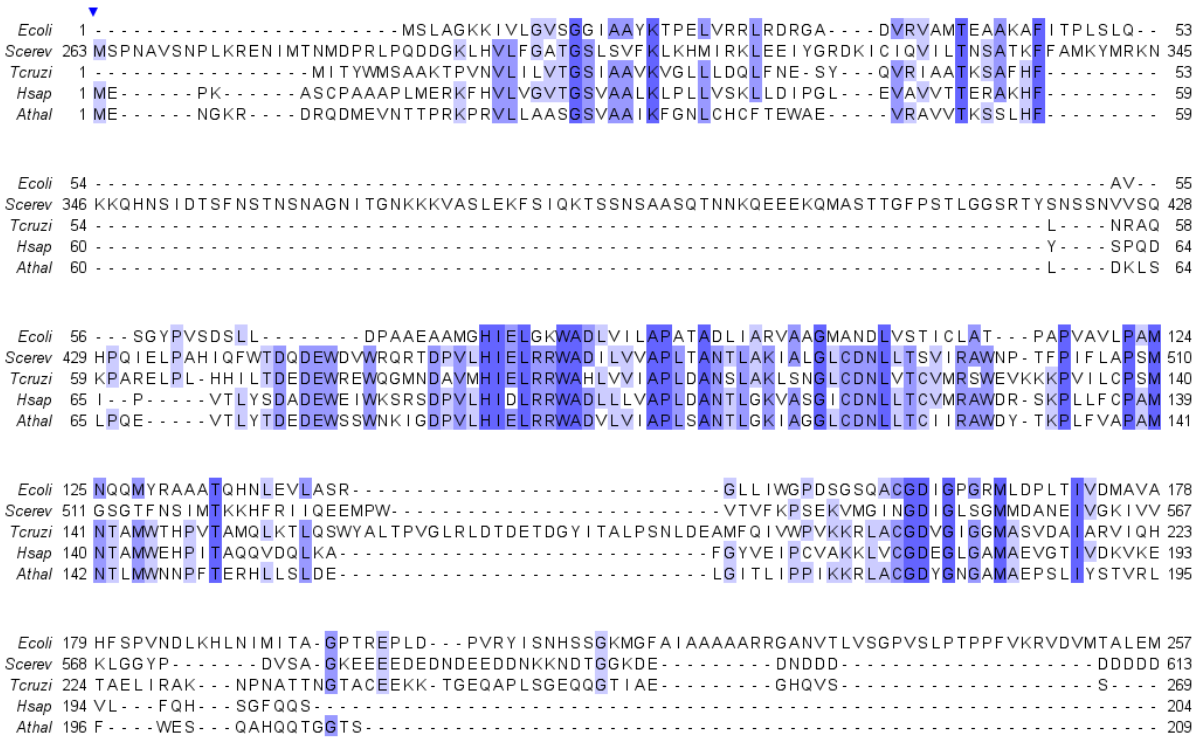


Fig. 3.2.6. PPAT alignment: *Putative T. cruzi PPAT (Q4D5T1) aligned with sequences from E. coli (P0A6I6), S. cerevisiae (P53332), H. sapiens (Q13057) and A. thaliana (Q9ZPV8). Shading as per Fig. 3.2.3.*

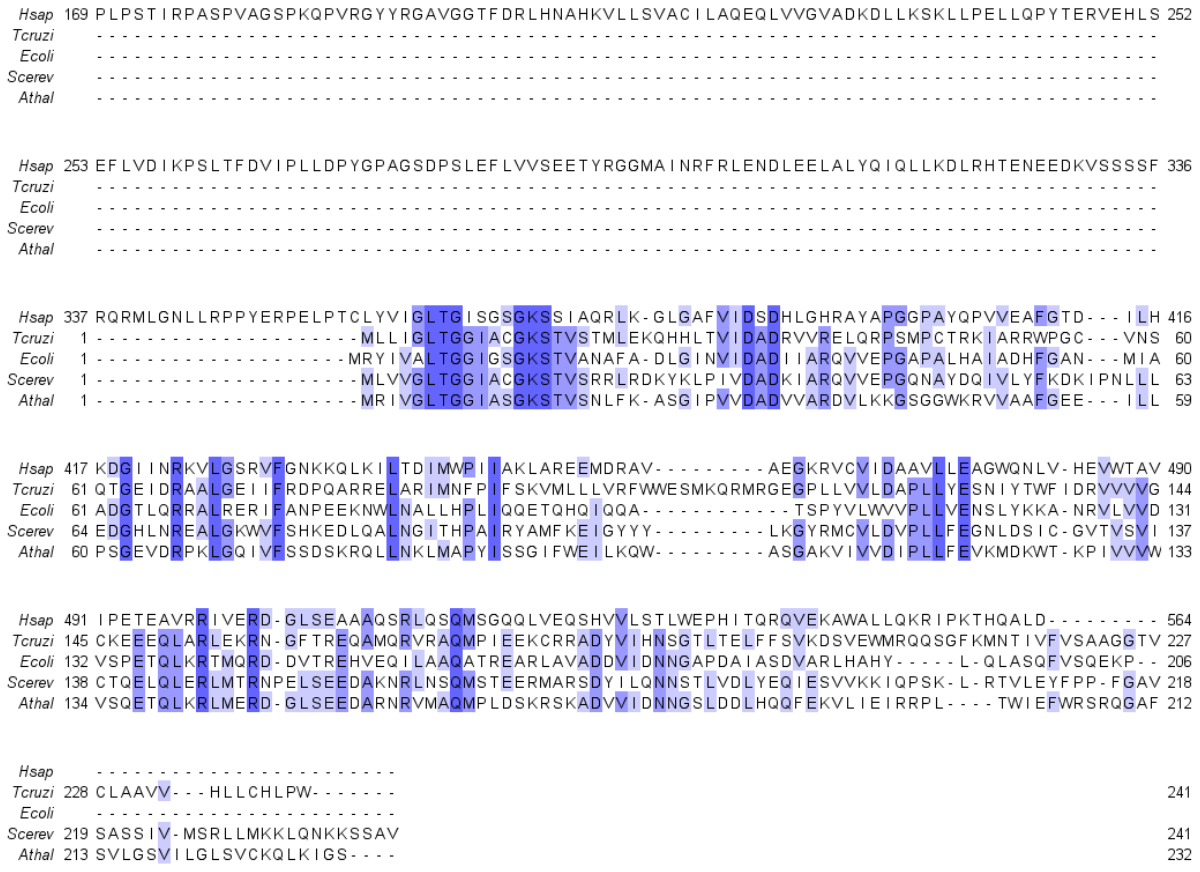
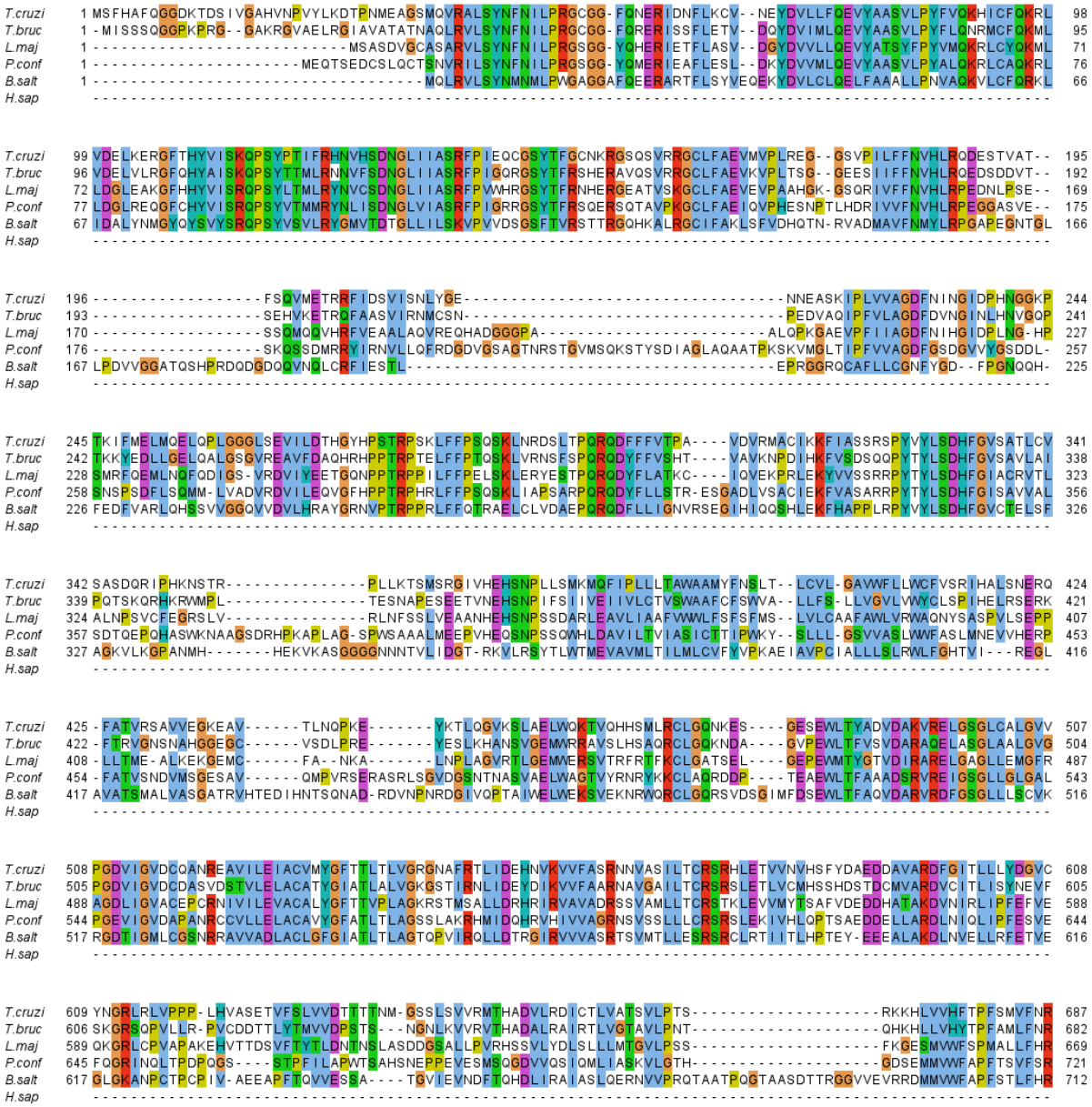


Fig. 3.2.7. DPCK alignment: *Putative T. cruzi DPCK (Q4DQ34) aligned with sequences from E. coli (P0A6I9), S. cerevisiae (Q03941), H. sapiens (Q13057) and A. thaliana (Q9ZQH0). Shading as in Fig. 3.2.3.*

3.3 Kinetoplastid pantothenate kinases contain multiple conserved domains

To more closely examine PanK in kinetoplastids, the *T. cruzi* CL Brener putative PanK (*Tc*PanK) protein sequence (Q4DHB2) was used as a query for a Blastp standard NCBI database search to identify homologues. This revealed that *T. cruzi* and other trypanosomatids possess at least one copy of a single PANK isoform, typically ranging from 1450-1480 residues in length. The first ~1000 residues seem to be unique to the kinetoplastid order, including bodonids.

This is demonstrated in Fig. 3.3.1, an alignment of putative PANKs from *T. cruzi, T. brucei* and *L. major,* the distantly related trypanosome *Paratrypanosoma,* the bodonid *Bodo saltans* and the characterised human PanK1 isoform (*Hs*PanK1). While PanK is highly conserved amongst kinetoplastids across >1400 residues, *Hs*PanK aligns only to residues 1075-1469 of *Tc*PanK and the corresponding C-terminus of kinetoplastid homologues. Conservation is generally high throughout the aligned region except for one major *Hs*PanK1 insertion (residues 133-185) and a small number of insertions or exclusively conserved regions in the kinetoplastid enzymes, including residues 965-1006, 1089-1096, 1215-1239 and 1346-1392 of *Tc*PanK. A number of insertions are present in *B. saltans* PanK, indicating that it is less related to those of other kinetoplastids.


T.cruzi
T.bruc
L.maj
P.conf
B.salt
H.sap

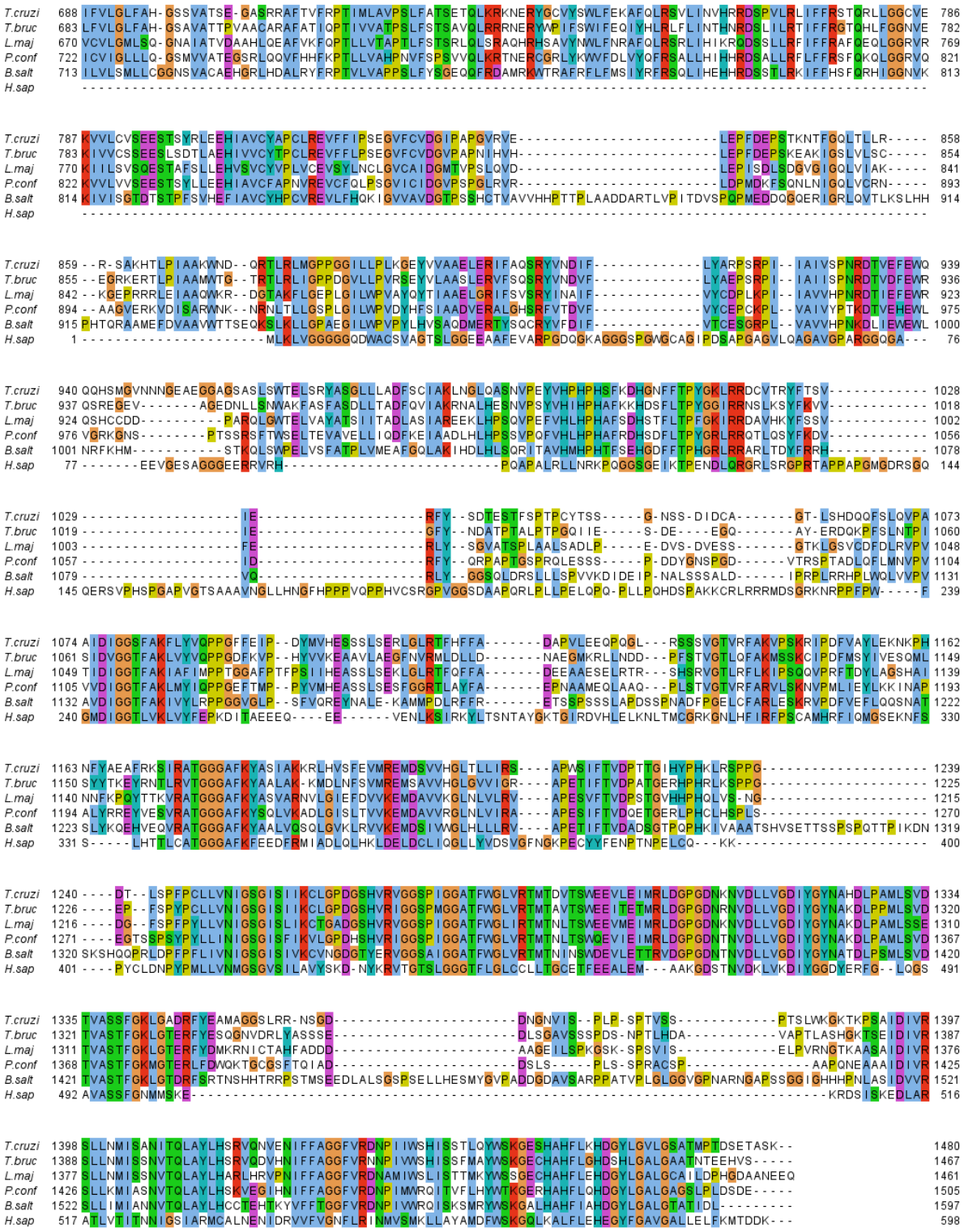


Fig. 3.3.1 Alignment between kinetoplastid and human pantothenate kinases: *ClustalO-aligned PanK sequences from T. cruzi (Q4DHB2), T. brucei (T. bruc, Q384T6), L. major (L. maj, Q4Q8Q6), P. confusum (P. conf, PCON_0021720*), B. saltans (B. salt, A0A0S4IR34) and H. sapiens (H. sap, Q8TE04) with ClustalX colouring (yellow: proline, orange: glycine, pink: cysteine, red: positive charge, magenta: negative charge, blue: hydrophobic, green: polar, cyan: aromatic, white: unconserved).*

To infer the functional significance of the kinetoplastid-specific PanK regions, domain analysis of *Tc*PanK was performed using the NCBI Conserved Domain Database (CDD). A number of

conserved domains show significant (E<0.001) similarity; the top hits are listed in Table 3.3. These hits suggest that *Tc*PanK contains three enzyme activities, encompassing residues ~38-340, ~450-1040 and ~1074-1480, which are predicted to belong to endo/exonuclease/phosphatase (EEP), adenylation and PanK superfamilies, respectively.

Table 3.3**: **TcPanK top domain hits listed by the conserved domains database.
Colours indicate domains according to the protein map below. EEP: endo/exonuclease/phosphatase

Name(s)	Accession	Description	Residues	E-value
Pantothenate kinase; fumble	PTZ00297	Provisional pantothenate kinase; trypanosomatids	28-1480	0e+0
	pfam03630	Fumble; required for cell division in drosophila; contains a domain with high similarity to fungal and mammalian pantothenate kinases; eukaryotes	1074-1468	1.0e-110
	PLN02902	Pantothenate kinase; embryophyte (land plants)	1035-1468	5.5e-51
	PLN02920	Pantothenate kinase 1; rosids (flowering plants)	1074-1468	1.8e-49
Neutral sphingomyelinase	cd09078	Phosphodiesterases that catalyze the hydrolysis of sphingomyelin to ceramide and phosphorylcholine; EEP superfamily; cellular organisms	38-339	3.6e-24
LC-FACS; LC-acyl-CoA synthetase	cd05927	Eukaryotic long-chain fatty acid CoA synthetase (LC-FACS) activates fatty acids with chain lengths of 12 to 20; forms fatty acyl-AMP molecule as an intermediate and fatty acyl-CoA as a product; eukaryota	480-1034	2.7e-23
	PLN02614	Long-chain acyl-CoA synthetase; arabidopsis	450-1041	9.5e-20
	PLN02861	Long-chain-fatty-acid-CoA ligase; magnoliophyta (flowering plants)	451-1039	5.7e-19
	COG1022	Long-chain acyl-CoA synthetase (AMP-forming); cellular organisms	455-783	1.0e-17
AMP-binding enzyme	pfam00501	AMP-binding enzyme; cellular organisms	478-726	7.4e-15
EEP-2	cd09084	Uncharacterised family within the EEP superfamily containing functionally diverse enzymes that share a common catalytic mechanism of cleaving phosphodiester bonds; bacteria	41-339	1.6e-09
RgfB-like	cd09079	Includes *Streptococcus agalactiae* RgfB (for regulator of fibrinogen binding) and related proteins; EEP superfamily; cellular organisms	56-339	8.5e-06
TDP2	cd09080	Phosphodiesterase domain of human TDP2, a 5'-tyrosyl DNA phosphodiesterase, and related domains; cleaves the DNA 5'-phosphodiester bond and restores 5'-phosphate termini; EEP superfamily; cellular organisms	71-339	4.1e-05

Of the EEP-family hits, the neutral sphingomyelinase (nSMase) domain family has the highest sequence similarity to the *Tc*PanK N-terminus (residues ~38-340, $E=3.56x10^{-24}$). Sphingomyelinases are phosphodiesterases that hydrolyse sphingomyelin, forming ceramide and phosphorylcholine. nSMases are involved in stress-induced cell responses, including cell death, adhesion, differentiation, and proliferation, via their lipid "second messenger" products. As well as phospholipids, nucleic acids are typical substrates of EEP enzymes, which have in common the ability to cleave phosphodiester bonds using metal ion cofactors (see Fig. 3.3.3 A).

The central domain (residues ~450-1040) hits are all members of the Adenylate Forming Domain class 1 superfamily (cl17068), a group of enzymes which include the adenylation domain of non-ribosomal peptide synthetases (NRPS) and acyl-CoA synthetases (Table 3.3). They catalyse an ATP-dependent condensation reaction which proceeds in two steps: first a carboxylate substrate is activated as an adenylate and then the carboxylate is transferred to the pantetheine group of coenzyme A or an acyl-carrier protein. The enzyme with the highest similarity is eukaryotic long-chain fatty acid CoA synthetase (LC-FACS, $E=2.67x10^{-23}$). Members of this family activate fatty acids with chain lengths of 12 to 20 carbons in a two-step reaction: a fatty acyl-AMP intermediate is formed, followed by a fatty acyl-CoA, which goes on to participate in catabolic and anabolic reactions (Fig. 3.3.3 B). There are typically multiple LC-FACS isoforms in an organism - e.g. mammals possess six and Arabidopsis have nine.

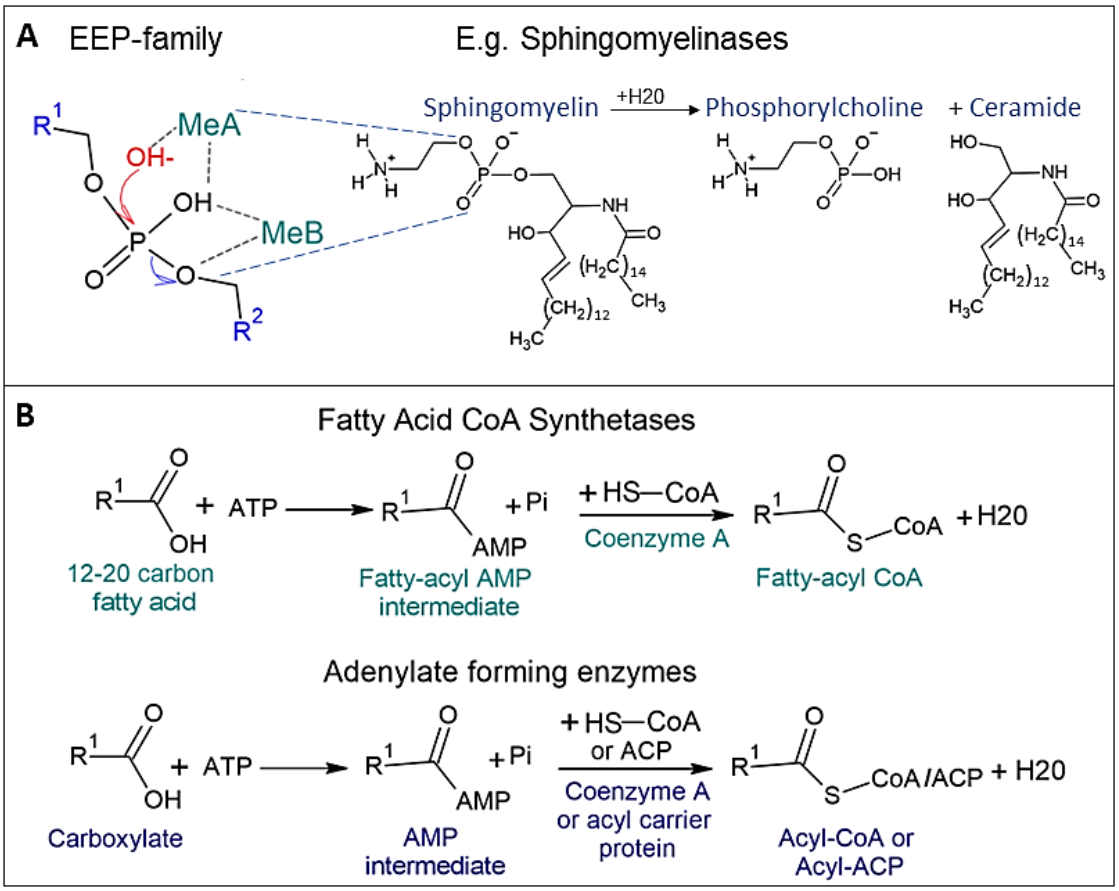


Fig. 3.3.3 General reaction mechanisms of the top domain hits.(A) *Endo/Exonuclease/phosphatase (EEP)-family proteins, such as sphingomyelinases which hydrolyse sphingomyelin to form phosphorylcholine and ceramide, break phosphodiester bonds using divalent metal ions (usually two metal ions, MeA and MeB) which trigger nucleophilic attack of the phosphate group. (**B**) Fatty acid CoA synthetases act via a two-step mechanism; first adenylating the fatty acid substrate and then transferring it to coenzyme A. The broader family of adenylate forming enzymes includes proteins which transfer the adenylated carboxylate to the pantetheine group of acyl-carrier protein (ACP), such as non-ribosomal peptide synthases.*

Finally, the C-terminus of *Tc*PanK (1074-1480) closely matches eukaryotic PanK domains, particularly the fumble domain (E=1.02x10^{-110}, Table 3.3). The entire sequence is recognised as a PTZ00297 domain, a putative pantothenate kinase of trypanosomatidae, for which there are no functional annotations due to lack of experimental data.

Next, the predicted domains were used as a guide for sequence-level analysis. We performed multiple-sequence alignments (MSA) using characterised enzymes belonging to EEP and FACS protein families and the unique kinetoplastid-specific domains 1 and 2, to determine the extent of homology at the residue level and identify conserved motifs with proven functional importance.

Firstly, MSAs were performed using the first 350 residues of *Tc*PanK and EEP family proteins. Representative alignments are illustrated in Fig. 3.3.4. The human DNA-binding proteins tyrosyl-DNA phosphodiesterase 2 (TDP2) and apurinic/apyrimidinic (AP) endonuclease-

1(APEX-1) and the thermophilic exonuclease III (ExoIII) of *Methanothermobacter thermautotrophicus* were chosen based on published experimental data confirming their functions. These enzymes have phosphodiesterase activities which are used in DNA repair. TDP2 repairs trapped topoisomerase II-DNA complexes by cleaving 5'-tyrosyl-DNA bonds [265], while APEX-1 and ExoIII cleave the DNA backbone at abasic sites during base excision repair [266,267]. In common to all three enzymes is a dependency on magnesium or other divalent cations for phosphodiesterase activity. At the sequence level, global conservation is low but four highly conserved motifs characterise cation-dependent endonucleases: S/TxN (where S or T is less commonly substituted with alanine (A) and x is commonly an aromatic amino acid), xQ/NE (where x is commonly leucine (L), valine (V) or isoleucine (I) and less commonly phenylalanine (F)), GDxN (where x is commonly F) and SDH. Mutation of each of these motifs has been shown to abolish enzyme activity *in vitro* [265,266]. As is clear from Fig. 3.3.4 A, all four motifs exist in *Tc*PanK: SYN (residues 42-44), FQE (residues 75-77), GDFN (residues 229-232) and SDH (residues 330-332), which strongly suggests that the protein possesses a magnesium (or other divalent cation)-dependent phosphodiesterase activity.

An alignment with neutral sphingomyelinases (nSMases) was also performed. *Tc*PanK was aligned with four nSMases: human neutral sphingomyelinase 2, *S. aureus* phospholipase C, *Bacillus cereus* sphingomyelinase C and *S. cerevisiae* Inositol phosphosphingolipid phospholipase C (ISC1) (Fig. 3.3.4 B). As well as the motifs which are broadly conserved amongst EEP proteins, another motif, SxxPI, is present in these sequences, but its functional importance has not been established. The xQ/NE, GDxN and SDH motifs have been assigned catalytic importance in *B. cereus* sphingomyelinase C *(Bc*-SMaseC) [283]. The proposed catalytic mechanism for the enzyme involves two or three metal ions, one bound to His-296 (SD**H**), which activates a water molecule, the nucleophile in SM hydrolysis, and another bound to Glu-53 (xQ/N**E**), which stabilizes a negatively charged transition state (see Fig. 3.3.5 A). Asp-195 and Asn-197 (G**D**x**N**) and Asp-295 (S**D**H) also participate in divalent metal ion binding at the active site. An aromatic-rich hydrophobic hairpin structure, which includes residues Trp-284 and Phe-285 of *Bc*SMase, was found to interact with the positive head group of 2-morpholinoethanesulfonic acid (MES), a structural mimic of the phosphocholine moiety of SM [283] (Fig 3.3.5 B). This hydrophobic patch is conserved in neither TcPanK nor human nSMases.

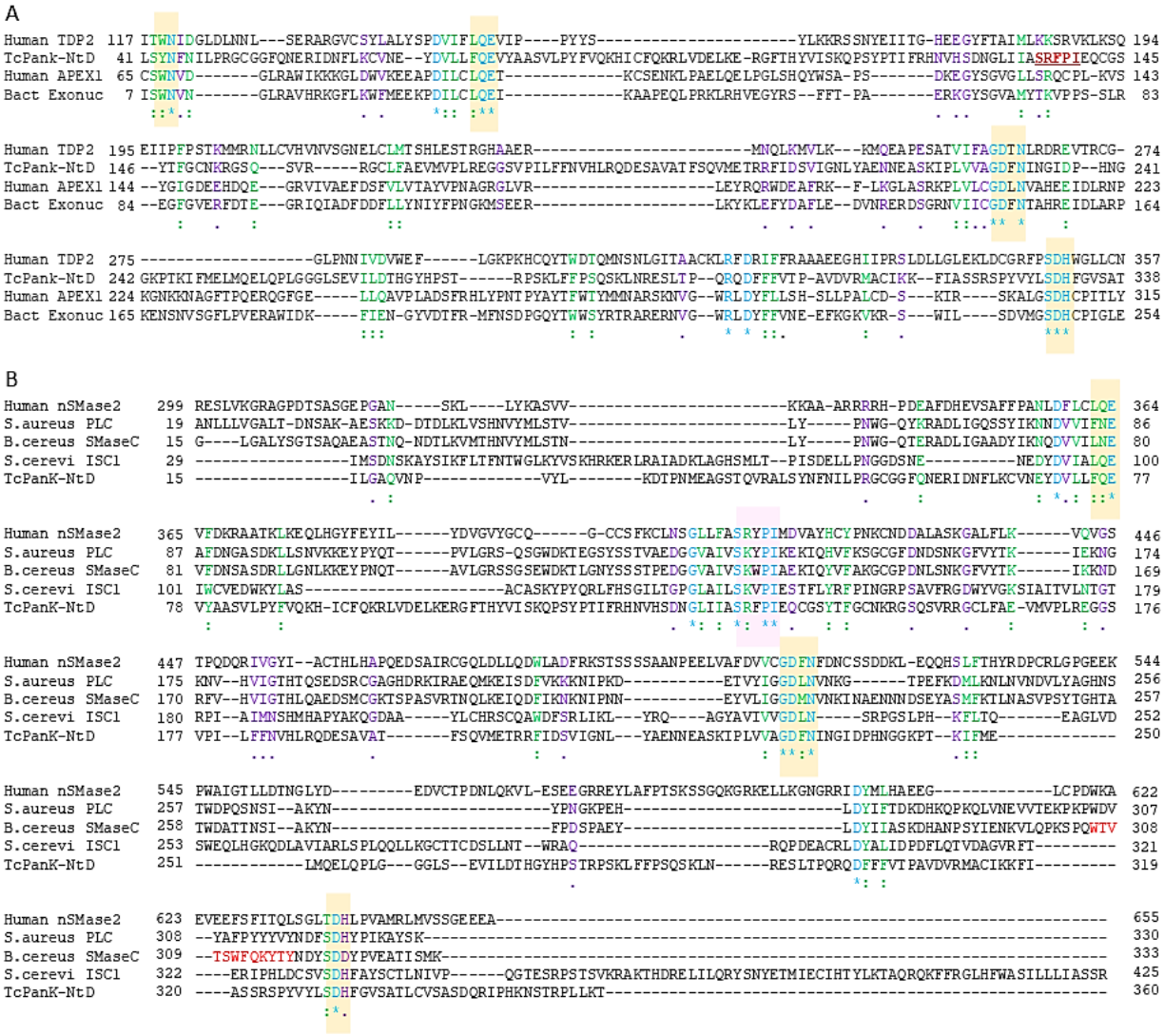


Fig. 3.3.4. Conserved motifs in EEP enzymes: (A) *A ClustalO alignment between the T. cruzi N-terminal domain (TcPank-NtD) and DNA-binding proteins of the EEP superfamily: human Tyrosyl-DNA phosphodiesterase 2 (TDP2, O95551) and APEX1 (P27695) and the thermophilic exonuclease III (Exonuc, 3FZI) of M. thermautotrophicus. Highlighted residues indicate conserved motifs associated with divalent metal ion binding. In TcPanK the SxxPI motif, which is characteristic of Smases, is coloured in red.* ***(B)*** *An alignment between TcPanK-NtD and neutral sphingomyelinases: human neutral sphingomyelinase 2 (nSMase 2, Q9NY59), S. aureus Phospholipase C (PLC, P09978), B. cereus sphingomyelinase C (P11889) and S. cerevisiae Inositol phosphosphingolipid phospholipase C (ISC1, P40015). Yellow highlighting indicates conservation within the endo/exonuclease/phosphatase superfamily and magenta highlighting indicates conservation within sphingomyelinases only. The hydrophobic hairpin of Bc-SMase is shown in red.*

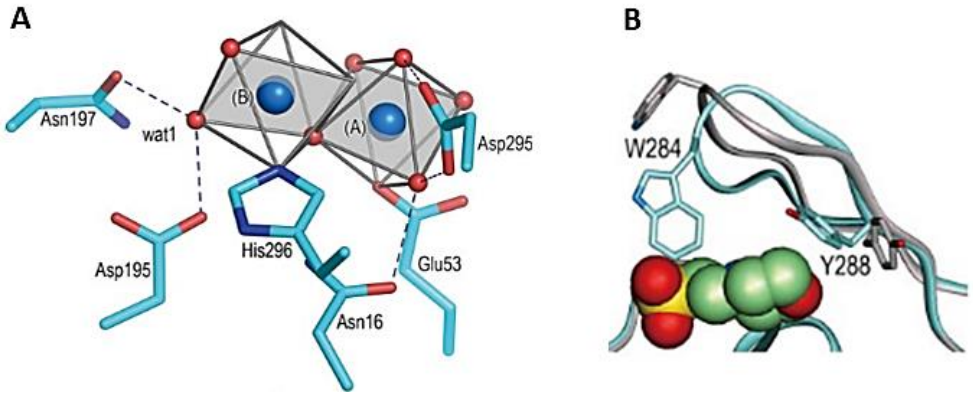


Fig. 3.3.5 Conserved residues in Bacillus cereus SMase crystal structure *(adapted from Ago et al., 2006): (**A**) The central metal-binding site of Cobalt-bound Bc-SMase. Bound cobalt ions and the coordinated waters are shown by blue and red spheres, respectively. Dashed lines represent the hydrogen bonds between ligand water molecules and amino acid residues. The amino acid ligands of the bound Co^{2+} ions are the carboxyl oxygen of Glu-53 (site A) and the imidazole nitrogen of His-296 (site B). (**B**) The β-hairpin of Bc-SMase twists in order to hold the bound substrate 2-morpholinoethanesulfonic acid (MES), shown as a space-filling model. The gray and cyan ribbons represent the hairpin of the free enzyme in cobalt and magnesium-bound enzymes, respectively. The positive head group of the MES molecule is a structural mimic of the phosphocholine moiety of sphingomyelin (SM).*

In trypanosomatids, the sphingolipid inositol-phosphoceramide (IPC) is more widely used than sphingomyelin. IPC hydrolysis is catalysed by inositolphosphosphingolipid phospholipase C (ISC), first characterised in fungi [284], generating ceramide and inositol phosphate. The *L. major* homologue ('inositol phosphosphingolipid phospholipase C-like' *Lm*ISCL) was found to hydrolyse sphingomyelin as well as IPC, which is not surprising due to structural similarities of IPC and SM [285]. The *T. brucei* orthologue is a nSMase with an even broader substrate range; it catabolises IPC, SM and ethanolamine-phosphoceramide sphingolipids as well as *lyso*-phosphatidylcholines (*lyso*-PCs) [286,287]. An orthologue is also found in the *T. cruzi* genome [288], but has yet to be characterised. Fig. 3.3.6 displays a ClustalO alignment between ISCs from *L. major, T. brucei* and *S. cerevisiae* and the first 660 residues of *Tc*PanK. In addition to the aforementioned 'metal-binding motifs', there appears to be conserved residues of ISC-type nSMases present in *Tc*PanK; Leu-41, Arg-59, Tyr-70, Asp-71, Phe-122, Gly-131, Ser-136, Pro-139, Phe-148, His-185, Gln-198, Ile-206, Gly-271, Leu-281, Pro-306 and Leu-339. However, the P-loop, which is characteristic of nucleotide-binding proteins and is present in ISC-type nSMases of *L. major* (His-114- Lys-121*), T. brucei* and *S. cerevisae,* does not appear in the N-terminal domain of *Tc*Pank.

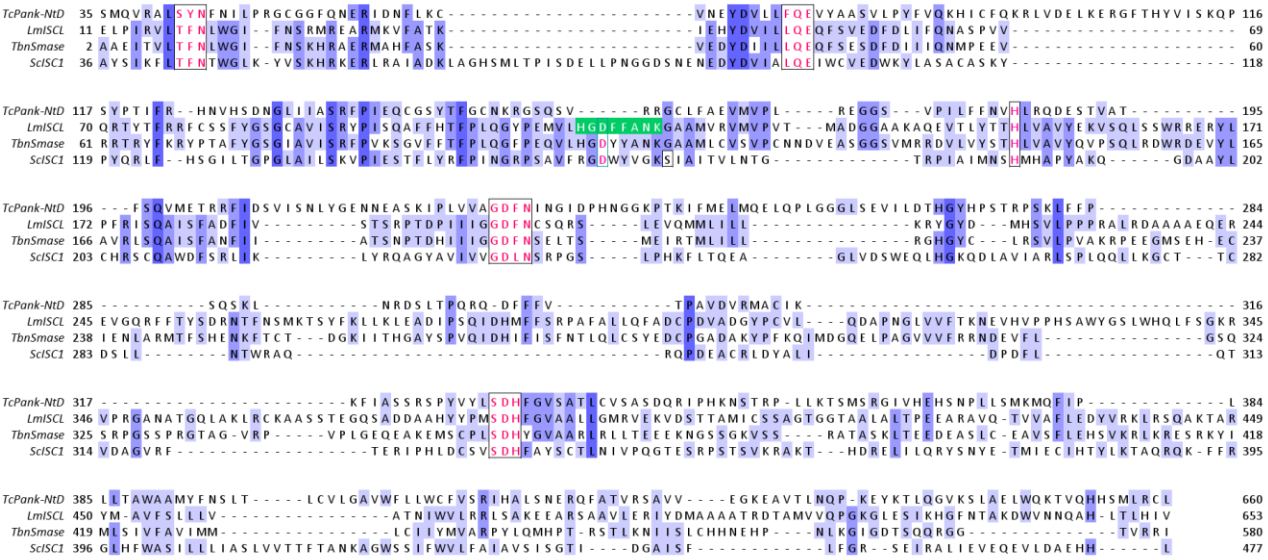


Fig. 3.3.6 Alignment of inositolphosphosphingolipid phospholipase C homologues. *Essential residues are shown in red, based on studies of B. cereus SMase, ScISC1p, and Hs-nSMase-1. The P loop domain is indicated in green in LmISCL. Shading is according to Fig. 3.2.3.*

Subsequently, the central domain (450-1040) of TcPanK was aligned with acyl-CoA synthetases to identify residues that according to the published literature are critical to their function. A representative alignment using sequences from *Dictyostelium discoideum* (slime mould), *Arachis hypogaea* (peanut plant), *C. elegans*, *Chlamydomonas reinhardtii* (alga) and humans revealed that the highly conserved DRKKxx or L-motif is absent from *Tc*PanK (Fig. 3.3.7). In LC-FACS sequences, the L-motif contains a peptide that acts as a linker between the large N-terminal domain and smaller C-terminal domain [289]. The IxxTSGTTGxPK motif which is associated with ATP binding is also missing from *Tc*PanK. Furthermore, a conserved stretch of residues starting with GExxxR/KGP and ending TGDI/VG (from Gly-483 to Gly-518 of the slime mould sequence Q1ZXQ4) is absent from *Tc*PanK.

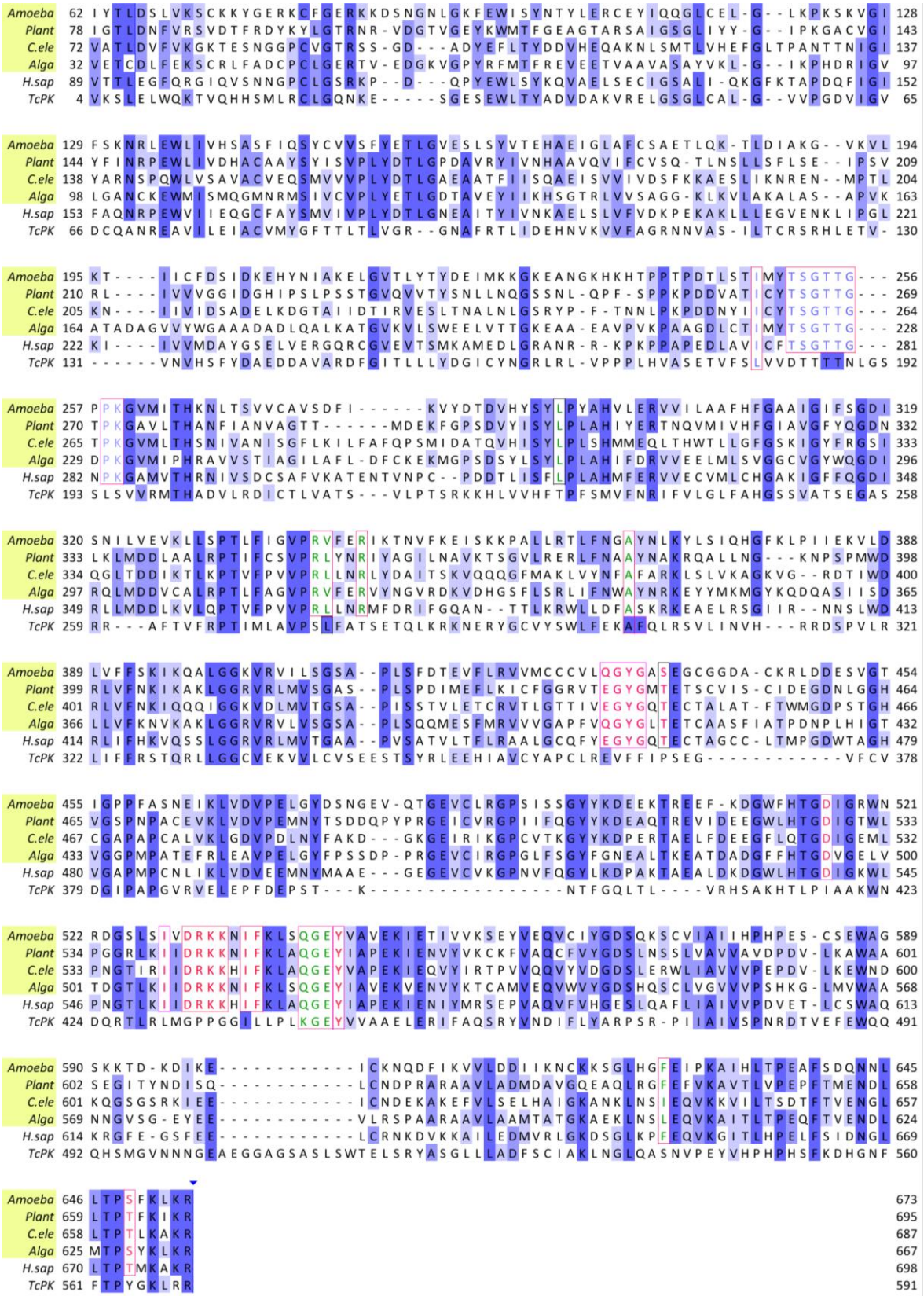


Fig. 3.3.7. TcPanK has limited sequence similarity with AMP-forming enzymes *The central domain of TcPanK (TcPK) aligned with fatty acyl-CoA synthetase B of D. discoideum (Amoeba, Q1ZXQ4), long chain acyl-CoA synthetase 6 of A. hypogaea (Plant, A0A0Y0AMC0), fatty acid CoA synthetase of C. elegans (C.ele, Q9XWD1), long chain acyl-CoA synthetase 1 of C. reinhardtii (Alga, A8JH58) and human long chain fatty acid CoA ligase 1 (H. sap, P33121). Conserved residues associated with AMP, CoA and ATP binding are shown in red, green and purple font, respectively. Shading as in Fig. 3.2.2.*

3.4 The predicted structure of *Tc*PanK

To support sequence-based analyses, structure models were obtained for *Tc*PanK using AlphaFold. This algorithm predicts a protein with three globular domains (Fig. 3.4). The accuracy of individual residues is estimated based on predicted local Distance Difference Test scores (pLDDT), whereby local Cα atomic interactions in the model are compared with a reference structure and used to score residues from 0-100. A score above 90 indicates high confidence in the model accuracy, 70-90 indicates a generally good backbone prediction, residues scoring 50-70 should be treated with caution and anything scoring below 50 should not be interpreted. Within each domain of *Tc*PanK, residues seem to be modelled well. At least 70% pLDDT values in each domain are higher than 70: 70% of the first (residues 38-340), 79% of the second (450-1040), 71% of the third domain (1074-1480). Another output of AlphaFold is the predicted aligned error (PAE). This is provided as a plot where the colour at (x, y) indicates AlphaFold's expected position error at residue x if the predicted and true structures were aligned on residue y. For *Tc*PanK, interdomain error is high (Fig. 3.4C), therefore confidence in the relative positions of each domain is low, which means that the likelihood of domain packing and interdomain interactions cannot be inferred accurately from this model.

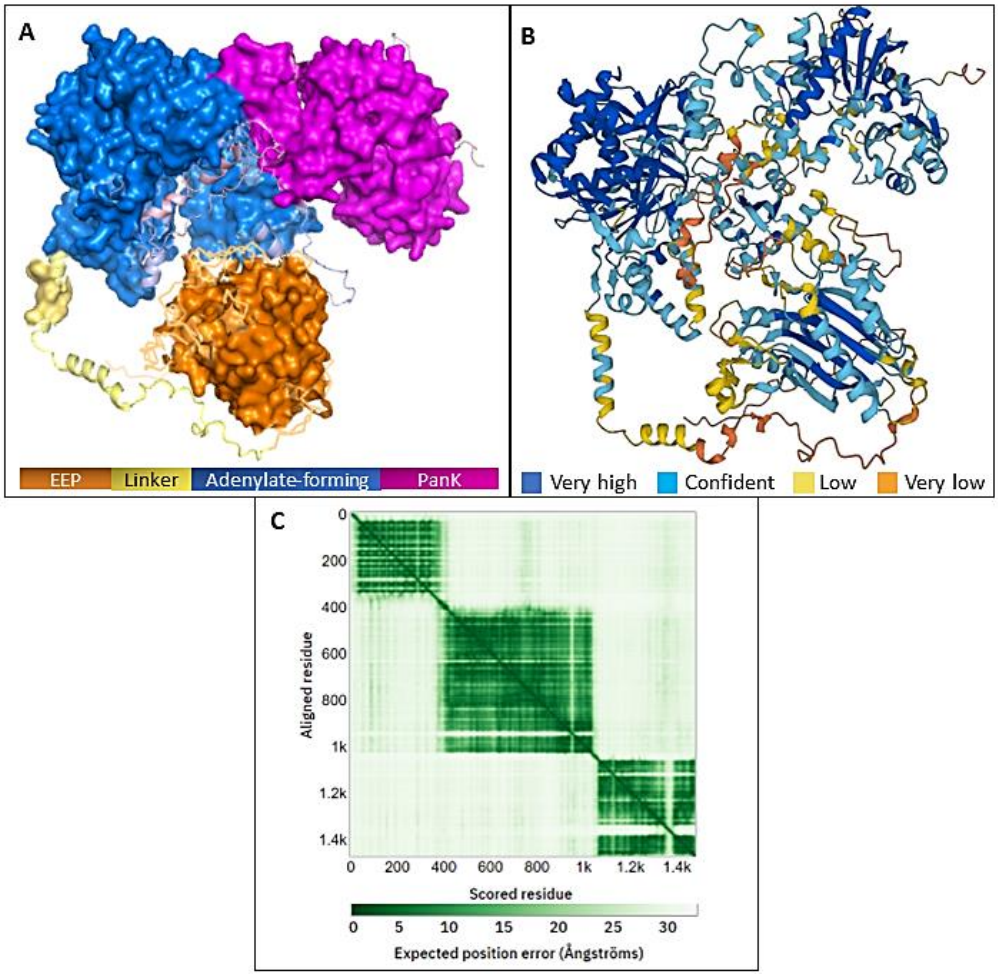


Fig. 3.4 AlphaFold predicts a multi-domain TcPanK. *AlphaFold predicted structure of TcPanK (Q4DHB2) A. Pymol surface representation with the putative EEP-family domain, linker region, putative adenylate-forming domain and PanK domain shown in orange, yellow, blue and magenta, respectively. Low confidence regions (pLDDT>70) are shown as pale ribbons.* ***B.*** *Ribbon representation coloured according to pLDDT confidence score, shown in the same orientation as in 3.4A.* ***C****. Predicted Aligned Error plot, whereby the colour at (x, y) indicates AlphaFold's expected position error at residue x if the predicted and true structures were aligned on residue y.*

3.5 Structure-based alignments also predict EEP and adenylate-forming activities

To predict function based on 3D structure, each subunit of the AlphaFold model was submitted to CO-FACTOR, which aligns a library of solved crystal structures to the query template, and based on homology, produces a list of structural homologues. These predictions further indicate that *Tc*PANK contains an EEP-type activity which is similar to sphingomyelinases, and an acyl-CoA synthetase-type function which is AMP-forming. Tables 3.5.1 and 3.5.2 show 10 proteins with the closest structural similarity to domains 1 and 2 of *Tc*Pank respectively, based on global structure alignments using the TM-align algorithm, which performs sequence-independent protein structure comparisons.

The domain 1 structural hits are characterised or predicted magnesium-dependent phosphodiesterases, the top 5 most similar being sphingomyelinases (Table 3.5.1). The domain 2 hits are adenylate forming domains, the most similar belonging to carboxylic acid reductases (CAR) and acyl CoA synthetases (Table 3.5.2).

Table 3.5.1. TcPanK domain 1 top structural hits

Colours indicate domains according to the protein map. EEP: endo/exonuclease/phosphatase

Protein Class	**Organism**	**PDB ID**	**TM**
Sphingomyelinase	*Bacillus cereus*	2uyr	0.795
Sphingomyelinase C	*Streptomyces griseocarneus*	3wcxA	0.786
Sphingomyelinase C (beta toxin)	*Staphylococcus aureus*	3i41A	0.779
Sphingomyelinase	*Listeria ivanovii*	1zwxA	0.767
Neutral sphingomyelinase 2	*Homo sapiens*	5uvgA	0.724
EEP-family Nuclease/deadenylase (CNOT6L)	*Homo sapiens*	3ngnA	0.681
Phosphodiesterase 12	*Homo sapiens*	4z0vA2	0.674
EEP family protein	*Bacteroides thetaiotaomicron*	3mprA	0.670
Tyrosyl-DNA phosphodiesterase 2	*Mus Musculus*	4gyzA	0.668
Tyrosyl-DNA phosphodiesterase 2	*Danio rerio*	4f1hA	0.667

Table 3.5.2. TcPanK domain 2 top structural hits

NRPS: Non-ribosomal peptide synthase, PDB and TM as in Table 3.5.1

Protein class	**Organism**	**PDB ID**	**TM**
Carboxylic acid reductase A domain	*Nocardia iowensis*	5msc	0.854
Carboxylic acid reductase A-PCP domain	*Segniliparus rugosus*	5mssA	0.722
Acetyl-CoA Synthetase	*Cryptococcus neoformans*	5ifiA	0.709
Acetyl-coenzyme A synthetase	*Saccharomyces cerevisiae*	1ry2A	0.679
Luciferase	*Luciola cruciata*	2d1rA	0.675
Acetoacetyl-CoA Synthetase	*Streptomyces lividans*	4wd1A	0.669
Anthranilate CoA ligase	*Stigmatella aurantiaca*	4wv3A	0.664
NRPS A domain	*Bacillus subtilis*	1mdbA	0.660
NRPS	*Bacillus stratosphericus*	6ulwA	0.646
NRPS A-PCP domain	*Microcystis aeruginosa*	4r0mB	0.642

Using the predicted *Tc*PanK 3D model and the most similar characterised structures for each domain based on the CO-FACTOR search, structure alignments were performed using Pymol. These are represented in Fig. 3.5.1. The N-terminal domain of *Tc*PanK (residues 38-340) was aligned to the crystal structure of a *Bacillus cereus* sphingomyelinase (UniprotKB code: 2uyr) (Fig. 3.5.1A). The putative adenylation domain of *Tc*PanK (residues 450-1040) was then aligned with a CAR structure of *Nocardia iowensis* (UniprotKB code: 5msc). CARs are multi-domain enzymes: their adenylation domains (A domains) are located between a large N-terminal reductase and smaller C-terminal PTP domain. To determine whether the reductase or PTP domains of structure 5msc aligned with either of the neighbouring domains of *Tc*PanK, the entire predicted *Tc*PANK was used in the CAR alignment. This indicated that only the A domain (residues 450-1040 of TcPANK) aligned with CARs. The CAR domains were delineated based on Gahloth et al., [290] .

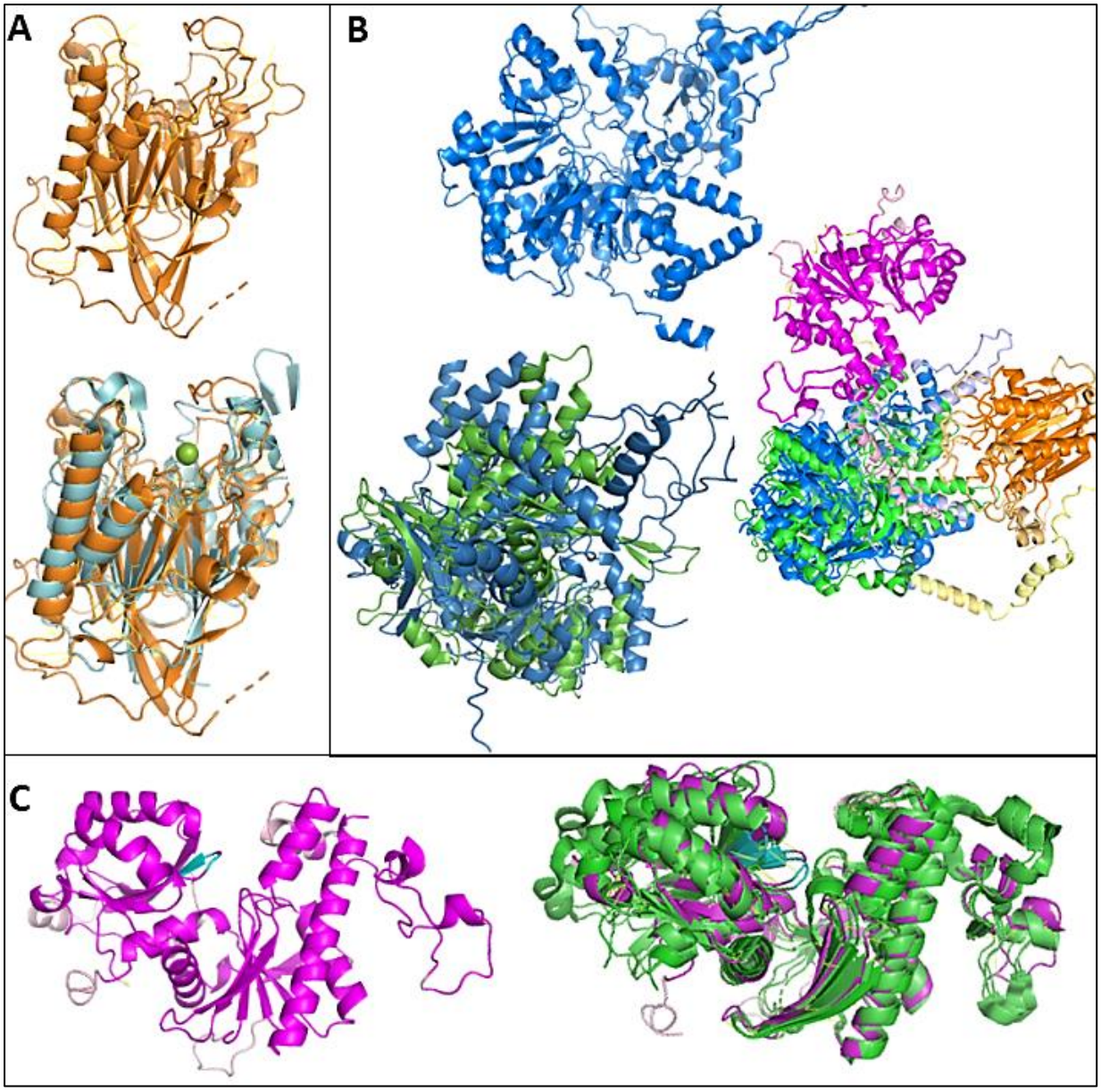


Fig. 3.5.1. Structural alignments of the TcPanK domains A. *Domain 1 (orange) aligned with sphingomyelinase (2uyr) of B. cereus (green).* ***B****. Domain 2 (blue) aligned with carboxylic acid reductase A domain (5msc) of N. iowensis (green).* ***C****. Domain 3 (magenta) aligned with human PanKs1-3 (green, Q8TE04, Q9BZ23, Q9H999)*

The programs COACH and CO-FACTOR were used to predict the location of ligand-binding sites within the 3D model using local structure alignment algorithms. The programs assign a confidence score (C score) for each ligand using a scale of 0-1, where a higher score indicates a more reliable prediction. The combined programs predicted the presence of magnesium and nucleic acid binding sites within domain 1, with a C score of 0.21 for both. AMP (C score=0.21), CoA (0.1) and phosphopantetheine (C score=0.05) sites were predicted for domain 2.

3.6 Summary

Bioinformatic analysis has revealed that CoA biosynthesis enzymes are present in trypanosomatid genomes, except for those required for *de novo* pantothenate synthesis. Uptake of extracellular pantothenate may occur via a major facilitator superfamily transporter, based on homology searches. Out of the identified proteins, PanK is deemed the most worthwhile candidate for characterisation in the context of drug discovery, based on features that appear unique to kinetoplastids. According to alignments performed at the level of amino acid sequence and predicted 3D structures:

- Kinetoplastid PanK is a multi-domain protein, consisting of three highly conserved activities, two of which are exclusively conserved amongst kinetoplastids.

- These additional activities are predicted to be associated with phosphodiesterase and adenylation domains, with the most similar proteins in the database being sphingomyelinases and fatty acyl CoA synthetases respectively.

4. Functional analysis of kinetoplastid PanKs using CRISPR-Cas9 and RNAi approaches

4.1 Introduction

CoA biosynthesis has been predicted or demonstrated to be essential in multiple pathogens, including bacteria, fungi and apicomplexa, using bioinformatic, genetic and biochemical methods. In the previous chapter, I identified the predicted enzymes of this pathway in kinetoplastid genomes and noted that the first enzyme, pantothenate kinase (PanK), is highly unusual in these organisms. This Chapter describes continued investigations of this enzyme by targeting the corresponding genes in *T. cruzi* and *T. brucei*, parasites that cause fatal diseases in humans and are a huge economic burden on afflicted countries. The findings from sequence analyses suggested that PanK may be expressed as a multi-domain protein. Here, I describe attempted verification of this by tagging the protein using CRISPR-spCas9 in *T. cruzi,* and by depletion of the protein in both organisms of interest using both CRISPR and RNA interference (RNAi) technology.

4.2 Is *Tc*PanK expressed as a single multi-domain protein?

First, I sought to tag PanK in *T. cruzi*, to determine both its subcellular location and to establish whether the three domains are expressed as a single protein, rather than undergoing post-translational processing to form multiple individual enzymes. A *T. cruzi* CL-Brener cell line was used, which had been engineered to express T7 RNA polymerase and SpCas9 nuclease, along with plasmids of the pPOT series for generating homology donor constructs by a simple PCR method [256,276]. Each terminus of the protein was targeted in separate transfections. Since *T. cruzi* CL-Brener is a diploid hybrid of two lineages (referred to as Esmeraldo-like and non-Esmeraldo-like) and therefore contains distinct alleles, both copies of *Tc*PanK were targeted in each transfection.

Cloning strategy

The homology donor constructs used for tagging consisted of a fluorescent protein or peptide epitope and an antibiotic resistance marker with the required RNA processing signals and flanking homology arms (Fig. 4.2.1). The fluorescence markers mNeonGreen or mScarlet, or the PTP epitope tag were inserted at the desired locus in addition to either puromycin or blasticidin resistance genes (puromycin N-acetyl-transferase or blasticidin deaminase) for

selection of transformed clones. Two of the constructs (mNeonGreen and PTP) also contained multiple Ty peptide repeats to visualise tagged *Tc*PanK using immunofluorescence with anti-Ty antibody.

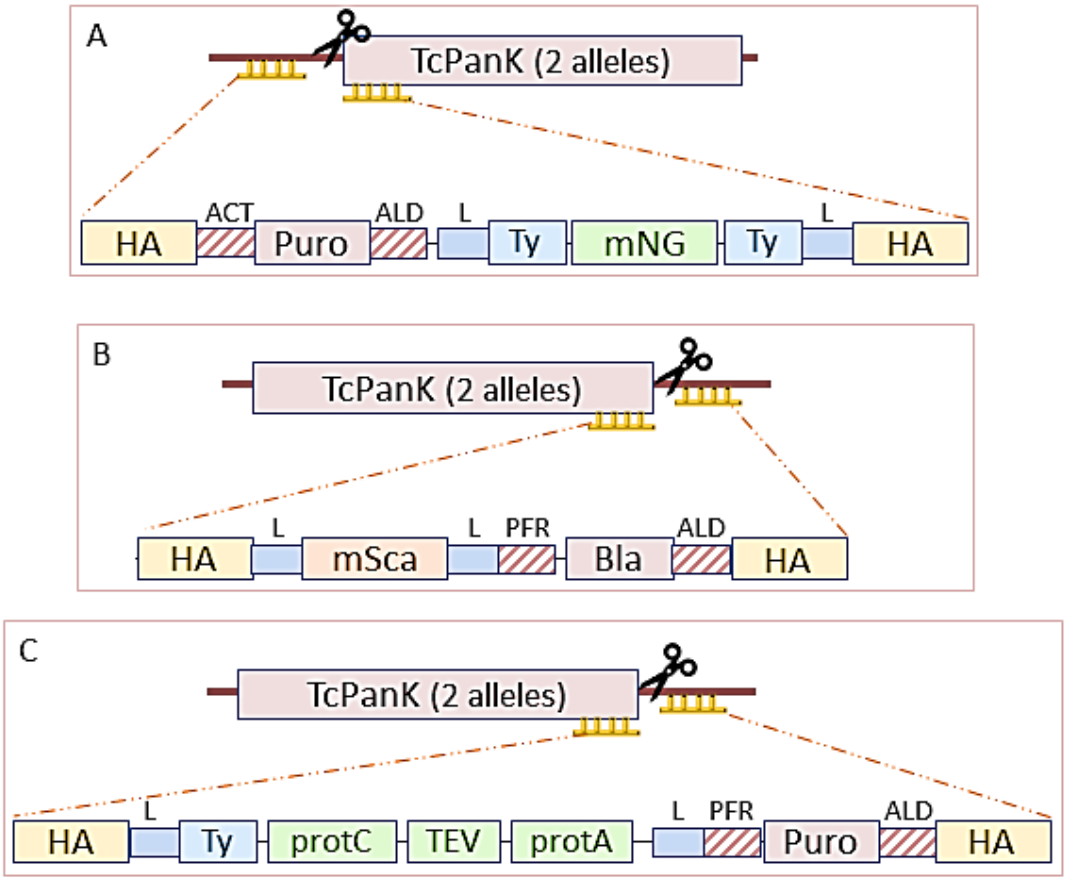


Fig. 4.2.1 CRISPR-Cas9 mediated tagging of TcPanK in TcCLB-T7Cas9 strain: *Representations of the TcPanK tagging method showing sgRNA binding sites (indicated by scissors), homology arms for integration (yellow combs) and maps of the homology donor constructs. (**A)** The N-terminus was targeted with an mNeonGreen (mNG) replacement cassette amplified from pPOTv6 puro-puro mNG. The C-terminus was targeted with **(B)** mScarlet (mSca) and **(C)** PTP (protein C, TEV cleavage site and protein A) homology donors amplified from pPOTv7 blast-blast mScarlet or pPOT-PTP respectively. To select for integration, constructs contained puromycin N-acetyl-transferase (Puro) or blasticidin deaminase (Bla) genes. HA: 30nt homology arms. RNA processing regions: ACT, actin; ALD, aldolase; PFR2, paraflagellar rod 2 intergenic sequence. Peptides: L, glycine-serine linker peptide; Ty, Ty repeats.*

A suitable donor plasmid containing the PTP tag was not available, therefore the plasmid for PTP tagging, named pPOT-PTP, was cloned as part of this project. This tag is useful for pulldown studies as well as for localisation and molecular size inferences. It is made up of protein A and C peptides, separated by a Tobacco Etch Virus (TEV) protease cleavage site, for two-step affinity purification of the tagged protein. The plasmid pC-PTP-NEO, developed by Gunzl et al., was used as a template (Appendix SF1) [273]. Primers which incorporated Hind-III (5') and Bsm-I (3') restriction sites were used to amplify the PTP sequence and insert this

into pPOTv6 puro-puro mNG (Appendix SF3) in place of the mNeonGreen tag (Materials and Methods, section 2.3.3; Appendix Table A.1, Fig. 4.2.2). To generate mScarlet or mNG-tagged constructs, plasmids pPOTv6 puro-puro mNG and pPOTv7 blast-blast mScarlet were used, which were a gift from Samuel Dean (Appendix SF3-4) [275].

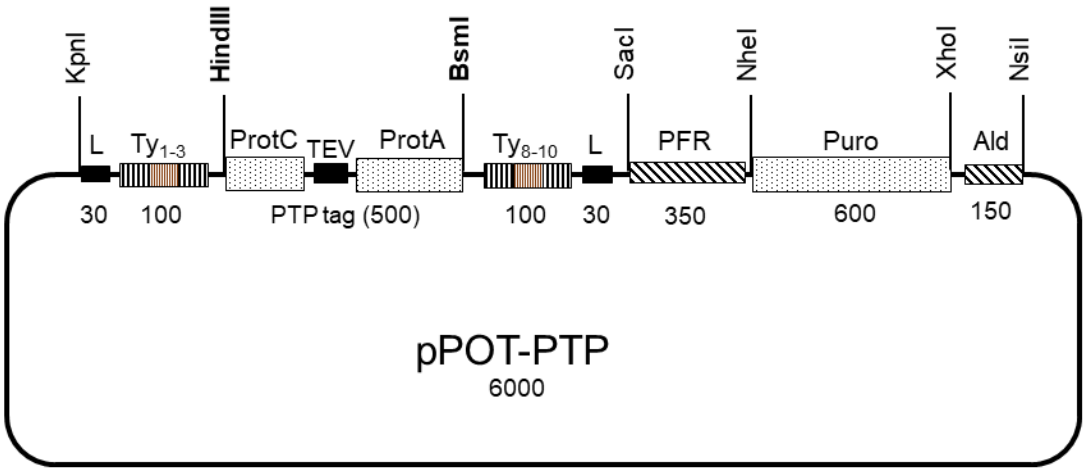


Fig. 4.2.2. pPOT-PTP: *For C-terminal tagging, the illustrated region of pPOT-PTP was targeted to the TcPanK 3' terminus. L: Glycine-serine linker, Ty: Ty epitope repeats, ProtC: protein C epitope (human hepatic plasma protein). TEV: tobacco etch virus (TEV) protease cleavage site, ProtA: IgG-binding unit of Staphylococcus aureus protein A, PFR2: Paraflagellar rod intergenic region, Puro: puromycin-N-acetyltransferase ORF, Ald: Aldolase 3' UTR. Approximate sizes are indicated in base pairs.*

To amplify targeting constructs, 30 bp homology arms - DNA fragments complementary to the genomic target sites for donor construct insertion - were incorporated into primers (Appendix Table A.2) and these primers were used to amplify the tag and resistance gene from the appropriate plasmid (Materials and Methods, section 2.3.4). Homology arm fragments were identical or near identical in both *Tc*PanK alleles, with up to two mismatches permitted.

To make single guide RNAs (sgRNAs), the online tool EuPaGDT (http://grna.ctegd.uga.edu) was used to locate a protospacer sequence within approximately 300 nucleotides of the desired integration sites [274]. A forward primer containing this protospacer, a T7 promoter and a partial RNA scaffold was combined with a reverse primer containing the complementary strand sgRNA scaffold in a second PCR reaction (Appendix Table A.2) [276]. The EuPaGDT tool also performs a BLAST search with the relevant genome to allow the selection of sgRNAs that have no secondary targets.

Generation and analysis of tagged cell lines

Epimastigotes of the *T. cruzi* CL-Brener T7-Cas9 cell line were transfected with the amplified sgRNA and homology donor constructs by means of electroporation (Materials and Methods, section 2.4.1). After a 24-hour recovery period, cultures were selected on blasticidin (10 µg/ml)

or puromycin (5 µg/ml). Within one month, parasites were growing well under the selective drugs and were subjected to molecular analyses to confirm integration of the tagging construct.

PCR and western blot analysis performed on isolated gDNA and protein indicated that *Tc*PanK tagging was successful (Fig. 4.2.3). To confirm integration in at least one cell line from each of the three transfections, PCRs were performed (Materials and Methods, section 2.6), amplifying across the gene-tag junction. Using one primer specific to *Tc*PanK and another to the tag ORF; mNG, Scarlet or PTP; expected product sizes of 1.2, 0.75 or 1.8 kbp, respectively, were obtained for each tagged cell line and were not amplified from parental gDNA (CLB T7-Cas9), demonstrating specificity (Fig. 4.2.3 A). Primer sequences can be found in Appendix Table A.4.

Western blots also indicated that tags had integrated at the *Tc*PanK locus, based on the visualisation of a ~170 kDa polypeptide in each transfected cell line, which is the expected mass of a PanK containing all three identified domains and the incorporated tag. For western blotting (Fig. 4.2.3 B) parasites were lysed in Laemmli buffer, then whole protein extracts were resolved by SDS-PAGE and transferred to nitrocellulose membranes (Materials and Methods, section 2.7). Detection was performed with antibodies reactive to mNeonGreen (1:2000), mCherry (1:2500) or Ty (1:1000) to detect the corresponding tags. HRP-conjugated secondary antibodies (1:5000) were applied for visualisation by chemiluminescence. In each case, a polypeptide of ~170 kDa could be distinguished in transfected cell lines, as well as fainter bands between ~50 and 100 kDa, which were in each case absent from parental cell lines. The smaller polypeptides observed are likely to be products of proteolytic or mechanical cleavage during the extraction process rather than endogenous cleavage of the protein since they are present in far lower abundance than the full-length protein.

These data are therefore consistent with *Tc*PanK being expressed as a large polypeptide containing 3 distinct functional domains.

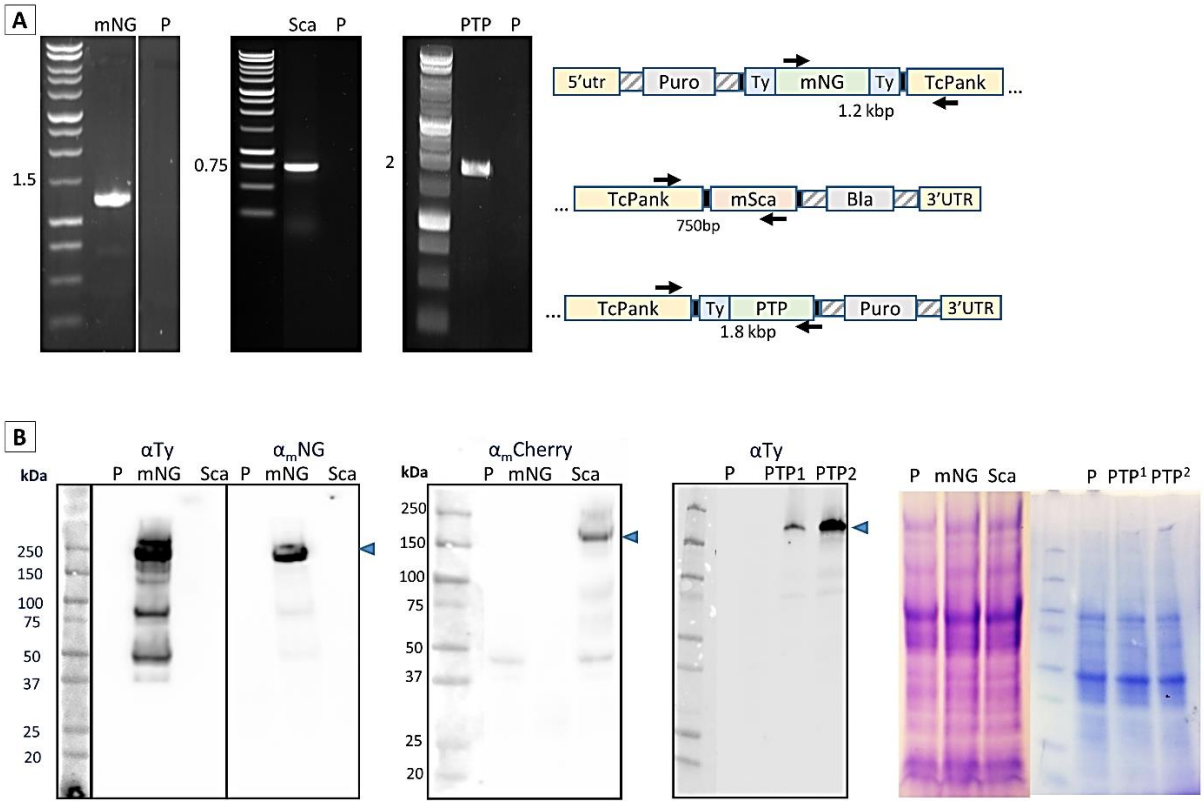


Fig. 4.2.3 Successful tagging of TcPanK mediated by CRISPR-Cas9 (A) *Expected PCR product sizes of 1.2, 0.75 and 1.8 kbp were obtained upon amplification across the Neon-TcPanK, TcPanK-Scarlet and TcPanK-PTP junctions respectively, demonstrating successful integration of tags. The gene maps indicate annealing sites of each primer pair.* ***(B)*** *Western Blots and Coomassie gels showing a ~170kDa polypeptide, which is absent from the parental cell sample, and is consistent with expression of mNeonGreen (mNG)-PanK, PanK-Scarlet and PanK-PTP. Loading of equal protein quantities in each lane was confirmed by Coomasie blue staining (bottom right). Arrowheads indicate expected size of tagged TcPanK. The original gel image of mNG amplification products is included in Appendix Fig. C.1.*

To localise *Tc*PanK, parasites were fixed in paraformaldehyde and imaged using an inverted Nikon Eclipse Ti epifluorescence microscope (Materials and Methods, section 2.9). Fluorescence was not conclusively detected in either of the fluorescently tagged cell lines, therefore immunofluorescence was performed on the PTP-tagged cell line containing intergenic Ty repeats. Fixed cells were incubated in mouse anti-Ty1 (1:500). The acquired images show fluorescent foci across the cytoplasm that are excluded from the nucleus (Fig. 4.2.4).

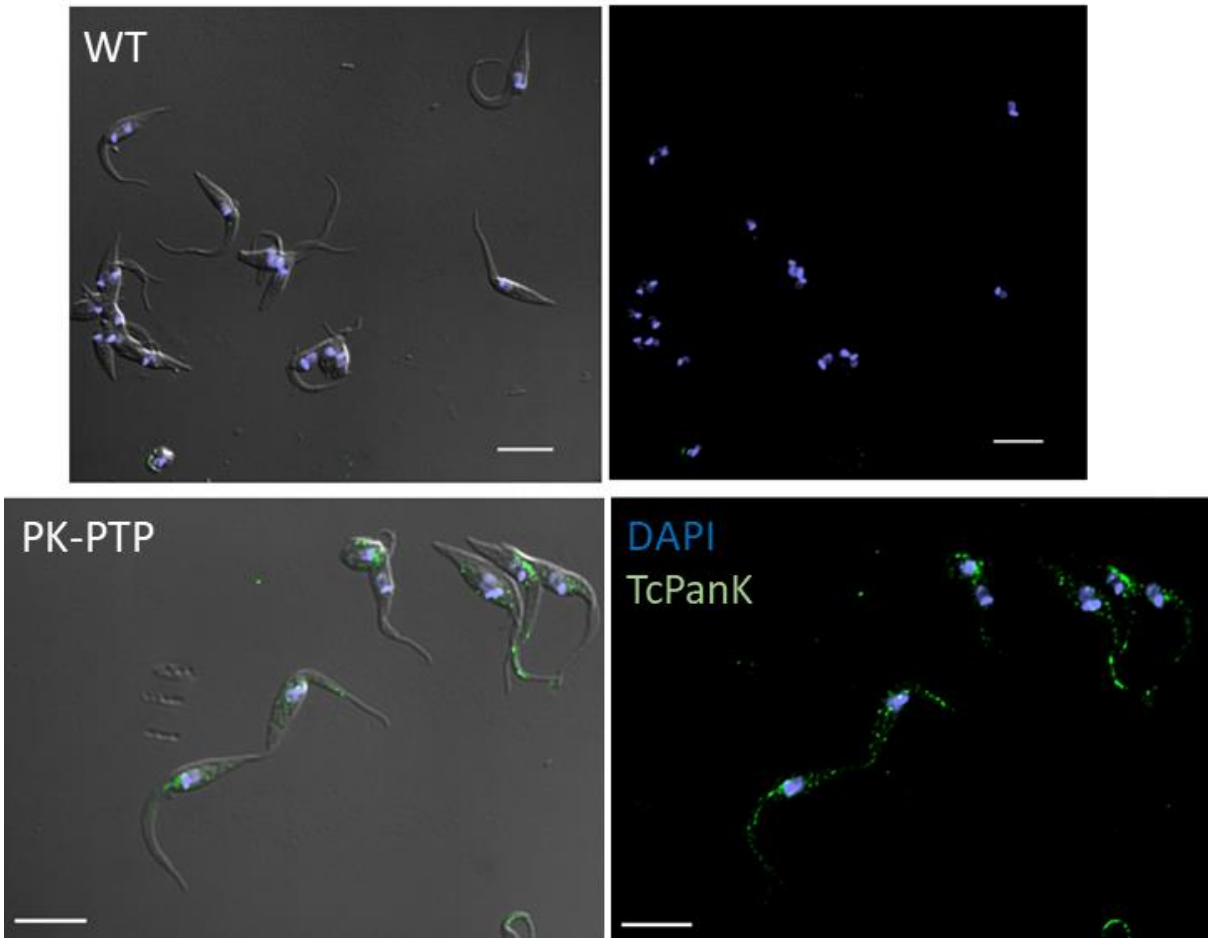


Fig. 4.2.4. TcPanK-PTP localisation *Immunofluorescence revealed punctate cytoplasmic staining in T. cruzi epimastigotes expressing Ty-PTP-tagged TcPanK detected with anti-Ty (1:500 [green]), which is absent from un-tagged (WT) cell lines. Merge of green signal and DAPI staining (blue) in differential interference contrast (left) and fluorescence overlay (right). Images acquired using an inverted Nikon Eclipse Ti epifluorescence microscope (Materials and Methods, section 2.9). Scale bar represents 10 µm.*

4.3 Disrupting the *Tc*PANK ORF in epimastigotes

To determine whether *Tc*PANK is essential in the CL Brener strain, we next used CRISPR-spCas9 to disrupt the ORF by replacing nucleotides 1-4286 of the coding sequence with an antibiotic resistance marker. The aim was to replace both alleles with blasticidin or puromycin resistance genes. This was attempted in three transfections by delivery of (1) blasticidin, (2) puromycin, or (3) blasticidin and puromycin genes. The sgRNAs and homology arms were designed to sequences that are identical in both *Tc*PanK alleles (Fig 4.3.1).

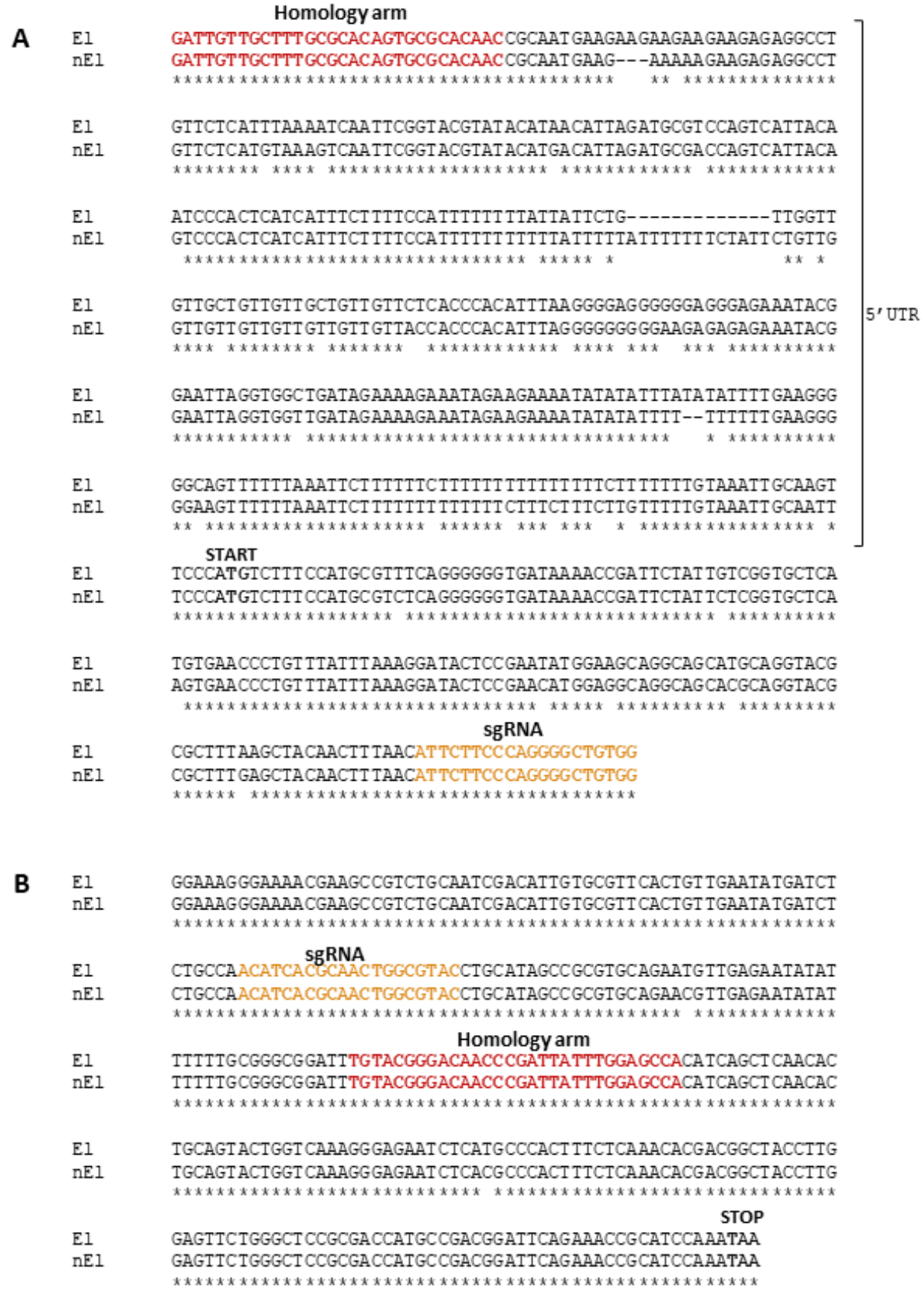


Fig. 4.3.1: Design of primers for TcPanK gene replacement using CRISPR SpCas9: *ClustalO-aligned T. cruzi* CL Brener *Esmeraldo-like (El) and non-Esmeraldo-like (nEl) TcPanK alleles, with sgRNA and homology arm sequences indicated as well as START and STOP codons. Partial sequences of 5' **(A)** and 3' **(B)** regions of the TcPanK ORF and 5'UTR (**A** only) are shown.*

The donor constructs for gene replacement were generated by the same PCR method and using the pPOT plasmids, as for tagging (see section 4.2), however two sgRNAs were used per transfection and the homology donor constructs contained only an antibiotic resistance gene and RNA processing signals (Fig 4.3.2A). See Materials and Methods, section 2.3.4 for PCR details and Appendix Table A.2 for primers.

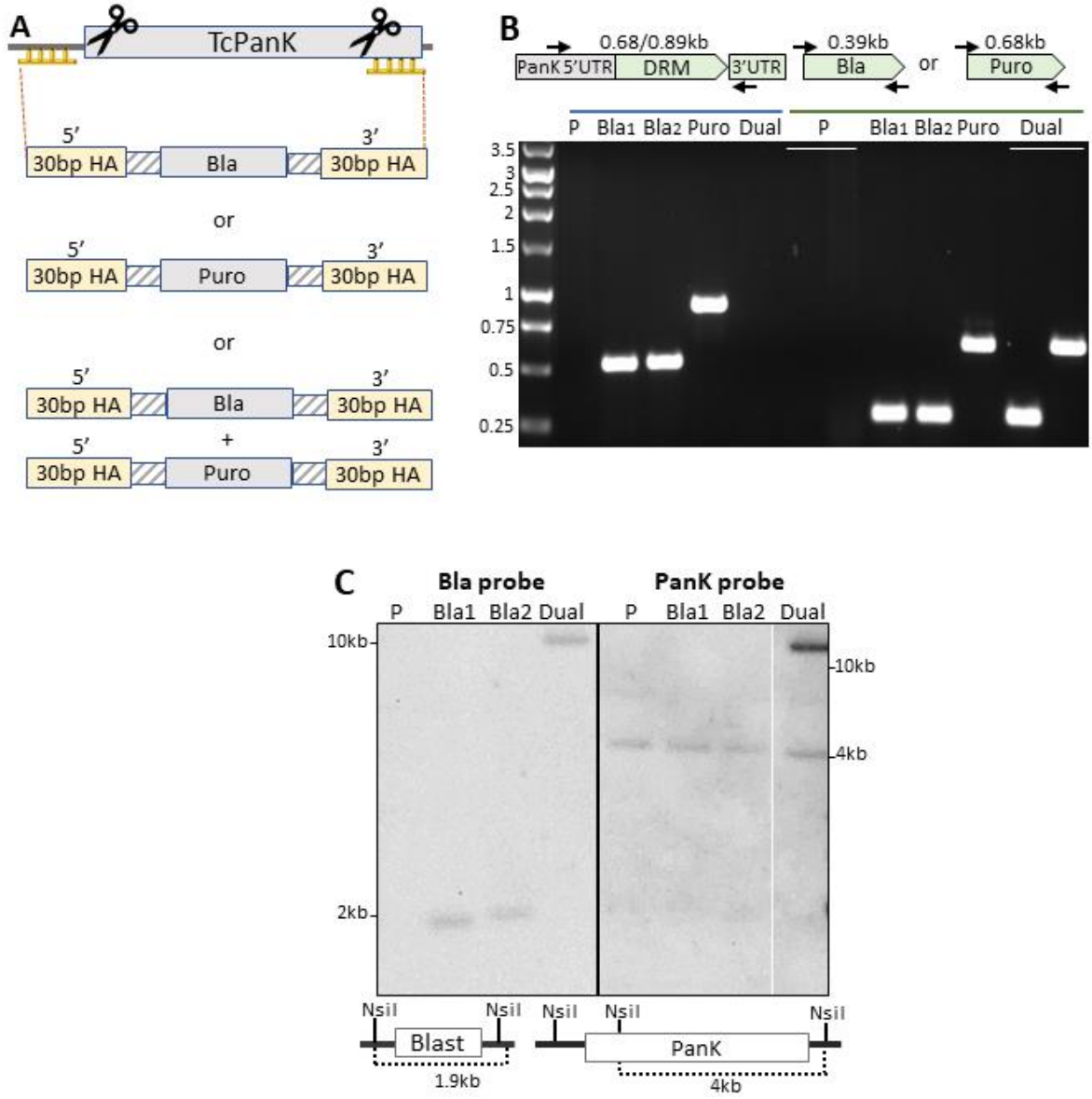


Fig. 4.3.2 Deletion of TcPanK by CRISPR-Cas9 (A) *Representative maps of TcPanK replacement cassettes and annealing sites for sgRNA (scissors) and homology arms (yellow combs) for each of three transfections performed. Donor constructs contained either Bla or Puro resistance genes to select for integration, flanked by RNA processing regions (striped boxes) and 30 nt homology arms targeting 360 nt upstream of the TcPanK start site (5' 30 bp HA) and 100 nt upstream of the TcPanK stop codon (3' 30 bp HA). RNA processing signals (striped boxes) of actin, aldolase or paraflagellar rod 2 intergenic sequence were used.* ***(B)*** *PCRs using primers targeting the 5'UTR of TcPanK and 3'UTR of the drug resistance marker (DRM) of the gene replacement cassette amplified products of 0.68 kb (Bla, flanking), 0.89 kb (Puro, flanking), 0.39 kb (Bla, internal) and 0.68kb (Puro, internal). Primer annealing sites and expected product sizes are indicated. Template gDNA was isolated from the following cell lines: P, parental; Bla1/2, Bla-replacement clone 1/2; Puro, Puro-replacement; Dual, Bla- and Puro-replacement. Ladder sizes are shown in kb.* ***(C)*** *Southern blots produced by NsiI-digestion and Bla- or PanK-probing of gDNA from Bla- or dual drug-resistance gene-replacement (Bla1 and 2, Dual) and parental (P) cell lines. Expected product sizes are indicated on the restriction maps. Bla: blasticidin deaminase; Puro: puromycin N-acetyltransferase. See Appendix Fig. C.2 for original Southern blot images.*

After 3-4 weeks, transfected parasites were evidently proliferating. Molecular analysis (Materials and Methods, sections 2.6 and 2.8) of the generated cell lines revealed that only a single allele was successfully deleted in each case. PCR analysis indicated that the antibiotic

resistance marker successfully replaced *TcPanK* in at least one allele, as products of the correct size were amplified by primers annealing to the flanking regions and the antibiotic resistance genes (Fig. 4.3.2 B, Appendix Table A.5). Southern blotting of NsiI-digested gDNA using a probe specific to the blasticidin resistance gene (blastR) revealed a product of approximately 2 kb in both blastR knock-in cell lines, which is the expected fragment size for NsiI digestion (Fig 4.3.2 C). The attempted deletion of both *TcPanK* copies was unsuccessful based on both analytic methods. Despite the fact that the resistance genes themselves could be amplified by PCR in all cell lines, primers targeting the UTRs of the dual knock-in sample did not amplify the expected product. Southern blotting showed a high molecular weight band (~10 kbp) with both blastR and PanK probing which was inconsistent with correct integration at the target locus. Instead it suggested that a rearrangement has occurred around the *TcPanK* allele. Southern blot banding patterns from *TcPanK* probing indicated that all lines still contained an unmodified allele.

An inability to generate null mutants suggests that the *TcPanK* gene is essential for viability in *T. cruzi* epimastigotes, although it does not represent unequivocal proof.

4.4 Inducible depletion of *Tb*PanK using RNAi and complementation with *TcPanK*

TbPanK depletion

As an alternative to *TcPanK* gene replacement by homologous recombination, we sought to induce PanK depletion in the related parasite *T. brucei* using RNA interference (RNAi). *T. cruzi* lacks the cellular machinery for RNAi and there have been no RNA level knockdown methods (e.g. Cas13) shown to work in this parasite. Fortunately, as well as being a pathogen of interest - responsible for sleeping sickness and nagana - *T. brucei* can be used as a tool for *T. cruzi* gene investigations thanks to add-back methods. For instance, it was previously shown that RNAi-mediated depletion of topoisomerase-IIa in bloodstream form (BSF) *T. brucei* is lethal and this lethality can be rescued by expression of the corresponding enzyme from *T. cruzi* [291]. Therefore, as well as determining essentiality of PanK in these cell lines, we also tested the functional homology of the *T. cruzi* pantothenate kinase orthologue.

T. brucei PanK (*TbPanK*) was targeted in the 2T1 BSF cell line, developed by Dr Sam Alsford, which constitutively expresses the tetracycline repressor protein (TetR), and has a partial hygromycin-resistance (*hyg*) gene inserted at a ribosomal RNA (RRNA) locus on chromosome 2a [254]. The targeting construct (pRPa^{Sli}) consisted of a stem-loop containing two inverted

copies of a 400 bp fragment of *TbPanK* (Lister 427 strain) and a 468 bp LacZ fragment, under the control of a tetracycline-regulatable ribosomal RNA (RRNA) promoter (Fig. 4.4.1). For integration into the tagged RRNA locus, pRPa^{Sli} contains a partial *hyg* fragment (5' fragment) and RRNA spacer sequences, designed to generate an intact *hyg* open-reading frame (ORF) upon integration, whilst deleting a puromycin N-acetyltransferase ORF.

Successful integration is indicated by hygromycin resistance and the restoration of puromycin sensitivity, resulting from constitutive *hyg* transcription driven by a procyclin promoter (Fig. 4.4.1). In the absence of tetracycline, transcription of the stem-loop is repressed by TetR binding to the Tet operator. Tetracycline administration causes conformational changes in TetR and lifts repression, resulting in expression of the stem-loop. The enzyme *Tb*Dcl-1 (Dicer-like 1) processes the RNA hairpin, generating siRNAs that trigger *Tb*PanK transcript depletion.

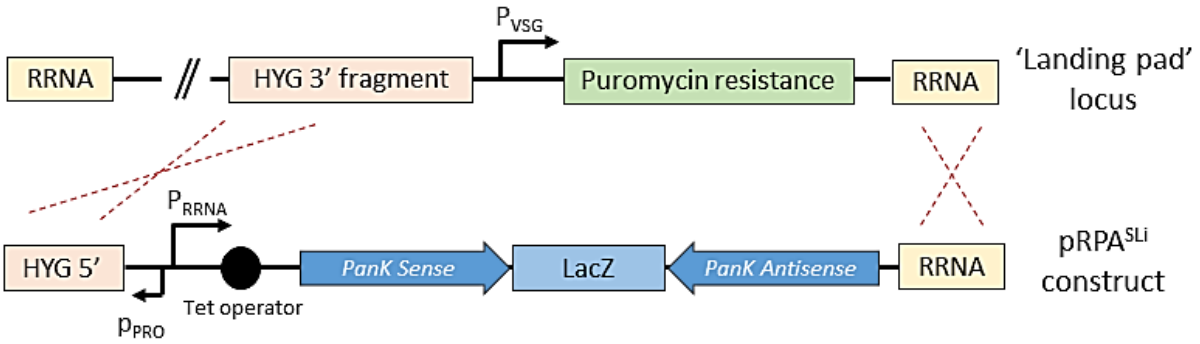


Fig. 4.4.1. pRPa^{Sli} stem-loop cassette for RNAi knockdown of TbPanK*: Hygromycin (Hyg 5') and ribosomal RNA (RRNA) spacer sequences target the stem-loop cassette to the tagged RRNA 'landing pad' locus of 2T1 cells, which consists of a partial (3') hygromycin resistance gene fragment and a variant surface glycoprotein expression-site promoter (P_{VSG}) driving constitutive expression of a puromycin N-acetyltransferase gene. Integration of pRPA^{Sli} creates a complete hygromycin resistance gene and deletes the puromycin resistance ORF. Tetracycline-induced expression of the stem-loop is driven by the RRNA promoter (P_{RNA}) when integrated in target cells. The stem-loop hairpin dsRNA is a substrate for TbDcl-1 which processes the double-stranded region into short interfering RNAs (siRNAs). The siRNAs bind to TbArgonaute and trigger PanK transcript degradation. P_{PRO}: procyclin promoter.*

To prepare the targeting construct, two complementary RNAi fragments corresponding to nucleotides 1028 to 1436 of *TbPanK* were inserted in opposite orientations into pRPai in a 2-step cloning process, using a single primer pair (Fig. 4.4.2). Each primer contained two restriction sites: internal sites (KpnI and BamHI) were used for cloning in a sense orientation into MCS1 and then external restriction sites (ApaI and XbaI) were used to integrate an antisense fragment into MCS2.

The linearised construct was transfected into 2T1 cells by electroporation (Materials and Methods, section 2.4.2). Hygromycin-resistant parasites were identified after 5 days and then clones were screened for puromycin sensitivity. Only clones displaying hygromycin resistance and puromycin sensitivity were taken forward for further analysis.

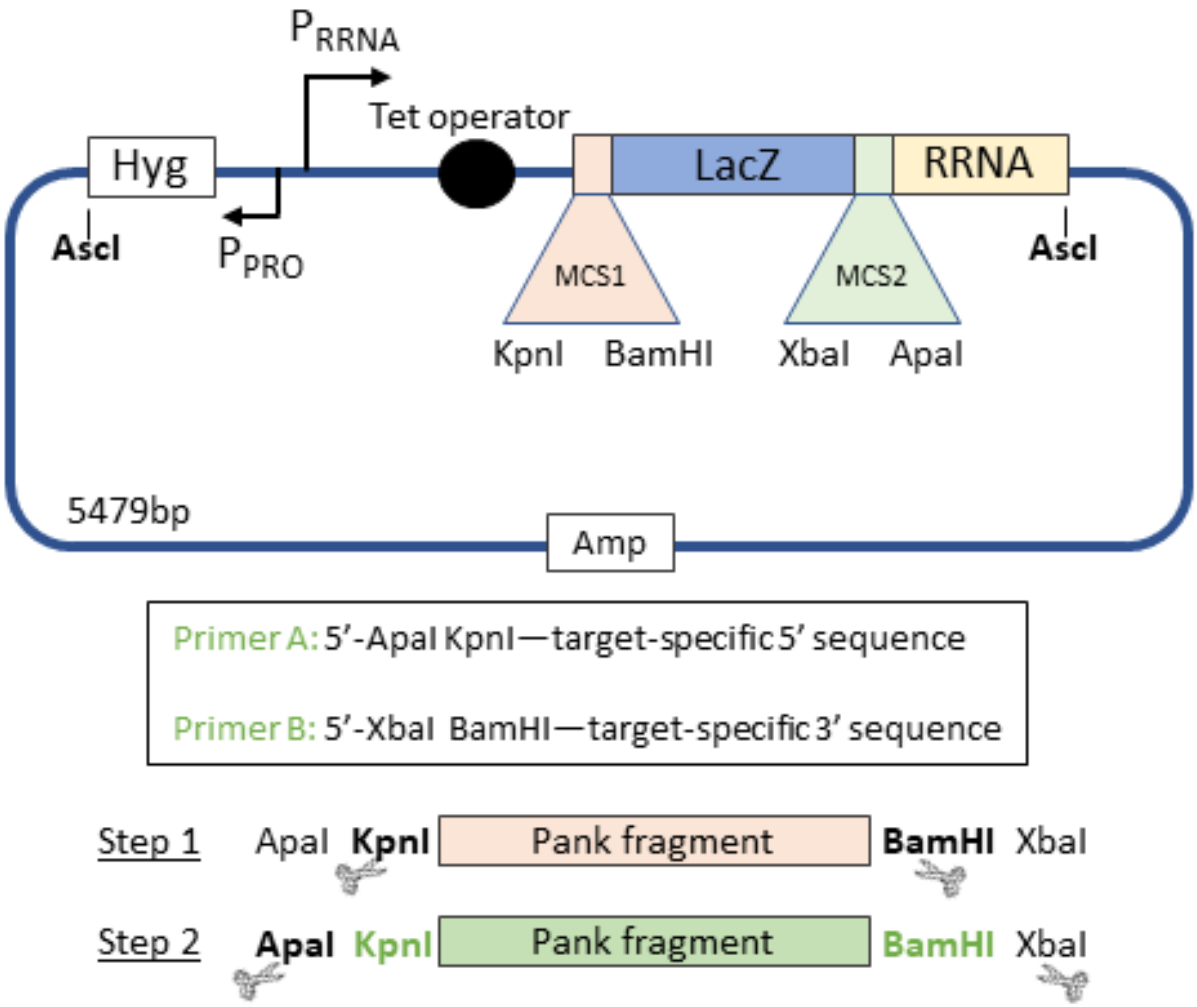
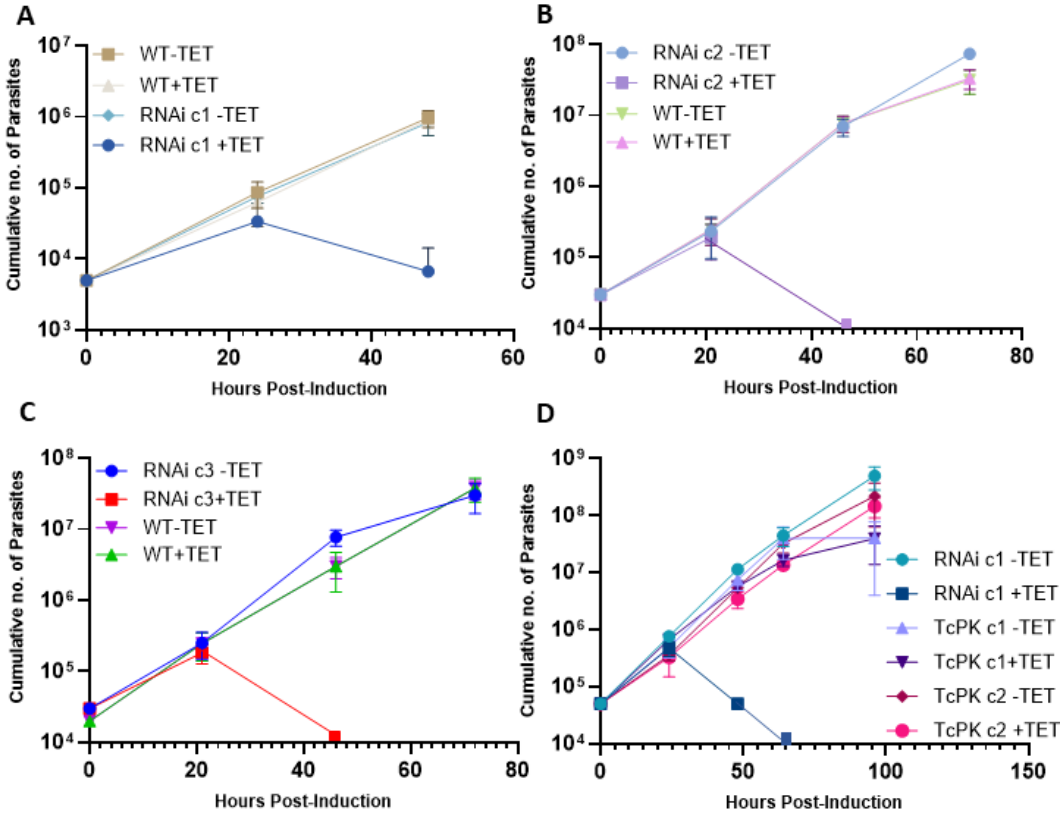


Fig. 4.4.2 pRPa^{Sli} cloning method: *A primer pair containing two restriction sites per primer was used to insert complementary TbPanK fragments in opposite directions into pRPai multiple cloning sites 1 and 2 (MCS1 and MCS2). Internal sites (KpnI and BamHI) allowed cloning in a sense orientation into MCS1 and external sites (ApaI and XbaI) were used for antisense fragment integration into MCS2. Hyg: partial hygromycin resistance gene fragment for locus integration, P_{PRO}: procyclin promoter, P_{RRNA}: ribosomal RNA promoter, RRNA: ribosomal RNA spacer for integration, Amp: Ampicillin resistance gene. Scissors indicate restriction digestion.*

To measure the effect of PanK knockdown on parasite growth and viability, growth curves were performed using transfected cell lines (Fig. 4.4.3 A). This involved seeding RNAi cell lines containing the *Tb*PanK depletion construct and parental cell lines in 24-well plates, with or without tetracycline (1 µg/ml) (Materials and Methods, section 2.10). Experimental groups were seeded in triplicate at an equal density (typically 5 x 10^4 parasites in 1 ml). Cumulative cell density was then measured every 24 hours, for a minimum of three consecutive time-points. Cells were diluted to 1 x 10^5 as required in order to maintain exponential growth. As Fig. 4.4.3 shows, there was a severe growth phenotype in induced RNAi cell lines (RNAi + TET). While untreated RNAi cell lines grew at the same rate as parental cells (WT), induced parasites had a clear proliferation defect within 24 hours of induction, and cell death occurred

within 48 hours. By this time, dead cells were readily observed in treated wells upon visual inspection under a light microscope. To verify that the tetracycline itself caused no toxicity, growth of the parental cell line in the presence of the inducer (WT + TET) was monitored in parallel. Results for three RNAi clones (RNAi c1-c3) are shown and are representative of several repeated growth curves (min n=2 per clone), which consistently displayed lethality of *Tb*PanK depletion within 48 hours of treatment.

To confirm at the molecular level that *Tb*PanK was targeted by RNAi in the induced cell lines, a loss of endogenous *Tb*Pank mRNA was quantified by qPCR (Fig. 4.4.3 E). RNA was collected from each experimental group at 16 hours post-induction, prior to the death of induced RNAi cell lines. To minimise experimental variability, each cell line was maintained at an equal density for at least 24 hours prior to and during tetracycline treatment. Untreated groups were grown at equal density to those receiving treatment. RNA was isolated from parasites using the Rneasy mini kit (Qiagen) and treated with Dnase prior to cDNA synthesis (Materials and Methods, section 2.11). cDNA synthesis reactions were performed in triplicate per RNA sample using 1 µg of total RNA and the SuperScript VILO cDNA Synthesis Kit (Thermofisher). An equal volume of each cDNA sample (cDNA 1-3) was then amplified in triplicate per primer pair in qPCR reactions containing SYBR Green master mix (Qiagen). Fragments of 108-112 bp were amplified using primers specific to *TbPanK* and telomerase reverse transcriptase (TeRT) The abundance of *Tb*PanK was calculated relative to the endogenous control TeRT using the $\Delta\Delta$Ct method [281]. The mean Ct values calculated from triplicate cDNA samples were similar between clones of each experimental group and calculated $2^{-\Delta\Delta Ct}$ values confirm that *Tb*PanK downregulation occurred in the tetracycline-induced RNAi cell lines, which contained on average 93% less *Tb*PanK transcript than the uninduced groups.



Group	TbPk Ct	TeRT Ct	ΔCT TbPk	ΔΔCT TbPk	$2^{-(\Delta\Delta Ct)}$ TbPk
RNAi c1	23.60	20.62	2.98	0.17	0.89
RNAi c2	23.18	20.54	2.64	-0.17	1.13
RNAi+T c1	29.85	22.42	7.43	4.62	0.04
RNAi+T c2	27.95	21.77	6.18	3.37	0.10
TcPK c1	26.02	21.86	4.15	0.08	0.95
TcPK c2	26.01	22.01	3.99	-0.08	1.06
TcPK+T c1	29.00	21.90	7.10	3.03	0.12
TcPK+T c2	28.96	22.21	6.75	2.68	0.16

Group	TcPk Ct	TeRT Ct	ΔCT TcPK	ΔCT TbPK	ΔΔCT TcPK	$2^{-(\Delta\Delta Ct)}$ TcPK
RNAi c1	35.36	20.62	14.73	2.98	11.93	0
RNAi c2	31.67	20.54	11.13	2.64	8.33	0.003
RNAi+T c1	32.91	22.42	10.49		7.68	0.005
RNAi+T c2	31.71	21.77	9.94		7.13	0.007
TcPK c1	22.09	21.86	0.23	4.15	-3.84	14.37
TcPK c2	21.18	22.01	-0.83	3.99	-4.90	29.92
TcPK+T c1	21.93	21.90	0.03		-4.04	16.46
TcPK+T c2	21.65	22.21	-0.55		-4.63	24.71

Fig. 4.4.3 RNAi knockdown of TbPanK causes cell death which is rescued by TcPanK expression: (A-D) *Tetracycline (1 µg/ml) addition to BSF T. brucei cells containing TbPanK-knockdown constructs (represented by RNAi c1-c3+TET) causes cell death* ***(A-C)****, whereas it has no effect on parental cells (WT+TET) or those expressing TcPanK in addition to the TbPanK-depletion construct (TcPK c1-2+TET,* ***D****). For comparison, experimental groups were included that did not receive tetracycline, either containing the RNAi construct only (RNAi -TET), both RNAi and TcPanK constructs (TcPanK c1-2 -TET), or lacking either construct (WT -TET). Cells counts were performed in triplicate per group at 24 hour intervals (mean values shown).* ***(E and F)*** *qPCR data for RNA extracted from parental (WT), TbPank RNAi (RNAi) and TcPanK add-back (TcPK) clones (16 hours post-induction) showing depletion of TbPanK (TbPK) transcript with tetracycline treatment (+, +TET) relative to untreated groups (-, -TET), using telomerase reverse transcriptase (TeRT) as a house-keeping gene and presence of TcPanK (TcPK) in add-back cells, shown as TcPK abundance relative to TbPK in uninduced lines.* ***(F)*** *Bars indicate mean 2^-(ΔΔCt) and standard error for each experimental group, calculated from triplicate cDNA samples from 2 experimental clones. Tables present mean Ct per clone.*

TcPanK Add-back

To provide additional evidence that the observed tetracycline-dependent growth phenotype was specifically due to depletion of the *Tb*PANK transcript, a complementation or add-back strategy was employed. This entailed inserting the *Tc*PanK ORF into the tubulin gene array of the *Tb*PanK RNAi cell line (RNAi c1) for constitutive expression. Firstly, PanK sequences from both trypanosome species were aligned to ensure that the RNAi construct would not target the *T. cruzi* transcript. The two genes were 59% identical at the nucleotide level and lacked conserved stretches >20 bp, suggesting that off-target silencing of *TcPanK* was unlikely to occur (Appendix D.1). We amplified the 4443 nucleotide *TcPANK* gene from *T. cruzi* CL Brener gDNA using primers which incorporated flanking SbfI and AscI restriction sites. Sequencing revealed that a chimeric *Tc*PanK sequence was amplified, containing a mixture of both alleles (Fig. 4.4.4). See Appendix D.2 for an alignment of both *Tc*PanK alleles and the cloned gene.

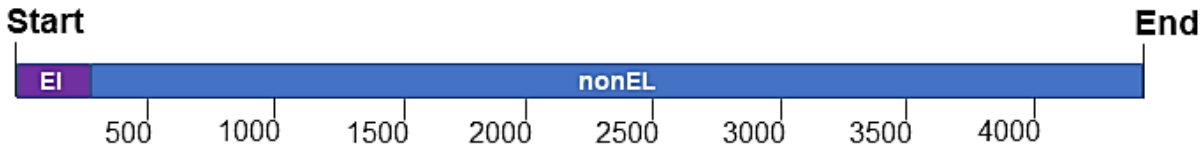


Fig. 4.4.4: Chimeric amplification of TcPanK*: The TcPanK ORF amplified from T. cruzi CL Brener gDNA consisted of both Esmeraldo-like (EL, nucleotides 1-279) and non Esmeraldo-like (nonEL, nucleotides 280-4443) alleles.*

The *Tc*PanK ORF was cloned into plasmid pTubEX, generating construct pTub-*Tc*PanK^{WT} (Material and Methods, section 2.3.5, Appendix SF6-7). Digestion with Not1 and Xho1 released a linear targeting fragment containing tubulin intergenic regions, an actin RNA processing signal and blasticidin resistance gene (Fig. 4.4.5).

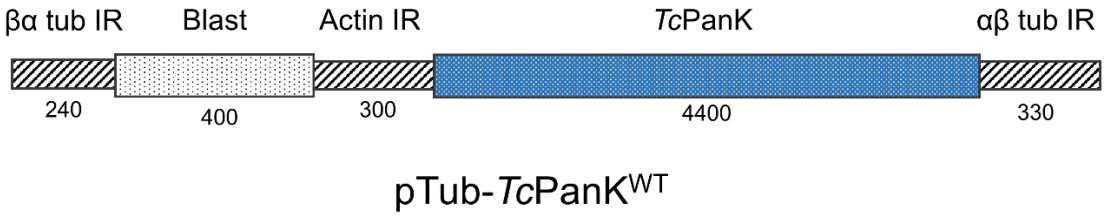


Fig. 4.4.5: Linearised pTub-TcPanK^{WT}*: Map of the linear tubulin array targeting construct pTubTcPanK^{WT} for constitutive expression of T. cruzi PanK. The TcPanK ORF is flanked by two intergenic β-α tubulin (βα tub IR, 240bp) and α-β tubulin (αβ tub IR, 330bp) sequences, a blasticidin deaminase gene (Blast) for selection of transformants and an actin intergenic region (actin) that provides intergenic RNA processing signals.*

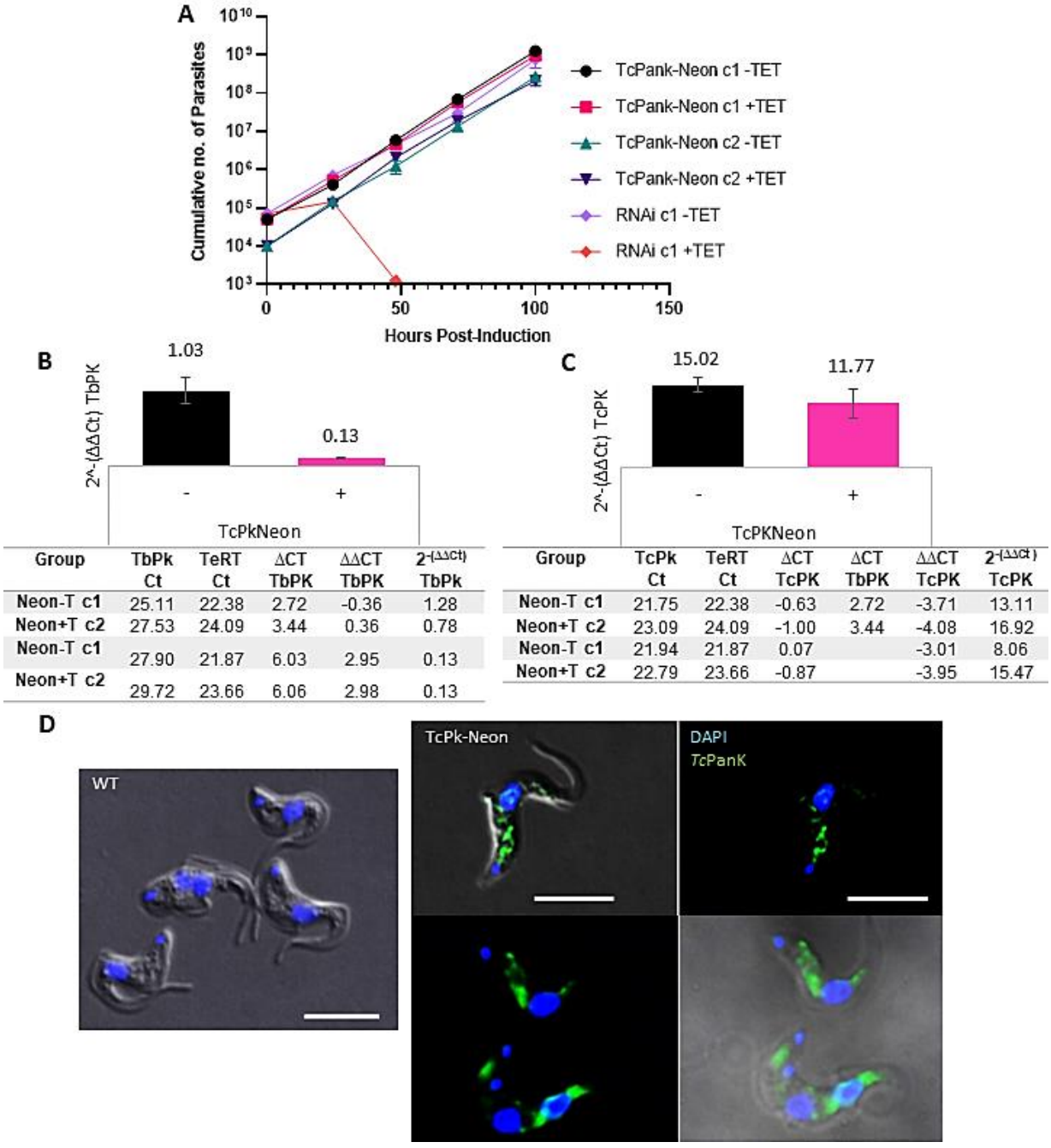
Parasites transfected with pTub-*Tc*PanK^{WT} were assessed for tetracycline-induced growth inhibition to determine functional homology of *Tc*PanK (Fig. 4.4.3 D). Two independent clones containing both the complementation and RNAi constructs (*Tc*PanK c1 and c2) and a clone containing solely the RNAi construct (RNAi c1) were seeded at a density of 5 x 10^4 parasites per well in a 24-well plate in triplicate, with and without tetracycline (1 µg/ml) (Materials and Methods, section 2.10). Cell densities of each group were measured at four consecutive 24-hour intervals (Fig. 4.4.3 D). As expected, tetracycline-treated RNAi cell lines began to die within 48 hours. In contrast, the proliferation rate of both *Tc*PanK-expressing cell lines was unchanged by RNAi induction, and grew at the same rate as the untreated groups. qPCR analysis confirmed that endogenous *Tb*PanK mRNA levels declined in treated groups by 86% relative to untreated parasites at 16 hours post-induction (Fig. 4.4.3 E). *Tc*PanK mRNA was also quantified in each cell line: $2^{-\Delta\Delta Ct}$ values calculated using *Tb*PanK levels in uninduced parasites as the calibrator value indicated that complemented lines expressed approximately 20-fold more *Tc*PanK than *Tb*PanK mRNA (Fig. 4.4.3 F).

TcPanK-mNeonGreen add-back

In light of the observed functional homology of *Tc*PanK, I next performed a genetic complementation experiment using the same ORF with the addition of a C-terminal green fluorescent tag to visualise the subcellular location of the add-back protein (Materials and Methods, section 2.3.5). The region encoding the C-terminus of *Tc*PanK in construct *pTub-TcPanK^{WT}* (Appendix SF7) was excised and replaced with a cloned fragment consisting of the missing *Tc*PanK C-terminus without a stop codon, followed by a glycine-serine linker peptide and the mNeonGreen (mNG) ORF, to generate construct pTub-*Tc*PanK-mNG (Appendix SF8).

Following transfection with construct pTub-*Tc*PanK-mNG, and selection with blasticidin (Materials and Methods, section 2.4.2), two *T. brucei* clones (*Tc*PanK-Neon clones 1 and 2) constitutively expressing fluorescently tagged *Tc*PanK, with the additional capacity to inducibly express the *Tb*PanK RNAi construct, were isolated. These parasites showed no proliferation defect following induction of RNAi targeted against the endogenous *Tb*PanK transcript (Fig. 4.4.6 A). Induction of the parental cell line (RNAi c1) caused the expected cell death phenotype, affirming persistent expression of the RNAi construct in the complementation cell lines. Depletion of *Tb*PanK was confirmed by qPCR (Fig. 4.4.6 B). On average, *Tb*PanK mRNA isolated from *Tc*PanK-Neon clones 1 and 2 was present at 88% of the uninduced level by 16 hours post-induction. At the same time-point, expression of *Tc*PanK in both induced and uninduced complemented cell lines was at least 11-fold greater than *Tb*PanK expression quantified from uninduced cells (Fig. 4.4.6 C).

To localise tagged *Tc*PanK in these clones, we performed confocal imaging on fixed parasites (Materials and Methods, section 2.9). The pattern of expression of exogenous *Tc*PanK in the *T. brucei* model was consistent with that observed in the PTP-tagged *T. cruzi* cell line, appearing to be associated with an organelle present throughout the cytosol, such as the endoplasmic reticulum (Fig. 4.4.6 D). PanK has also been tagged in *T. brucei* as part of the TrypTag project [292]. mNeonGreen tagging performed in this project indicated a cytoplasmic or endoplasmic reticulum locus for *Tb*PanK tagged at N- or C-termini (available on http://tryptag.org/?id=Tb927.11.7290).



Group	TbPk Ct	TeRT Ct	ΔCT TbPK	ΔΔCT TbPK	$2^{-(\Delta\Delta ct)}$ TbPk
Neon-T c1	25.11	22.38	2.72	-0.36	1.28
Neon+T c2	27.53	24.09	3.44	0.36	0.78
Neon-T c1	27.90	21.87	6.03	2.95	0.13
Neon+T c2	29.72	23.66	6.06	2.98	0.13

Group	TcPk Ct	TeRT Ct	ΔCT TcPK	ΔCT TbPK	ΔΔCT TcPK	$2^{-(\Delta\Delta ct)}$ TcPK
Neon-T c1	21.75	22.38	-0.63	2.72	-3.71	13.11
Neon+T c2	23.09	24.09	-1.00	3.44	-4.08	16.92
Neon-T c1	21.94	21.87	0.07		-3.01	8.06
Neon+T c2	22.79	23.66	-0.87		-3.95	15.47

Fig. 4.4.6. Fluorescently-tagged TcPanK complements TbPank knockdown. (A) *Induction of RNAi targeting TbPanK had no effect on growth of BSF T. brucei constitutively expressing fluorescently tagged (C-terminal) TcPanK, based on 2 clones (TcPanK-Neon c1 and c2). Induction of RNAi was lethal in parental cells containing only the TbPanK-depletion construct (RNAi c1). Tetracycline (TET) 1 µg/ml) was provided at T=0 hours to indicated cell lines. Cells counts were performed in triplicate per group at 24-hour intervals (mean values shown).* ***(B and C)*** *qPCR data for RNA extracted from TcPanK-Neon clones 1 and 2 at 16 hours post-tetracycline induction, showing:* ***(B)*** *Depletion of the TbPanK (TbPK) transcript with tetracycline treatment (+, +TET) relative to untreated groups (-, -TET), using telomerase reverse transcriptase (TeRT) as a house-keeping gene;* ***(C)*** *Expression of TcPanK (TcPK) with ∆CT values calculated relative to TeRT and ∆∆CT calculated relative to TbPK in untreated groups. Bars indicate mean 2^-(∆∆Ct) and standard error for each experimental group, calculated from triplicate cDNA samples from two experimental clones. Tables present mean Ct values per clone.* ***(D)*** *mNeonGreen fluorescence (green) in T. brucei RNAi cell lines constitutively expressing C-terminally-tagged TcPanK. Merge of green signal and DAPI staining (DNA) (blue) in differential interference contrast images. Images acquired using an inverted Nikon Eclipse Ti epifluorescence microscope (Materials and Methods, section 2.9). Scale bar represents 10 µm.*

4.5 Does Coenzyme A rescue lethality of *Tb*PanK depletion?

To validate the specific targeting of the CoA pathway in RNAi experiments, we performed *Tb*PanK depletion in the presence of the final pathway product. *In vitro* growth curves were performed using *T. brucei* BSF cell lines containing the *Tb*PanK depletion construct in the presence of tetracycline and/or exogenous CoA.

The dosage of CoA was chosen based on a published *in vitro* experiment performed on BSF *T. brucei* [219]. In this study, three compounds believed to target the CoA pathway were found to have a growth inhibitory effect on *T. brucei*. Specific targeting of the pathway was assessed by CoA supplementation at a concentration of 0.6-2 mM. The most effective rescue of growth inhibition was achieved using 1 mM CoA. This concentration was therefore chosen for the PanK RNAi experiments.

In the first rescue experiment, RNAi clone 1 (c1), which was engineered to facilitate *Tb*PanK mRNA depletion, received exogenous CoA at two time-points (Fig. 4.5A). In one experimental group, CoA was provided at the time of RNAi induction (T=0 hours) with tetracycline (1 µg/ml), while another group received CoA at 20 hours post-induction. Two control groups were included: one receiving tetracycline alone (RNAi c1 +TET) and another that received neither compound (RNAi c1). As expected, the untreated control group grew exponentially while cell death occurred in the RNAi c1 +TET group by 48 hours post-induction. The CoA-treated groups escaped cell death, instead transitioning to a slow-growth state. Based on the shape of the cumulative growth curves in Fig. 4.5 A, the timing of treatment had no effect on rescue, as 0 hour and 24 hour experimental groups follow a similar growth pattern. While neither of these groups grew exponentially, they did undergo a low level of proliferation, which is consistent with their morphologically healthy appearance when examined under a light microscope.

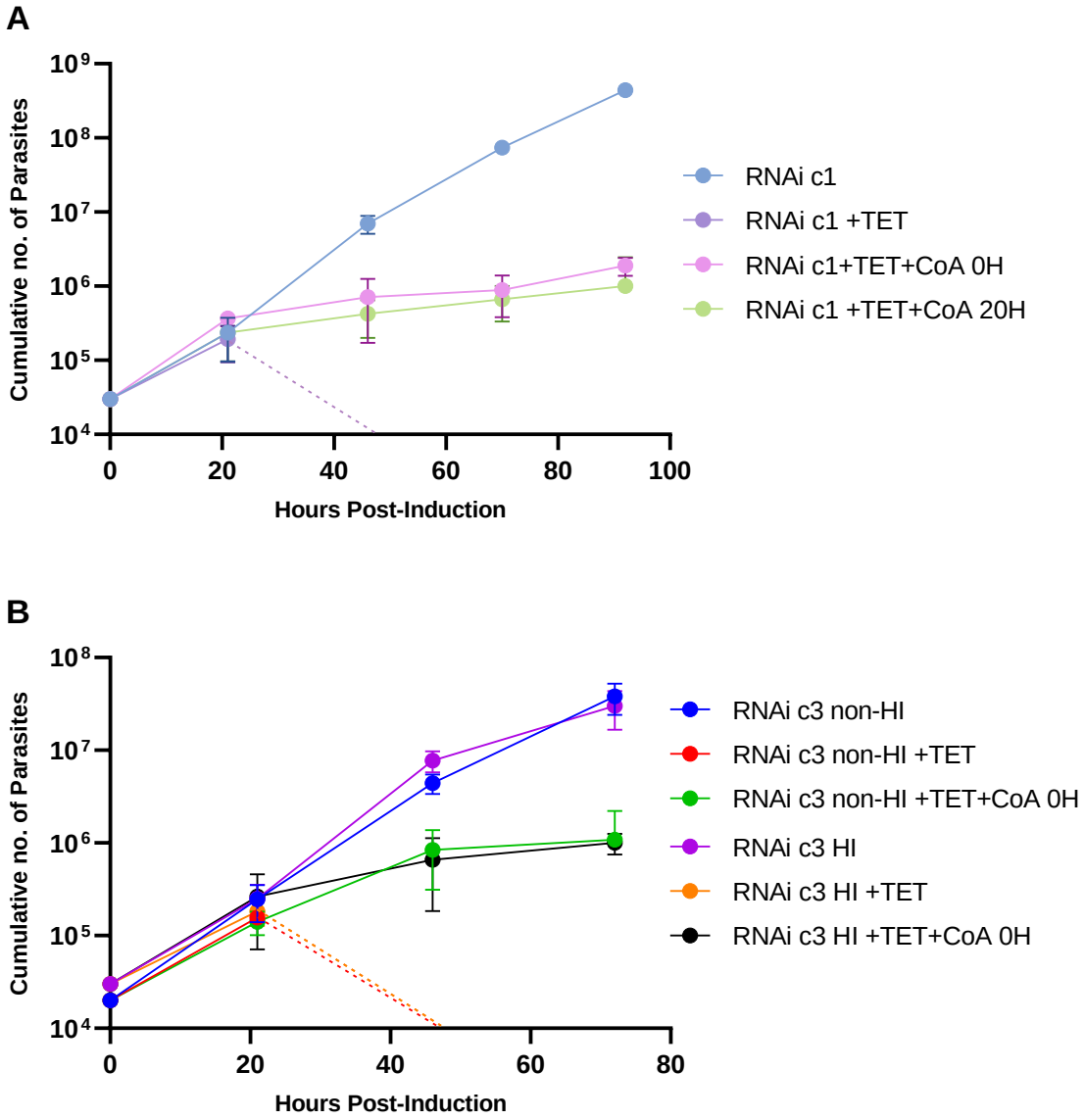


Fig. 4.5: Rescue of RNAi lethality by Coenzyme A supplementation: *Growth curves of T. brucei BSF RNAi clones in the presence or absence of tetracycline (TET, 1 µg/ml) and CoA (1 mM), established as outlined in Materials and Methods (section 2.10).* ***(A)*** *CoA was added at 0 or 20 hours post-induction to RNAi clone 1 (c1).* ***(B)*** *CoA was added at 0 hours post-induction to the two serum-supplemented groups: non heat-inactivated (non-HI) and heat-inactivated (HI), using RNAi clone 3 (c3). Both clones 1 and 3 contain an RNAi construct which triggers the depletion of TbPanK in the presence of tetracycline.*

Srinivasan and colleagues made the observation that CoA used in *in vitro* supplementation experiments was unstable in serum, being hydrolysed by pyrophosphatases to 4′-phosphopantetheine [293]. 4′-phosphopantetheine, but not CoA, was shown to passively diffuse across biological membranes of cultured Drosophila S2 and HEK293 cells, using a parallel artificial membrane-permeability assay. Once inside cells, 4′-phosphopantetheine replenished intracellular CoA levels. Due to the limited extent of rescue observed in our CoA

supplementation experiment, we tested whether the heat-inactivated FBS used in our HMI-9 medium might prevent rescue through the absence of active serum enzymes.

Using a distinct RNAi clone (RNAi c3) to demonstrate reproducibility, we included both heat-inactivated (HI) and non-heat-inactivated (non-HI) serum groups (Fig. 5.4 B). CoA (1 mM) was added at the time of RNAi induction. The effect of supplementation was consistent with the results for the previously tested clone. While the induced, untreated group (+TET, 0mM CoA) underwent cell death, CoA-treated groups instead increased in cell number over the 72 hour assessment period. The use of non-HI serum seemed to have only a slight effect that was confined to the period of 24-48 hour post-induction, based on a single experiment using two cell lines.

4.6 Summary

The experiments described in this chapter provide an initial functional characterisation of the unique kinetoplastid pantothenate kinase and is the first comprehensive study to demonstrate the importance of coenzyme A synthesis in trypanosomes. My observations are:

- Tagging of *Tc*PanK in epimastigote life-stages of *T. cruzi* CL Brener indicates a protein of approximately 170 kDa

- This suggests three putative *Tc*PanK activities (phosphodiesterase, adenylation and PanK) are associated with a single polypeptide with no further cleavage into individual functional domains

- PanK is extra-nuclear in both *T. cruzi* and *T. brucei*

- *T. cruzi* null mutants could not be generated by CRISPR-SpCas9-mediated replacement of both *TcPanK* alleles, consistent with the gene being essential

- RNAi-mediated *Tb*PanK depletion in bloodstream form *T. brucei* resulted in mortality, indicating the essential nature of the protein

- The PanK gene is functionally conserved between *T. cruzi* and *T. brucei.* Expression of *TcPanK* in *T. brucei* following RNAi-mediated depletion of the endogenous enzyme was able to rescue the parasite from a lethal outcome

- Supplementing the growth medium of bloodstream form *T. brucei* parasites that had undergone RNAi-mediated depletion of *Tb*PanK with exogenous CoA was able to rescue parasites from a lethal outcome. The failure to fully restore normal growth may indicate an impact of the putative activities associated with the other functional domains in the large polypeptide (phosphodiesterase and adenylation)

5. Importance of fused domains to the essential role of *T. cruzi* pantothenate kinase

5.1 Introduction

In the previous chapters, I established the importance of pantothenate kinase to the *in vitro* proliferation of clinically important trypanosomes and predicted the enzyme to be multi-functional. In this chapter, I investigate the importance of the fused domains in the multi-functional *Tc*PanK. The approach taken exploited the *T. brucei* RNAi model, combined with add-back of modified *Tc*PanK constructs containing truncations or residue substitutions. Specific residues were selected for substitution based on their conservation amongst characterised enzymes with predicted homology. Substitutions were then incorporated by site-directed mutagenesis into add-back constructs. Where possible, I accompanied our *in vitro* findings with biochemical validation of protein activity, by expressing recombinant *Tc*PanK in bacteria and conducting time-course assays with purified extracts.

In this work I have identified an essential residue for PanK domain activity, which is also necessary for *in vitro* proliferation of parasites. In addition, I found that the PanK domain itself, whilst essential, was not sufficient to sustain normal growth, indicating that the fused domains contribute to the essential function(s) of *Tc*PanK. This further demonstrates the suitability of *Tc*PanK as a candidate for drug development.

5.2 Arginine 1270 is an important *Tc*PanK residue

Computational predictions

My first aim was to identify and mutate residues within *Tc*PanK predicted to be essential for activity and then determine the effects of these mutations *in vitro*. Since the *T. brucei* RNAi cell lines had a very clear phenotype upon *Tb*PanK depletion, that was entirely rescued by expression of the *T. cruzi* homologue (Fig. 4.4.3), I was able to use this *in vitro* model for mutation analysis. By delivering constructs containing modified *TcPanK* genes into cell lines containing the *TbPanK* depletion construct, I was able to assess the impact of mutations by the extent of rescue of RNAi-mediated lethality.

To select amino acids for mutagenesis, conserved residues with experimentally validated importance were identified. An alignment of the *Tc*PanK amino acid sequence with PanKs from distant relatives revealed multiple highly conserved residues within the PanK domain (residues 1074-1480). Amongst sequences from *H. sapiens*, *D. melanogaster*, *S. cerevisiae* and *A. thaliana,* 62 amino acids in the corresponding region are conserved with *Tc*PanK (Fig. 5.2.1 A). A literature search was performed to obtain experimental evidence for residues that participate in substrate interactions and therefore might be essential in *T. cruzi*. The crystal structure of human PanK isoform 3 has been resolved in complex with pantothenate, phosphopantothenate, magnesium, AMP-PNP (a non-hydrolysable version of ATP) and ADP, revealing active site residues [239]. Amongst these, residues which align with S1254 and R1270 of *Tc*PanK were shown to interact with substrates in the active site. S192 *(H. sapiens)* which aligns with S1254 (*Tc*PanK), interacts with ATP (via a single hydrogen bond), while R207, which aligns with R1270, forms two hydrogen bonds to pantothenate. The interactions of these residues were reconstructed with Pymol [262], using the human PanK3 resolved structure (Fig. 5.2.1 B-C). The AlphaFold predicted structure of *Tc*PanK was aligned to human PanK3, which confirmed that the *T. cruzi* residues S1254 and R1270 structurally align to the human active site residues. On the basis of these analyses, S1254 and R1270 were chosen for site-directed mutagenesis.

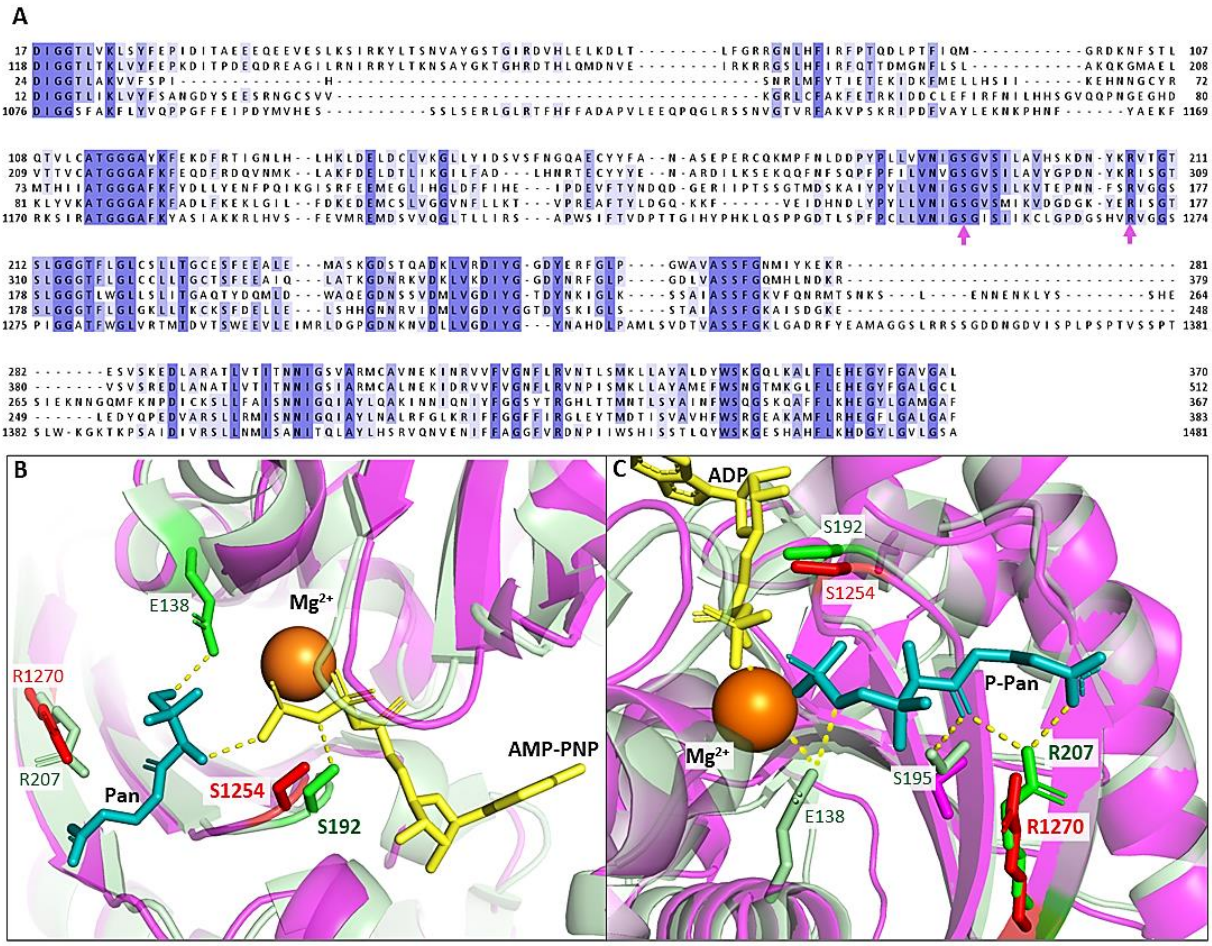


Fig. 5.2.1 Conserved interactions in the pantothenate kinase active site: (A) *ClustalO alignment of pantothenate kinase amino acid sequences from H. sapiens, D. melanogaster, S. cerevisiae, A. thaliana and T. cruzi (in order from top to bottom), with shading representing 50, 75 or 100% identity in order of intensity. Arrows indicate residues selected for mutagenesis in TcPanK (S1254 and R1270).* ***(B)*** *Active site structure of the PANK3·AMPPNP·Mg^{2+}·pantothenate complex (H. sapiens, PDB code: 5KPR [239]) aligned with the AlphaFold predicted TcPanK structure.* ***(C)*** *PANK3·ADP·Mg^{2+}·phosphopantothenate complex (PDB code: 5KPZ) aligned with the AlphaFold predicted TcPanK structure. Ligand side-chains of PANK3 and TcPanK are displayed in bright green and red, respectively, and backbones are displayed as pale green (PANK3) and magenta (TcPanK) ribbons. Substrates/products: Pan: pantothenate, P-Pan: Phosphopantothenate (in blue); ADP and AMP-PNP (in yellow); Mg^{2+} (in orange). Interactions are represented by yellow dotted lines and based on Subramanian et al., 2016 [239].*

Site-directed mutagenesis

A two-step PCR protocol was employed to substitute the conserved polar (S1254) or charged (R1270) residues with non-polar alanine in separate constructs (Materials and Methods, section 2.3.5; Fig. 5.2.2). For the S1254/A transition, the codon TCC was replaced with GCC, and for the R1270/A transition, the codon CGC was replaced with GCC (for primer sequences, see appendix Table A.3). As outlined in Fig. 5.2.2, in the first step, primer F incorporating the alanine codon at the desired sites, was used in conjunction with primer R and the cloned *TcPanK* template in a PCR reaction that generated a partial gene fragment containing the

targeted substitutions. The product was then used as a primer for extension across the remaining gene sequence, producing a complete *TcPanK* ORF with the incorporated modifications and flanking restriction sites for cloning into pTub-EX (Appendix. SF6) (Materials and Methods, section 2.3.5). The cloned products were then sequenced to confirm that the correct mutations had been introduced and that no other changes had occurred as a result of the amplification process (Fig. 5.2.3 A-B). After linearization with enzymes NotI and XhoI the purified product was used to transfect the *T. brucei* cell line RNAi c1 (Fig. 4.4.3 A) by electroporation. 5 days following transfection, blasticidin-resistant clones were selected (Materials and Methods, section 2.4.2) and at least 2 of each were used for subsequent analysis.

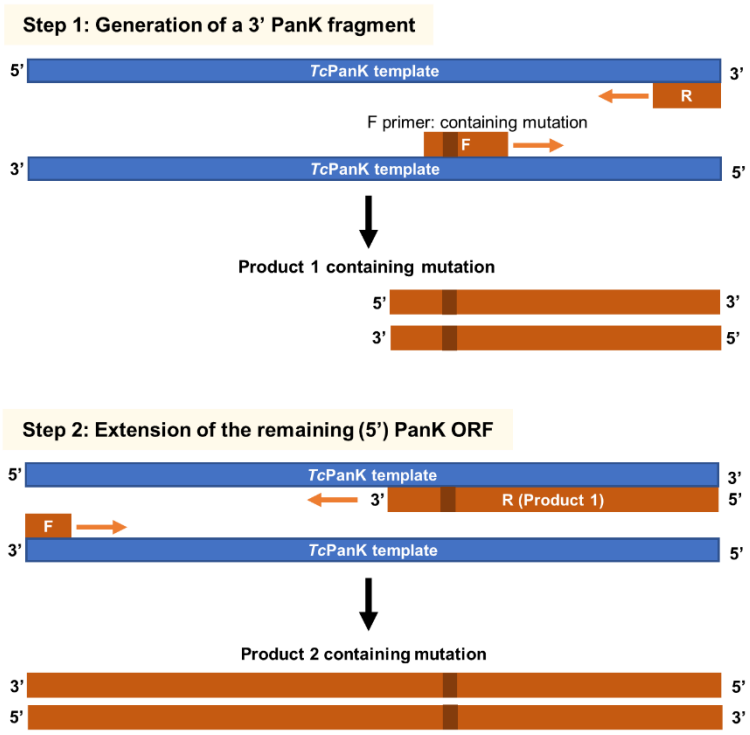


Fig. 5.2.2. Two step-PCR mutagenesis: *In step 1, primers incorporating an alanine codon at the desired sites within the TcPanK template, generate a partial fragment containing the substitutions. In step 2, the product of PCR 1 was used as a primer for extension across the remaining gene sequence, producing a complete TcPanK ORF with the incorporated modifications.*

Functional analysis of mutant versions of TcPanK

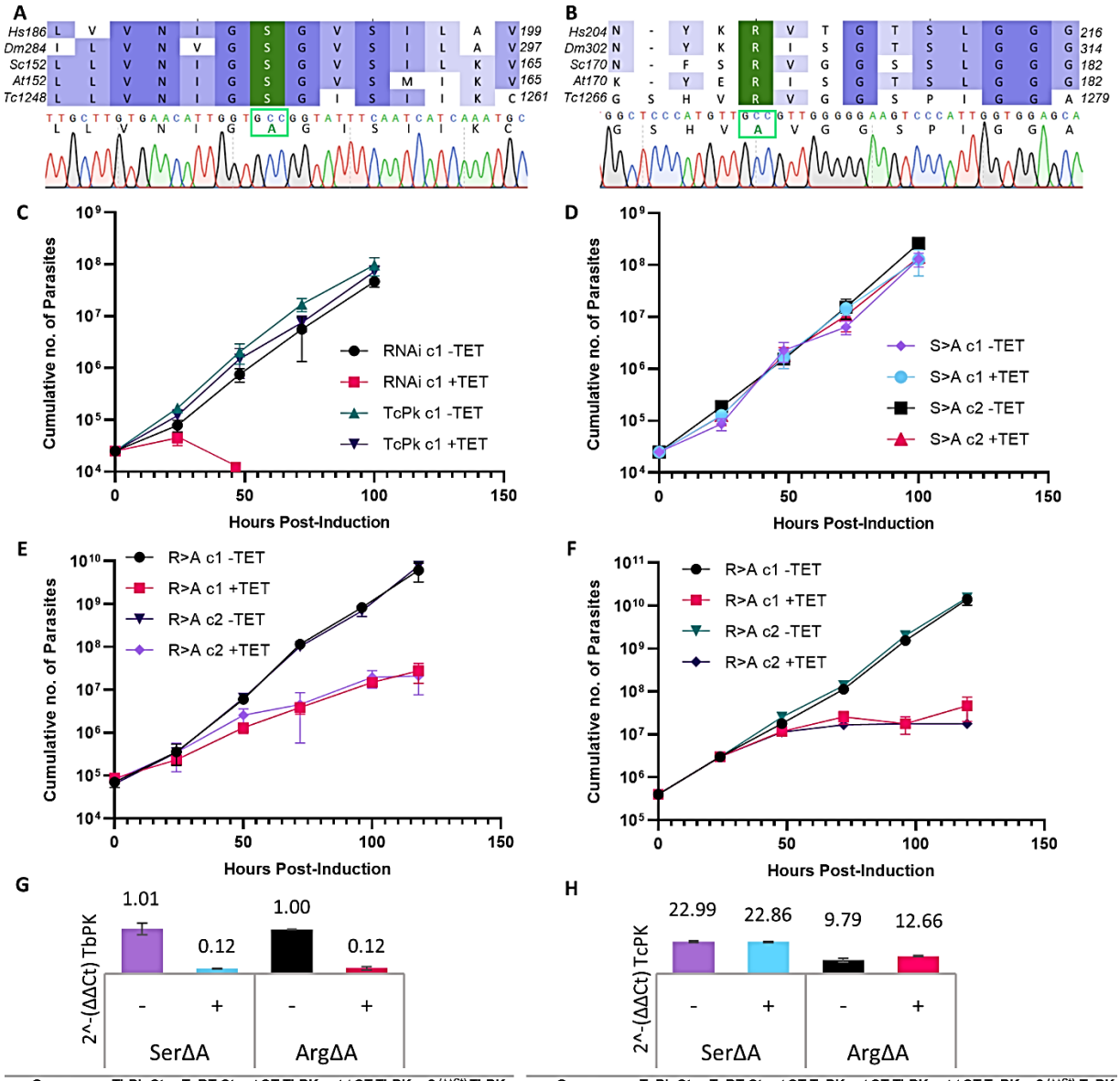
Growth curves were performed using the parental bloodstream form *Tb*PanK RNAi cell line (RNAi c1), a WT *Tc*PanK expressing cell line (TcPK c1) and the cell lines that had been manipulated to express the mutant *Tc*PanK enzymes. As in Materials and Methods section

2.10, the experimental groups were seeded at equal densities with or without tetracycline, and cumulative cell densities were calculated at 24-hour intervals for 4 days post-induction.

Growth data revealed that, as expected, induction of *Tb*PanK depletion was lethal with the RNAi c1 cell line and had no effect on *Tc*PanK c1 (Fig. 5.2.3 C). Two clones expressing S1254A constructs (S>A c1 and c2) showed no growth defect in response to *Tb*PanK knockdown, growing at the same rate as the uninduced groups and *Tc*Pank c1 (Fig. 5.2.3 D). In contrast, induction of RNAi in two clones transfected with *Tc*PanK R1270A constructs (R>A1 and 2) severely impaired proliferation (Fig. 5.2.3 E). By 5 days post-induction, the cumulative density (CD) of induced clones R>A1 and R>A2 was 220 and 365 fold lower than uninduced groups, respectively. Growth curve analysis for R>A1 and 2 was repeated to confirm this result (Fig. 5.2.3 F). The repeat curve revealed CD of induced cell lines to be 301-864 fold lower than that of uninduced parasites by day 5 post-induction.

To confirm the specificity of RNAi targeting, *Tb*PanK mRNA was measured in each cell line by qPCR at 16 hours post-induction (Fig. 5.2.3 G). The fold change in *Tb*PanK mRNA ($2^{\Delta\Delta Ct}$ *Tb*PanK) in induced cell lines relative to uninduced controls is shown in the bar graph in Fig. 5.2.3 G. *Tb*PanK mRNA levels were more than 80% lower in induced add-back cell lines compared with uninduced controls. In all cases, *Tb*PanK was quantified relative to *T. brucei* telomerase reverse transcriptase (TERT) resulting in the ΔCt *Tb*PanK values shown in the table (Fig. 5.2.3 G). To verify the expression of the add-back constructs themselves, *T. cruzi* PanK mRNA was also quantified under the same experimental conditions (Fig. 5.2.3 H). To compare expression of *Tc*PanK with endogenous *Tb*PanK, ΔΔCt *Tc*PanK (shown in the table in Fig. 5.2.3 H) was normalised to the ΔCt *Tb*PanK value obtained from uninduced cell lines. Based on these calculations, *Tc*PanK was expressed in all add-back cell-lines, at more than 8x endogenous *Tb*PanK levels: S>A mutants contained ≥22- fold more *Tc*PanK than *Tb*Pank and R>A mutants contained 8-13 fold more *Tc*PanK mRNA than *Tb*Pank.

These data therefore demonstrate that Serine1254 does not have an essential role in *Tc*PanK activity. In contrast, R1270 appears to be functionally important, since expression of a mutated protein with an alanine residue at this position does not complement *T. brucei* cells in which the endogenous enzyme *Tb*PanK has been depleted by RNAi.



Group	TbPk Ct	TeRT Ct	ΔCT TbPK	ΔΔCT TbPK	$2^{-(\Delta\Delta Ct)}$ TbPK
S>A -TET c1	26.55	22.40	4.15	-0.20	1.14
S>A -TET c2	26.59	22.05	4.54	0.20	0.87
S>A +TET c1	29.42	21.91	7.51	3.16	0.11
S>A +TET c2	29.54	22.18	7.35	3.01	0.12
R>A -TET c1	25.33	21.96	3.37	-0.01	1.01
R>A -TET c2	25.40	22.01	3.38	0.01	0.99
R>A +TET c1	28.08	22.16	5.92	2.55	0.17
R>A +TET c2	28.99	21.85	7.14	3.77	0.07

Group	TcPk Ct	TeRT Ct	ΔCT TcPK	ΔCT TbPK	ΔΔCT TcPK	$2^{-(\Delta\Delta Ct)}$ TcPK
S>A -TET c1	22.25	22.40	-0.15	4.15	-4.49	22.54
S>A -TET c2	21.84	22.05	-0.20	4.54	-4.55	23.44
S>A +TET c1	21.77	21.91	-0.14		-4.49	22.44
S>A +TET c2	21.99	22.18	-0.19		-4.54	23.28
R>A -TET c1	22.42	21.96	0.46	3.37	-2.91	7.53
R>A -TET c2	21.80	22.01	-0.21	3.38	-3.59	12.04
R>A +TET c1	21.94	22.16	-0.22		-3.59	12.07
R>A +TET c2	21.50	21.85	-0.35		-3.73	13.24

Fig. 5.2.3. Complementation of TbPanK knockdown with kinase domain mutants:
(A,B) *Aligned PanKs from H. sapiens, D. melanogaster, S. cerevisiae, A. thaliana and T. cruzi (top to bottom) with conserved residues highlighted. S1254 and R1270 of TcPanK (in green) were substituted for alanine in the mutant TcPanK complementation constructs as the chromatograms indicate.* ***(C-F)*** *Cumulative growth curves showing the effect of tetracycline (TET, 1 µg/ml)-induced TbPanK depletion on T. brucei cell lines containing RNAi hairpin constructs, either alone (RNAi c1) or in addition to a complementation construct consisting of WT TcPanK (TcPk c1) or TcPanK containing S1235A (S>A c1-2) or R1270A (R>A c1-2) substitutions. Two growth curves were performed for R>A clones to confirm the proliferation defect upon tetracycline induction. Each data point is an average of 3 wells, counted in triplicate.* ***(G-H)*** *qPCR data for RNA extracted from mutant add-back clones at 16 hours post-induction. Bar charts indicate mean 2^-(ΔΔCt) and standard error for each experimental group (based on 2 clones) and tables provide mean Ct per clone.* ***(G)*** *Depletion of TbPanK (TbPK) transcript with tetracycline treatment (+, +TET) relative to untreated groups (-, -TET): mean $2^{\Delta\Delta Ct}$ TbPK was calculated for two cell lines per group (+/- TET), using telomerase reverse transcriptase (TeRT) as a house-keeping gene.* ***(H)*** *Levels of TcPanK (TcPK) in add-back cells, relative to TbPK levels in uninduced add-back groups.*

5.3 Arginine 1270 is required for recombinant *Tc*PanK-domain activity

Generation of constructs for in vitro expression of TcPanK

To ascertain whether the residue R1270 is required for protein activity, I aimed to express and purify recombinant 'wildtype' (WT) *Tc*PanK and a mutant containing the R1270A substitution and then measure their activities using a kinase assay.

To express recombinant *Tc*PanK (rec*Tc*PanK) I tested three expression systems: a cell-free bacterially-derived system (NEBExpress), *E. coli* and the mammalian cell line HEK293T. Expression of the entire protein (168 kDa) and the putative 51 kDa PanK catalytic domain (rec*Tc*PanK-domain, residues 1044-1480) was attempted in each system. For the single domain protein, both WT and R1270A mutant constructs were expressed.

In all cases the expression systems were designed to incorporate hexa-histidine tags onto the expressed polypeptides. Initially both N and C-terminal tags were trialled using the NEBExpress system, however to simplify the methodology, N-terminal tags were exclusively chosen for most experiments, due to convenience of N-terminal tagging in the chosen vectors and lack of observed differences in protein yields. Furthermore, complementation experiments in *T. brucei* suggested that both N- and C-terminally tagged *Tc*PanK (full-length) is functional, based on successful rescue of lethality of *Tb*PanK depletion (Fig. 5.3.1). For this experiment, *TcPanK* gene fragments with an N- or C-terminal hexa-histidine tag coding sequence were generated (Materials and Methods section 2.3.5; Appendix Table A.3) and inserted into the pTubEX construct (Appendix Figs. SF9,10). These were used to transfect the *Tb*PanK RNAi cell line (RNAi c1). The endogenous gene was then depleted following tetracycline-induced RNAi-mediated knockdown, and the ability of both the N- and C-terminal histidine tagged *Tc*PanK to maintain cell viability and growth was confirmed.

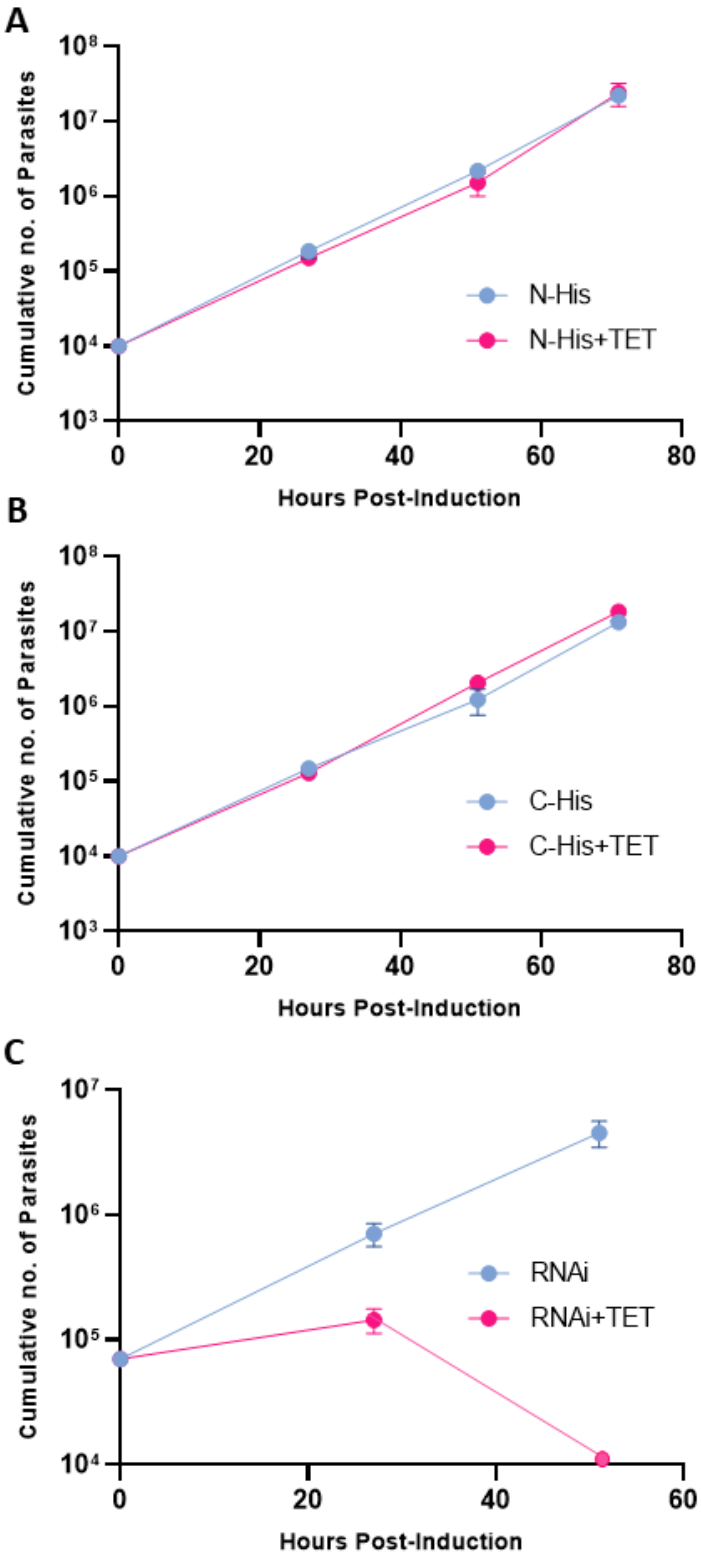


Fig. 5.3.1: His-tagged TcPanK rescues lethality of TbPanK depletion: *Cumulative growth curves showing the effect of tetracycline (TET, 1 µg/ml)-induced TbPanK mRNA depletion on cell lines expressing RNAi hairpin constructs and full-length TcPanK containing an N-* ***(A)*** *or C-terminal* ***(B)*** *hexa-histidine tag.* ***(C)*** *The parental cell line containing the RNAi TbPanK depletion construct without TcPanK undergoes cell death. Experimental conditions were as outlined previously (Fig. 5.2.3)*

For the cell-free expression system, fragments containing the PanK sequence were ligated with the NEBExpress plasmid (Appendix, SF11), such that the recombined gene came under the control of a T7 promoter. The primers used for the amplification reaction are listed in Appendix Table A.8. For expression, the construct was added to a buffer containing S30 ribosomal extract, T7 RNA Polymerase and the required substrates, salts and cofactors (Materials and Methods, section 2.12.1).

For mammalian expression, *Tc*PanK was cloned into a pcDNA expression vector (Appendix SF12; Materials and Methods, section 2.12.2; Appendix table A.9). As well as a histidine tag and interrupting linker sequence, the insert contained a Kozak sequence for translation initiation. The vector drives expression by the strong human cytomegalovirus promoter. It also contains an SV40 promoter and origin components, which allow copy number amplification in the HEK293T cell line, which possesses SV40 large T antigen. The vector was delivered to HEK cultures by polyethyleneimine (PEI) incubation using cells at approximately 60% confluency (Materials and Methods section 2.4.3). Cells were left to incubate at 37°C for 24-, 48- or 72- hours.

For bacterial expression, the *E. coli* strain BL21 (DE3) was used, which is deficient in Lon and OmpT proteases and expresses T7 RNA polymerase from a lac promoter. Two vectors were trialled for bacterial expression of *Tc*PanK: pTrcHisC (Appendix SF13), which drives expression by the Trc promoter, a hybrid tryptophan and lac promoter optimised for high-level expression, and pET (Appendix SF14), which contains a T7 promoter. Two methods of induction were tested: isopropyl-β-thio-galactopyranoside (IPTG) and autoinduction using diauxic growth media. Several incubation duration and temperature combinations were tested (Materials and Methods, section 2.12.3).

Expression and purification of recombinant TcPanK

Purification of recombinant protein from cell lysates was by nickel affinity chromatography (NAC) (Materials and Methods, section 2.12.4). Empty vector controls were included in all expression and purification procedures to assess the specificity of expression and purification products. The outcome of these trials was first assessed by SDS-PAGE. In all cases, based on stain-free visualisation or Coomassie blue staining, it was not possible or very difficult to identify the recombinant polypeptide in crude lysates separated by PAGE, due to low levels of expression and the presence of similar sized native proteins.

Western blotting was used in an attempt to visualise expression products. Gels were transferred to nitrocellulose membranes, which were then incubated overnight with αHis antibody (Materials and Methods, section 2.7). However, the full-length (168 kDa) *Tc*PanK was not detected using any of the three expression systems (Fig 5.3.2, as example). In most cases, no product was visible, although in one case a truncated (~25 kDa) product was expressed and visualised. The truncation product was expressed in *E. coli* BL21 using the pTrcHis vector, and detected following SDS-PAGE-separation of NAC eluates and western blotting of crude lysates (Fig. 5.3.2). An attempt was made to codon-optimise nucleotides 226-232 of *TcPanK* which contains a potentially interfering Shine-Dalgarno sequence (AGGAGG) by substituting A**A**GA**A**G. However, this was also unsuccessful, resulting in expression of a truncated product of 20-30 kDa.

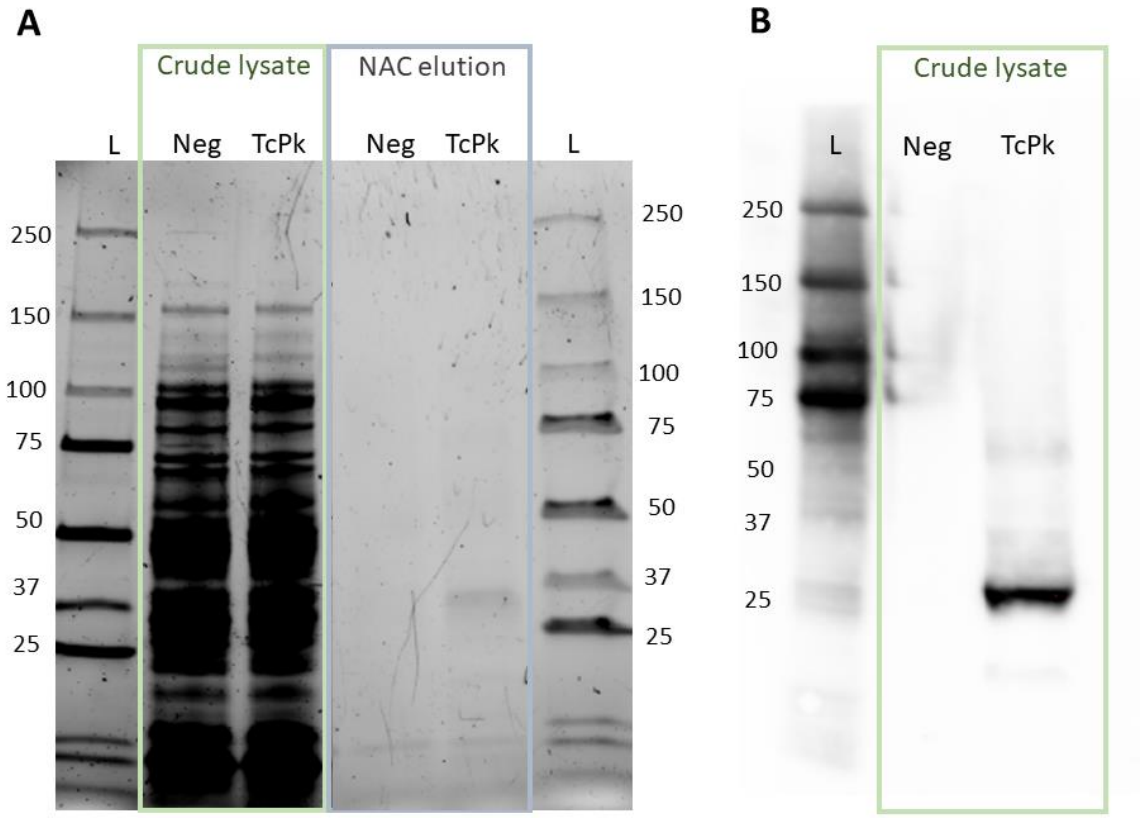


Fig. 5.3.2 Inability to express full-length recTcPanK: *E. coli BL21(DE3) transformed with pTrcHis containing an insert encoding N-terminally tagged WT TcPanK (residues 1-1480), or 'empty' pTrcHis, were induced with 50 μM IPTG and incubated at 37°C for 3 hours. Soluble lysates were purified by nickel affinity chromatography (NAC), using 300 mM imidazole elution (Materials and Methods, section 2.12.4).* ***(A)*** *Crude cell lysate preparations (~10 μg), NAC eluates (10x concentrated) and protein ladders were resolved by SDS-PAGE using a stain-free 4-20% polyacrylamide gel. TcPanK eluates contained a 25-37 kDa band which was absent from the negative control.* ***(B)*** *Crude lysates separated by SDS-PAGE were transferred to nitrocellulose membrane and probed with αHis (1:1000). The blot indicated expression of a histidine-reactive ~25 kDa protein in TcPanK cell lanes. Neg: empty pTrcHis samples; TcPK: pTrcHis-WT-TcPanK samples; L: protein ladder.*

I hypothesised that the size and multi-domain nature of the full-length *Tc*PanK protein (1480 residues) may have been a confounding factor that limited recombinant expression. Attempts

were therefore made to specifically express the putative PanK catalytic domain. The His-tagged rec*Tc*PanK-domain (1044-1480) was expressed in all three systems. Fig. 5.3.3 displays representative gel and western blot images showing consistent and specific detection of a ~50 kDa product using the cell-free (CF), *E. coli* BL21 and mammalian HEK expression systems. CF and HEK expression yielded comparable levels of protein (Fig. 5.3.3 B). The yield from BL21 expression was much higher than for HEK and CF based on visualisation by western blotting (5.3.3 C-E). The pET-*Tc*PanK vector encodes an N-terminal 60 kDa NUS tag, therefore the expected product size was ~110 kDa. As Fig. 5.3.3E indicates, two products were observed with apparent masses of 75 and 100-130 kDa.

By testing several variable parameters, I identified that the highest yield was achieved with bacterial expression using the pTrcHisC-PanK-domain construct combined with 50 µM IPTG induction and a 3-4 hour incubation period at 37°C. **Therefore, this system and protocol was employed for recombinant protein expression**. The quantified yield of the purified rec*Tc*PanK-domain by pTrcHis/BL21 expression was 150 µg product per 100 ml culture.

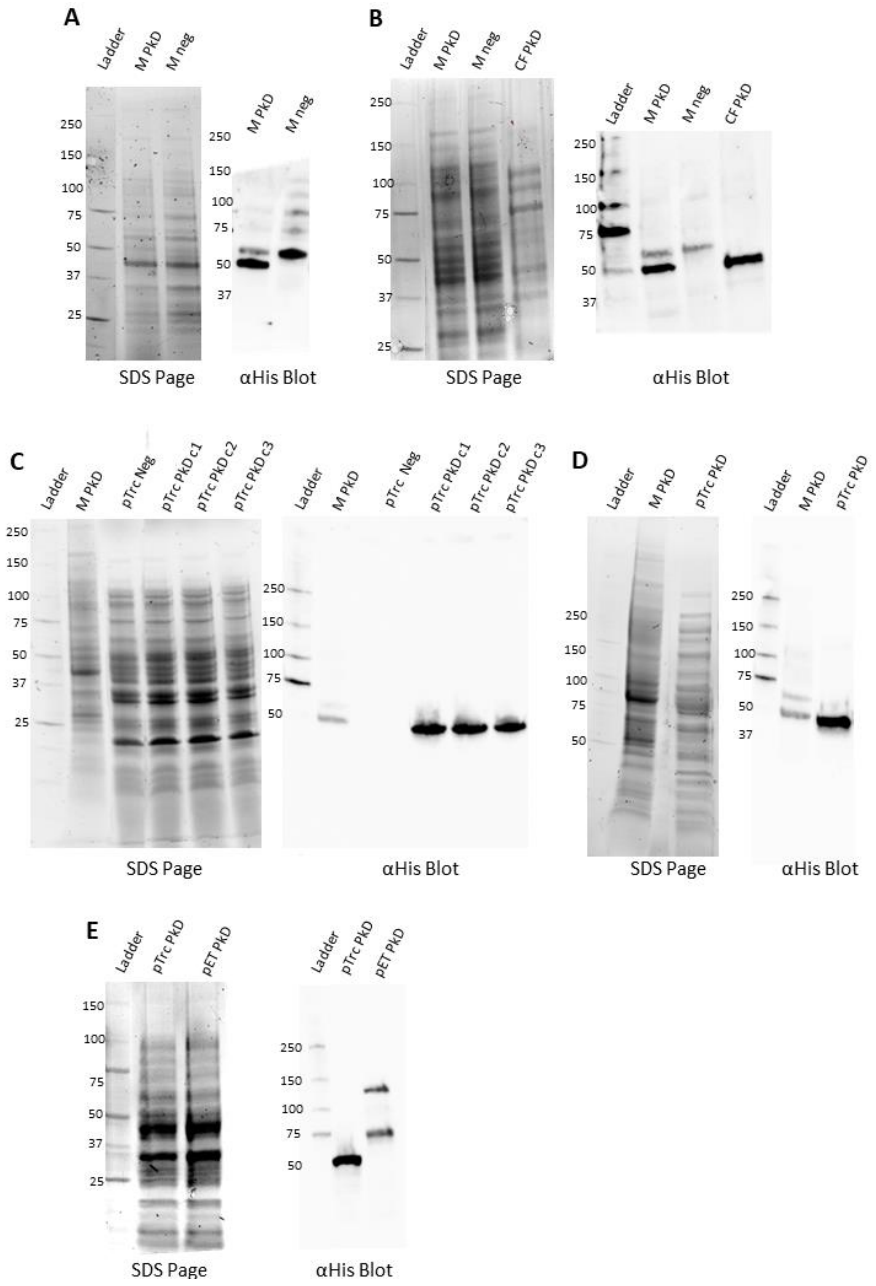


Fig. 5.3.3. Expression of the WT RecTcPanK-domain: *SDS-PAGE-separated whole cell lysates and western blots probed with anti-His tag antibodies (diluted 1:1000) demonstrate successful expression of the WT recTcPanK-domain (1044-1480) (**A)** Mammalian expression: HEK293 expression using pcDNA-PanK-domain vector and a 48 hour incubation (M PkD) resulted in a specific α-His-reactive protein, consistent with the expected size of 51kDa and a non-specific higher molecular weight protein (55-60kDa) that was present in empty pcDNA lysates (M neg). **(B)** A repeat experiment using HEK293 (48 hour expression) yielded a specific ~50kDa and non-specific ~55-60 kDa product using pcDNA-PanK-domain (M PkD). Again the non-specific product was observed using empty vector (M neg). Cell-free expression using the NEBExpress system (lane marked CF PkD) yielded an α-His-reactive protein of equal size to HEK293-derived expression products. **(C)** Expression using three E. coli BL21 clones transformed with a pTrcHisC-PanK-domain construct (pTrc PkD c1-c3) yielded ~50kDa α-His-reactive products, which were of equal size to products of pcDNA-PanK-domain expression (M PkD) and absent from cells transformed with empty vector (pTrc Neg). **(D)** A repeat E. coli BL21 expression using pTrcHis-PanK-domain (pTrc PkD) yielded a ~50kDa α-His-reactive band, again equal in size to HEK293 products (M PkD). **(E)** BL21 expression using the pET-PanK-domain vector yielded two products ~75 and 100-130kDa in size, due to NUS tag fusion the expected product size was ~110kDa. ~10ug crude cell lysates were loaded per SDS gel lane. Protein ladder sizes are shown in kDa.*

Having optimised the *E. coli* BL21 expression based on western blotting analysis, the next step was optimising purification by NAC. Testable parameters included binding conditions, number of washes and imidazole concentration of buffers. To optimise binding, different pH, salt and imidazole conditions (listed in table 5.3) were tested. A pH 7.4 buffer containing 700 mM NaCl, 5 mM imidazole, 0.02% tween and 10% glycerol in 20 mM Tris provided an optimal yield of pure protein and minimised non-specific binding of contaminant proteins. 5 mM $MgCl_2$ was also added to potentially improve enzyme stability, as a general co-factor of PanKs. Another critical factor for binding specificity was the concentration of lysate used per volume of resin. 300 µl resin was used per 100 ml starting culture volume (condensed to 5 ml lysate) with the aim of saturating the resin with the target protein.

For washing, a 50 mM sodium phosphate buffer (pH 7.4) containing 10 mM imidazole and 750 mM NaCl was effective at removing weakly bound non-specific proteins. 5% glycerol was added to maintain solubility and stability of *Tc*PanK. Eight washes were performed to sufficiently remove unwanted proteins, which was determined by testing consecutive washes with Bradford reagent to ensure the disappearance of proteins. This was then confirmed by SDS-PAGE visualisation. The minimum required amount of imidazole was chosen for elution, to reduce the likelihood of imidazole interference in the downstream assay. A step-wise gradient of 40-100 mM (40 mM, 50 mM, 100 mM) was used in elution buffers and > 50% of the protein was eluted with 50 mM imidazole (Fig. 5.3.4).

The purest and most concentrated elution fractions, determined by SDS-PAGE, were exchanged to storage buffer-designed to maintain protein solubility and contained 50 mM Tris-HCl pH 7.5, 5 mM MgCl2, 300 mM NaCl and 10% glycerol. The process of buffer exchange was performed using a centrifugal filter that also concentrated the sample and further purified it by removing residual proteins of molecular weight <10kDa. A final concentration of 5 mg/ml was stored with 1 mM DTT and 0.5 mM EGTA to minimise protein oxidation.

Table 5.3.1: Optimisation of NAC binding conditions*: To improve the efficacy of recTcPanK-domain purification by nickel affinity chromatography, buffers were optimised according to pH and concentrations of NaCl, Imidazole, Tween-20 and Glycerol.*

Condition	pH	NaCl	Imidazole	Tween-20	Glycerol
Function	Protein stability	Optimise non-specific binding, protein stability	Reduce non-specific binding	Protein solubilisation and stability	Protein solubilisation and stability
Range	7-8	150-700 mM etc	0-15mM	0-0.1%	5-10%
Optimum	7.4-7.5	700mM	5mM	0.02-0.1	10%
Final buffer concentration	7.4	700mM	5mM	0.02	10%

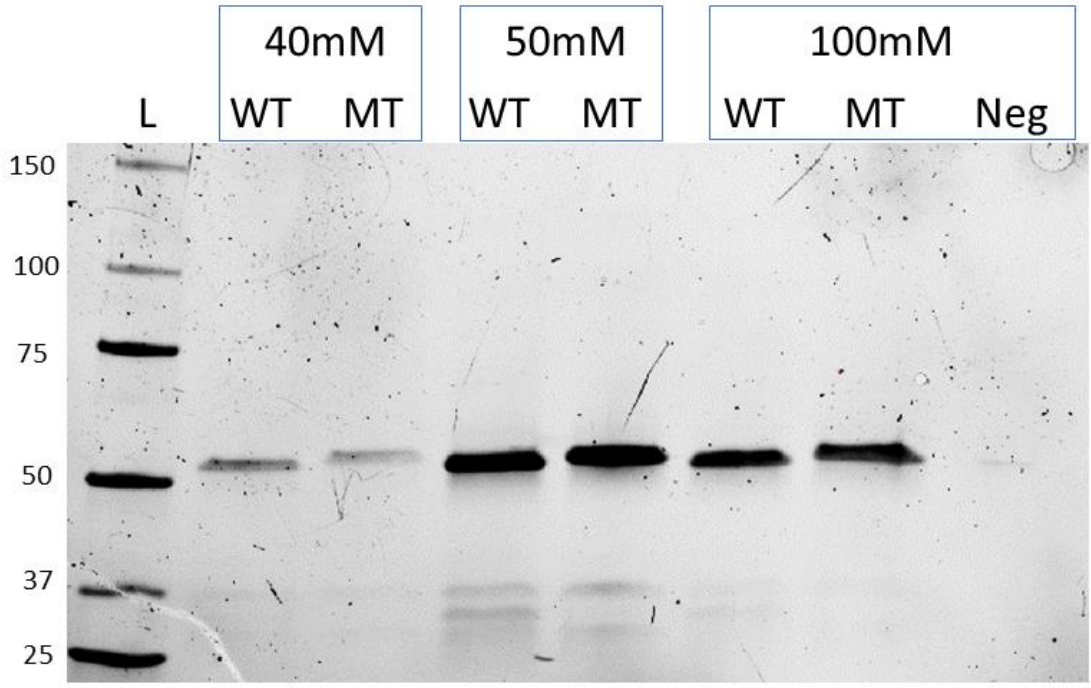


Fig. 5.3.4 Elution of NAC-bound TcPanK-domain using 40-100 mM imidazole: *SDS-PAGE reveals 50 kDa bands corresponding to His-tagged recTcPanK-domain (WT) and R1270A mutant recTcPanK-domain (MT) following purification of E. coli lysates by nickel affinity chromatography (NAC) and using stepwise imidazole elution concentrations of 40, 50 and 100 mM. Application of lysate from empty pTrcHis-transformed colonies to NAC resin and subsequent elution with 100 mM imidazole did not yield a 50kDa product (Neg). Protein ladder (L) sizes are indicated in kDa.*

Biochemical characterisation

The activity of purified *Tc*PanK-domain samples was measured indirectly using a pyruvate kinase-lactate dehydrogenase coupled assay (Fig. 5.3.5). In step 1, PanK catalyses the reaction: **Pantothenate + ATP -> 4'Phosphopantothenate +ADP.** The ADP generated in step 1 is then converted to ATP by pyruvate kinase (PK) in the reaction: **ADP + phosphoenolpyruvate (PEP) -> pyruvate +ATP**. Finally, pyruvate is converted to lactate by lactate dehydrogenase (LDH), which in the process reduces NADH to NAD+. This reduction of NADH is measured spectrophotometrically at 340nm wavelength.

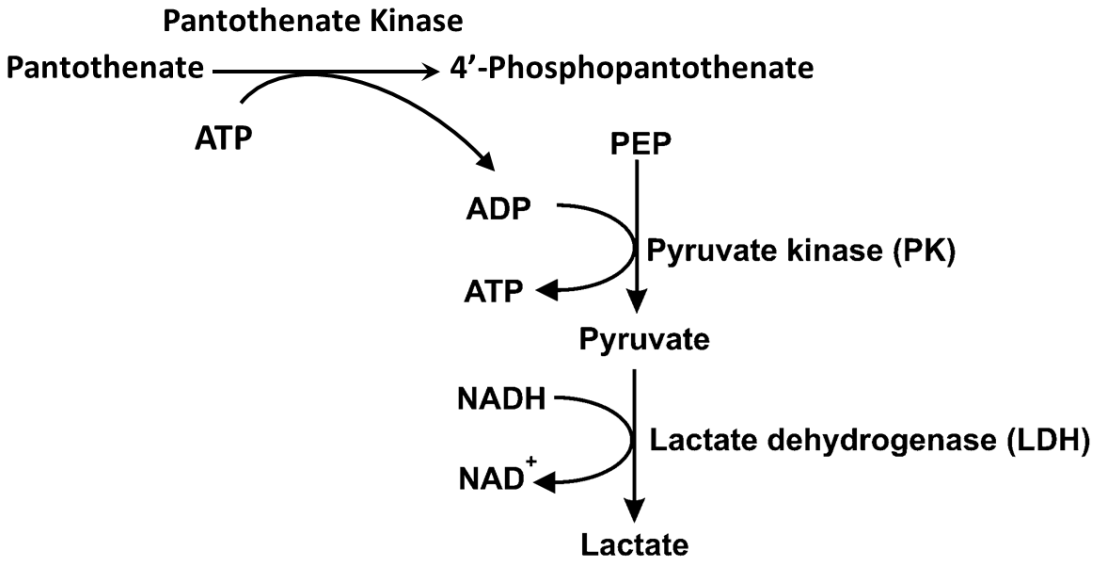


Fig. 5.3.5: PK-LDH coupled assay for PanK activity measurement.

The assay reactions were performed using 1 mM PEP, 33U PK/LDH, 200 μM NADH, 10 mM $MgCl_2$, 100 mM Tris (pH 7.4) and 8 μg rec*Tc*PanK-domain in a volume of 150 μl (Materials and Methods, section 2.13). Kinetic parameters for substrates ATP and pantothenate (Pan) were determined using 0.1-5 mM ATP and 4 mM Pan or 0.05-4 mM Pan and 3 mM ATP. Each assay was set up in triplicate in a 96-well plate at 24°C. 340 nm absorbance readings were taken every 30 seconds before and after adding ATP to initiate the reaction. Before ATP addition, readings were taken for approximately 5 minutes to obtain a rate for background activity. After adding ATP, absorbance was measured for 20 minutes or until absorbance readings reached a baseline. Absorbance values were plotted against time (mins) using GraphPad prism and reaction rates, before and after ATP addition, were calculated using regions of constant slope (Fig. 5.3.6). Background rates were subtracted from post-ATP reaction rates to obtain initial velocities (V_o) for each substrate concentration (n=3). A decrease in absorbance was indicative of PanK activity. Initial velocities of the mutant rec*Tc*PanK-domain were calculated for 0.5, 1, 2 and 3 mM ATP using 4 mM pantothenate and 8 μg purified protein. At each ATP concentration tested, MT *Tc*PanK-domain was >90% less active than WT rec*Tc*PanK (Table 5.3.2, Figs. 5.3.6-7).

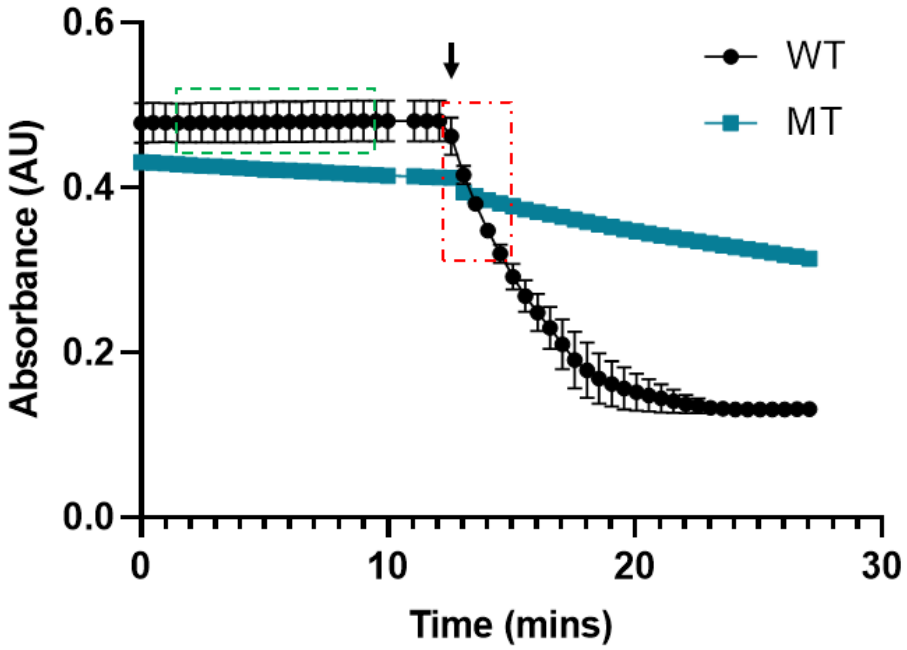


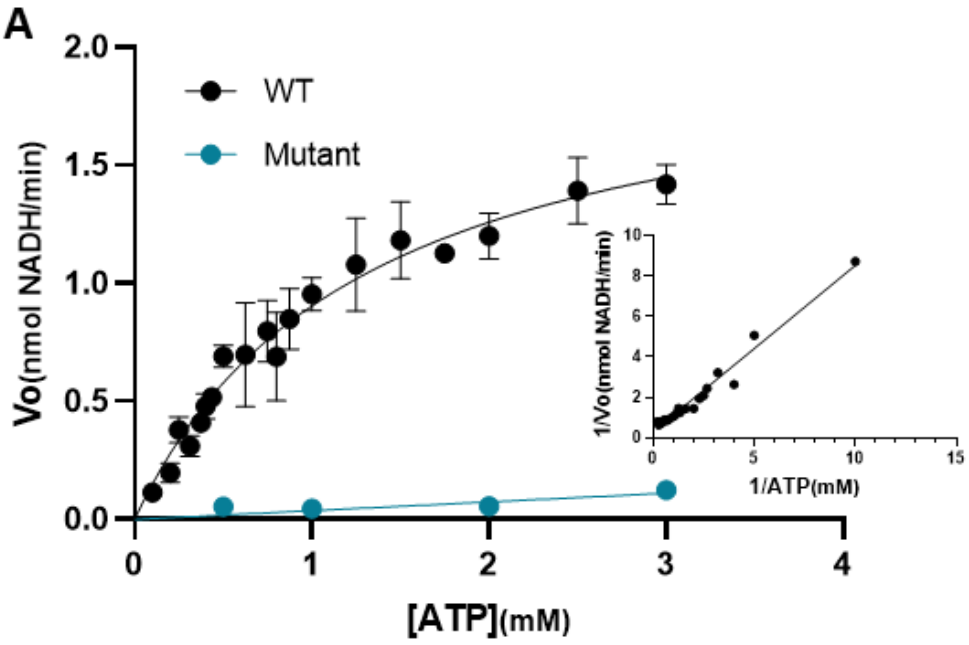
Fig. 5.3.6. Activity of recTcPank-Domain: *Plot of 340 nm absorbance change over time for assay reactions (n=3) containing 8 µg recombinant protein, 3 mM ATP and 4 mM pantothenate. Arrow indicates addition of ATP to initiate the reaction. WT protein refers to reactions containing wildtype recTcPanK-domain, MT is the mutant recTcPanK-domain containing an R1270A substitution. Boxes indicate slope regions from which background (green) and post-ATP (red) WT reaction rates were calculated. Error bars indicate SD.*

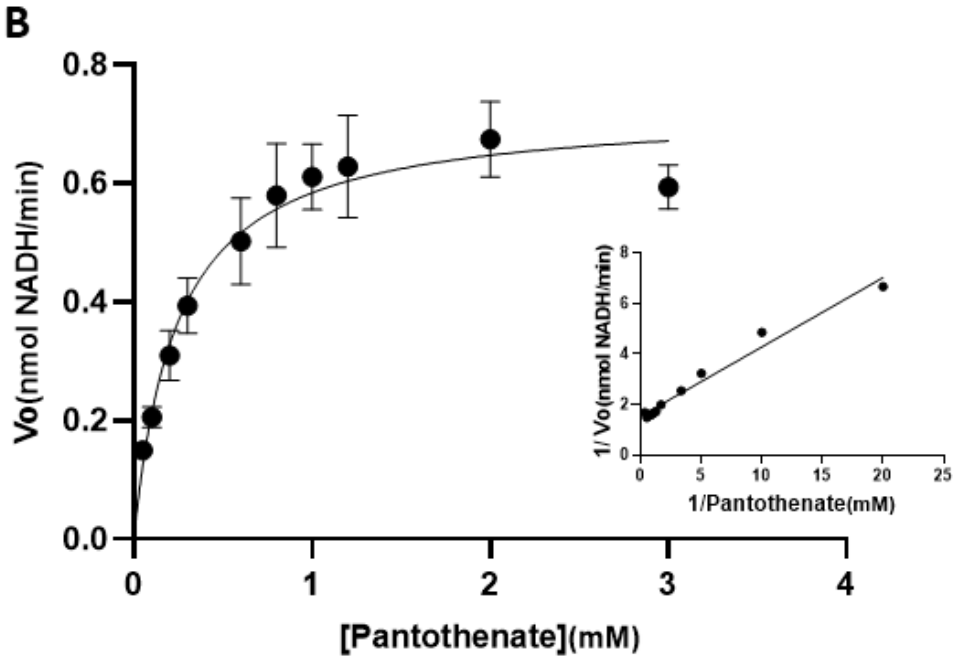
Table 5.3.2. Initial rate (V_o) of MT and WT recTcPanK-domain at different ATP concentrations

[ATP] (mM)	V_o MT (nmol NADH/min)	V_o WT (nmol NADH/min)
0.5	0.054	0.56
1	0.046	0.95
2	0.056	1.2
3	0.12	1.42

To calculate kinetic parameters for the WT *Tc*PanK-domain, mean initial velocities (n=3) were plotted (in units of nmol NADH/min) against ATP or pantothenate concentrations (mM) and curve-fitting was performed (Fig. 5.3.7 A-C). Non-linear regression curves were fitted to the data using the Michaelis-Menten model: $\mathbf{Y = V_{max}*X/(K_m + X)}$, where Y is initial velocity (nmol NADH/min), V_{max} is the maximum enzyme velocity (nmol NADH/min) , X is substrate concentration (mM) and K_m is the Michaelis-Menten constant (the substrate concentration (mM) required to achieve a half-maximum enzyme velocity). Based on this curve, V_{max} was more than twice as high for ATP (1.79 nmol NADH/min) than for pantothenate (0.73 nmol NADH/min). K_m was x4.1 higher for ATP (0.99 mM) than for pantothenate (0.24 mM). The number of substrate molecules converted to product per minute (K_{cat}) was calculated as $\mathbf{V_{max}/E_t}$. E_t (concentration of catalytic sites) assumed a molecular weight of 51.3 kDa (based on

amino acid sequence) and 8 µg enzyme per reaction. The derived values indicate that substrate turnover is more than twice as high for ATP (K_{cat}=11.48min^{-1}) than for pantothenate (K_{cat}=4.65min^{-1}). Lineweaver-Burk plots (1/Vo against 1/[substrate]) were generated for both substrate datasets to demonstrate that each produced a linear curve, consistent with Michaelis-Menten kinetics.





C

	ATP	Pantothenate
Km(mM)	0.99(0.81-1.2)	0.24(0.19-0.31)
Vmax(nmol NADH/min)	1.79(1.66-1.94)	0.73(0.68-0.78)
Kcat(min^{-1})	11.48(10.62-12.44)	4.65(4.33-5.01)

Fig 5.3.7 Biochemical characterization of recTcPanK-domain: (A) *Michaelis–Menten curves for ATP with 8 µg of WT recTcPanK-domain (residues 1044-1480) or recTcPanK-domain R1270A mutant. Inset: Lineweaver–Burk plots for ATP with WT protein* ***(B)*** *Michaelis–Menten curves for pantothenate with 8 µg of WT recTcPanK-domain. Inset: Lineweaver–Burk plots for pantothenate with WT protein Error bars in* ***A*** *and* ***B*** *represent SD.* ***(C)*** *Kinetic parameters (Km, Vmax and Kcat) for WT recTcPanK-domain, with 95% Confidence intervals provided in brackets. Reactions were performed in triplicate.*

Characterised PanKs from other organisms have been reported to be regulated by CoA, acetyl-CoA, and other acyl CoAs [223,224,294]. Crystallographic studies of human PanK3 indicated that the adenine and pantothenate moieties of acetyl-CoA (Fig. 5.3.8) competitively inhibit PanK activity by occupation of the ATP and pantothenate binding pockets respectively [239]. Binding of Acetyl-CoA induces a conformational change in human Pank3 that stabilises the inactive conformation. To further characterise rec*Tc*PanK-domain we determined the extent of feedback inhibition of this enzyme by pathway product CoA and the derivative acetyl-CoA.

Inhibition assays were performed using serial dilutions of each compound, within the range of 0-24 μM. Each reaction also contained 1 mM PEP, 33U PK/LDH, 200 μM NADH, 10 mM $MgCl_2$, 100mM Tris pH 7.5, 8 μg purified rec*Tc*PanK-domain and saturating concentrations of pantothenate (4 mM) and ATP (2 mM). Each inhibitor concentration was assayed in duplicate (CoA) or triplicate (acetyl-CoA) and inhibition values were estimated from nonlinear fitted curves using GraphPad Prism.

As is evident from Fig. 5.3.9, *Tc*PanK was strongly inhibited by CoA and even more strongly by acetyl-CoA. The IC50, which is the concentration of inhibitor required to reduce activity to half of the uninhibited value, was calculated using the model:

$$R = 1 \div 1 + (IC50 \div [l])^n$$

Where R is the response (the fraction of the biological activity that has been inhibited), $[l]$ is the inhibitor concentration and n is the Hill coefficient (which describes how steep the curve is).

At saturating substrate concentrations, IC50 values were twice as low for acetyl-CoA (0.31 μM) than for CoA (0.66 μM), suggesting that acetyl-CoA it is about twice as inhibitory (Table 5.3.3). The inhibitor constant (K_i), a more accurate measure of dissociation which accounts for tight binding, was calculated for each substrate based on the model:

$$Q = \left(Ki \times \left(1 + (S \div Km)\right)\right)$$

Where S is substrate concentration.

$$Y = Vo \times \left(1 - \left(\left(\left((Et + x + Q) - \left(((Et + x + Q)^2) - 4 \times Et \times x\right)^{0.5}\right)\right) \div (2 \times Et)\right)\right)$$

Based on a graph of inhibitor concentration (x) plotted against enzyme activity (Y).

For each substrate, the following constrained values were applied for each substrate:

ATP: Et =1.04 μM, S =2000 μM, and Km =990 μM

Pantothenate: Et =1.04 μM, S =4000 μM, and Km =240 μM

For both inhibitors the Ki for ATP (CoA: 151.5 nM; acetyl-CoA: 6.3 nM) was around 6x higher than for pantothenate (CoA: 25.9 nM; acetyl-CoA: 1.1 nM) at saturating concentrations of both substrates. Ki values suggest acetyl-CoA to be around 25x more inhibitory than CoA.

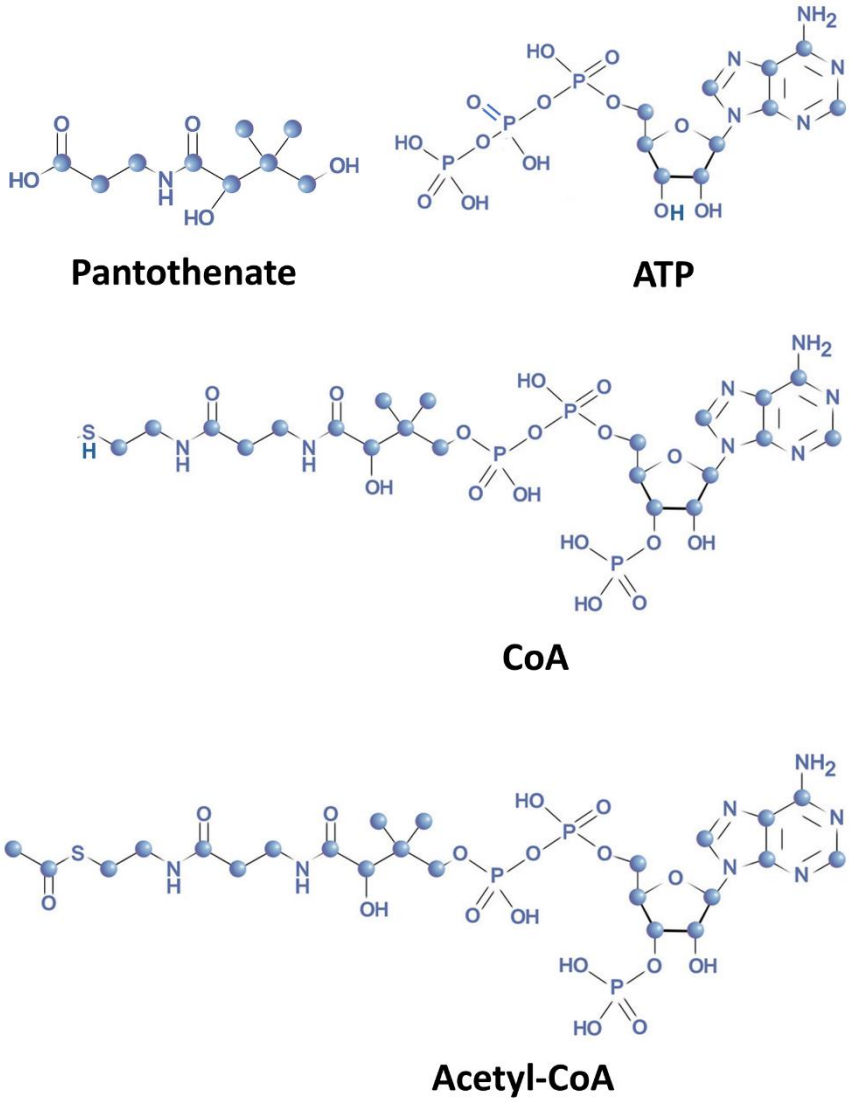


Fig. 5.3.8. Structures of PanK substrates and feedback inhibitors CoA and acetyl-CoA. *Structures were obtained or modified from https://biologydictionary.net.*

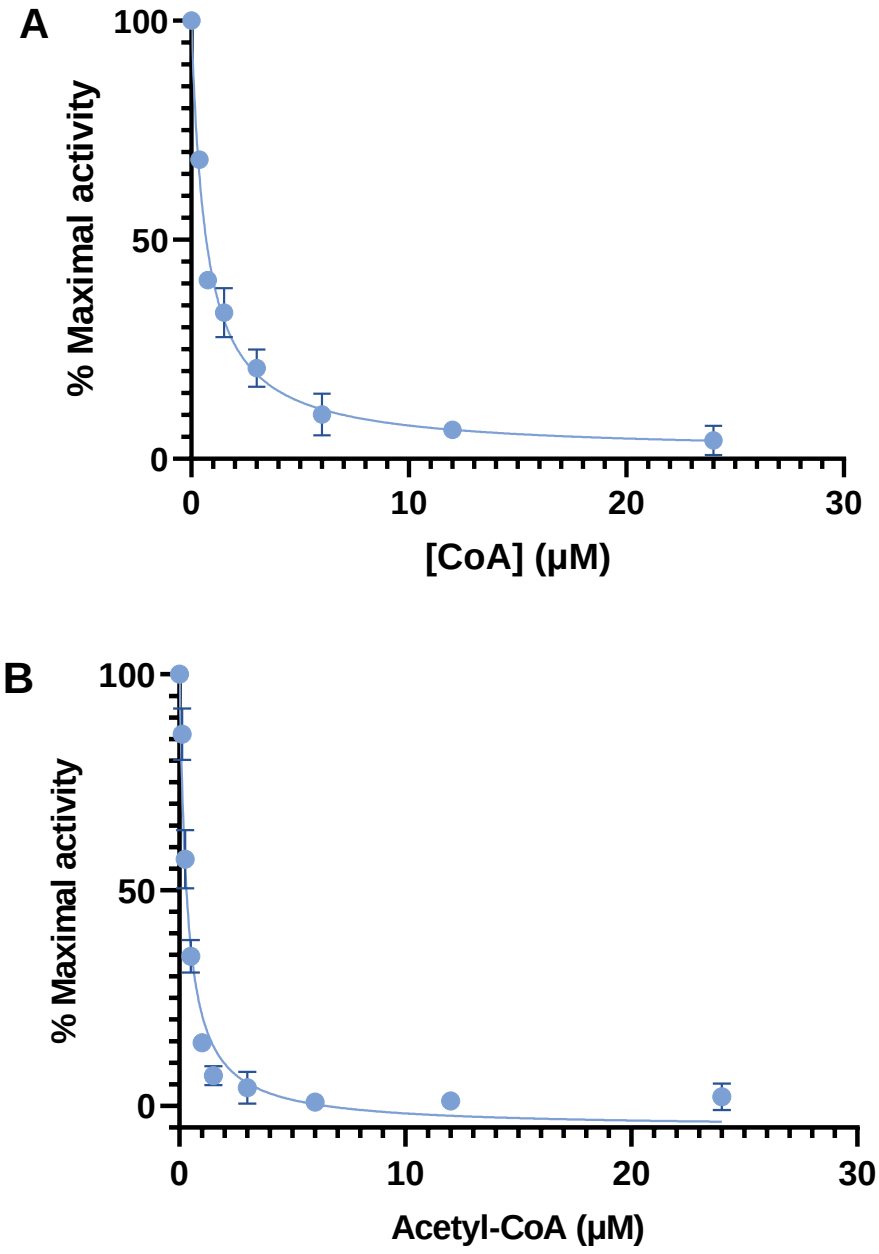


Fig. 5.3.9. Feedback inhibition of TcPanK: *Effect of CoA* ***(A)*** *and acetyl-CoA* ***(B)*** *on WT recTcPanK-domain activity. Relative activities in the presence of various concentrations of inhibitors to those without inhibitors are shown. The assay was performed in 100 mM Tris, pH 7.5, 10 mM $MgCl_2$, 1 mM PEP, 33U PK/LDH, 200 µM NADH, 4 mM pantothenate, and 2 mM ATP with various concentrations of CoA or acetyl-CoA at 24°C.The assay was carried out two (CoA) to three (acetyl-CoA) times independently, and the results are shown as means ± SEM.*

Table. 5.3.3. Inhibitory parameters of CoA and acetyl-CoA. *Measurements of WT recTcPanK-domain inhibition by CoA and acetyl-CoA. 95% confidence intervals are indicated in brackets.*

	CoA	**Acetyl-CoA**
IC50 (nM)	660 (487.6-902.2)	311.3 (247.9-393)
Ki(ATP) (nM)	151.5 (88.7-250.8)	6.3 (0.3-18.7)
Ki(Pantothenate) (nM)	25.9 (15.2-42.9)	1.1 (0.06-3.2)

5.4 Coenzyme A rescue of R1270A *T. brucei* clones

Having shown that R1270 is essential for optimal rec*Tc*PanK-domain activity and that *T. brucei* clones expressing *Tc*PanKR1270A fail to undergo normal proliferation in the absence of *Tb*PanK, we asked whether growth of these mutants might be restored with Coenzyme A (CoA) supplementation. The purpose of this experiment was to better understand the function of *Tc*PanK domains. Specifically, we were asking the question: do the fused domains, which are predicted to possess phosphodiesterase and adenylation activities, play a role in CoA metabolism?

The same approach was used as described previously (section 4.5). This utilised bloodstream form *T. brucei* that had been modified to facilitate constitutive expression of a mutated version of *Tc*PanK (R1270/A transition) and inducible RNAi-mediated knockdown of *Tb*PanK. The rescue experiment was performed by measuring cumulative cell densities of two R>A mutant clones (c1 and 2) over 5 consecutive 24 hour time-points (Fig. 5.4). One experimental group for each clone received 1 mM CoA at the time of RNAi induction (T=0) by tetracycline (1 µg/ml). Two control groups were included per clone: one receiving tetracycline alone (R>A +TET) and another which received neither compound (R>A). The routinely heat-inactivated (HI) serum was used in all growth media. Consistent with previous RNAi experiments, the untreated control groups grew at a healthy exponential rate (doubling time of 6-8 hours). In comparison, doubling times of the R>A +TET groups ranged from 18-26 hours and CoA-treated groups doubled in 8-9 hours, over the period of 120 hours.

These data indicate that addition of CoA is able to partially rescue the growth defect that results from RNAi-mediated knockdown of the full-length *Tb*PanK in a background where the *T. cruzi* R1270A mutated version of the multi-domain enzyme is being expressed.

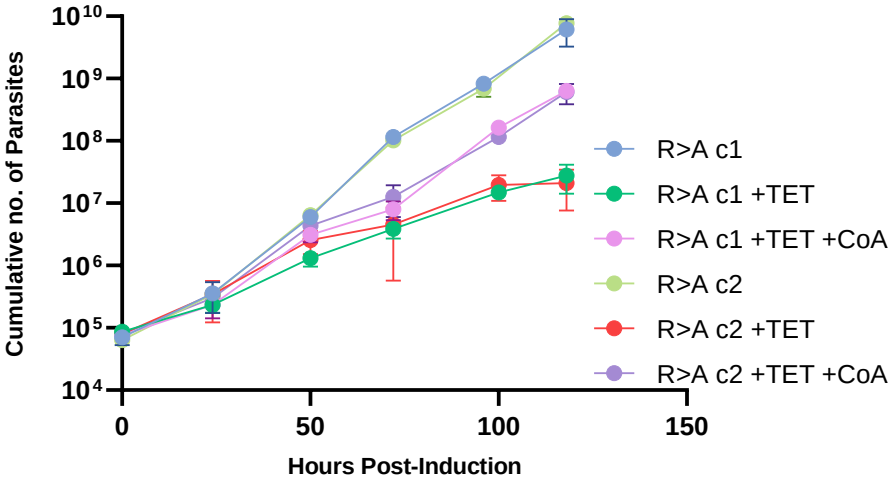


Fig. 5.4. Coenzyme A supplementation partially rescues growth arrest induced by TbPanK depletion: *T. brucei* clones *containing the TbPanK depletion construct (section 4.4) and which constitutively expressed the R1270A TcPanK mutant construct (R>A c1-2) were grown in the presence or absence of tetracycline (TET, 1 µg/ml) and Coenzyme A (CoA, 1 mM). Cumulative cell density of each experimental group was measured in triplicate.*

5.5 The PanK domain alone is insufficient to maintain normal growth of *T. brucei*

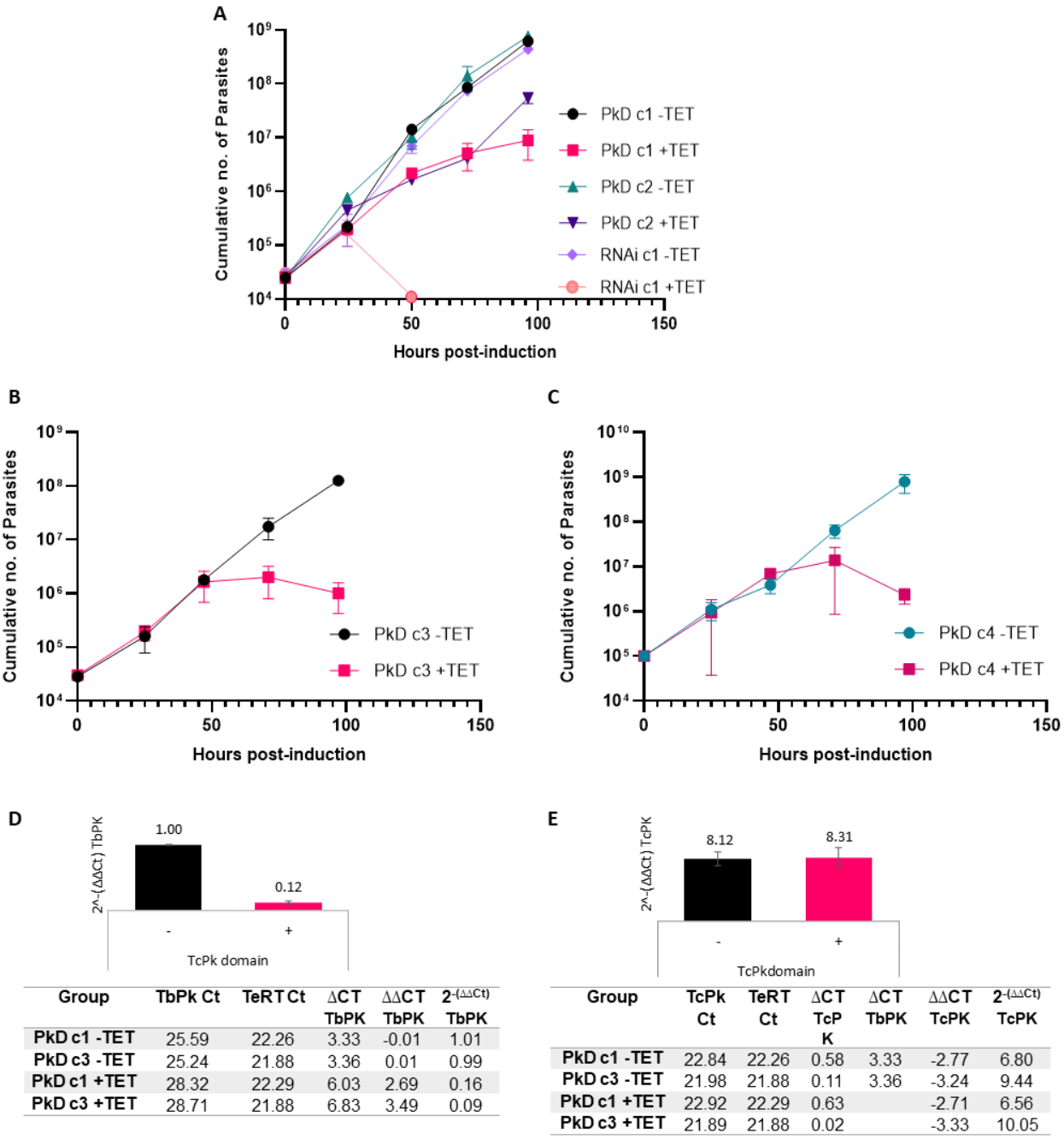
The question of whether the pantothenate kinase domain is essential was followed by 'is it sufficient?'. By expressing only the PanK domain (residues 1044-1480) of *Tc*PanK in the RNAi cell line, we aimed to establish whether this activity alone can sustain 'normal' growth of these parasites or whether the fused domains contribute to the essential roles of *Tc*PanK.

The PanK domain (nucleotides 3130-4443), which was demonstrated to be active in recombinant protein assays, was amplified from cloned *Tc*PanK template using primers which inserted SbfI (5') and AscI (3') restriction sites and a preceding start codon (Appendix Table A.3). It was cloned into the pTubEX vector and transfection of the *T. brucei* RNAi cell line was performed as previously described (section 5.2).

Two transformed *T. brucei* clones (PkD c1 and 2) containing the active *Tc*PanK domain in addition to the *Tb*PanK RNAi hairpin fragment were subjected to growth curve analysis alongside the parental RNAi cell line (RNAi c1). The results, presented in Fig. 5.5 A, showed as expected that parasite proliferation was impeded in the absence of the multidomain *Tc*PanK and knockdown of native *Tb*PanK, with cell death occurring within 48 hours post-induction. The effect of RNAi induction on the PkD clones that expressed the single *Tc*PanK domain was mild in comparison. In the case of PkD c2, parasite growth rate returned to normal levels in

the 72-96 hour period, although this was possibly due to an outgrowth of revertants that had lost expression of the hairpin fragment, a common outcome after several days growth in this type of RNAi experiment. The transfection was repeated to derive another two independent clones (PkD c3 and 4), which were also monitored by growth curve analysis (Fig. 5.5 B and C). RNAi induction in both clones led to a growth plateau by 72 hours post-induction, followed by a decline in cell density by 96 hours. Depletion of *Tb*PanK mRNA in induced cell lines was confirmed by qPCR (Fig.5.5 D). *Tb*Pank mRNA levels were 85-91% lower in induced clones PkD c1 and c3. Expression of the *Tc*PanK domain construct was indicated by quantification of mRNA in add-back cell lines using primers which annealed to the *Tc*PanK domain (Fig. 5.5 E) Add-back clones contained 7 to 10-fold more *Tc*PanK than *Tb*PanK transcript.

This experiment therefore demonstrates that expression of the catalytically active *Tc*PanK domain on its own is able to prevent rapid cell death, but is not sufficient to support long-term growth in this context.



Group	TbPk Ct	TeRT Ct	ΔCT TbPK	ΔΔCT TbPK	$2^{-(\Delta\Delta Ct)}$ TbPK
PkD c1 -TET	25.59	22.26	3.33	-0.01	1.01
PkD c3 -TET	25.24	21.88	3.36	0.01	0.99
PkD c1 +TET	28.32	22.29	6.03	2.69	0.16
PkD c3 +TET	28.71	21.88	6.83	3.49	0.09

Group	TcPk Ct	TeRT Ct	ΔCT TcPK	ΔCT TbPK	ΔΔCT TcPK	$2^{-(\Delta\Delta Ct)}$ TcPK
PkD c1 -TET	22.84	22.26	0.58	3.33	-2.77	6.80
PkD c3 -TET	21.98	21.88	0.11	3.36	-3.24	9.44
PkD c1 +TET	22.92	22.29	0.63		-2.71	6.56
PkD c3 +TET	21.89	21.88	0.02		-3.33	10.05

Fig. 5.5 TcPanK single domain complementation partially rescues TbPanK depletion: (A-C) *Growth curves showing the effect of tetracycline (TET, 1 µg/ml)-induced TbPanK depletion on cell lines containing RNAi constructs either alone (RNAi c1) or in addition to a TcPanK domain complementation construct (PKDc1-4) which causes constitutive expression of residues 1044-1480. Each data point is the mean of three wells, counted in triplicate. Error bars indicate SD. (**D-E**) qPCR data for RNA extracted from PkD complementation clones at 16 hours post-tetracycline. **(D)** Depletion of TbPanK (TbPK) transcript with tetracycline treatment (+, +TET) relative to untreated groups (-, -TET), using telomerase reverse transcriptase (TeRT) for normalisation. **(E)** Levels of TcPanK (TcPK) in add-back cells, relative to TbPK levels in uninduced parasites. Bar charts indicate mean 2^-(ΔΔCt) and standard error for each experimental group (2 clones per group) and tables provide mean Ct values per clone. Calculations are based on triplicate cDNA samples from each experimental clone.*

5.6 Does the phosphodiesterase domain contribute to *Tc*PanK essentiality?

Since the fused domains in PanK may contribute to its essential function in *T. brucei*, we sought to determine the contribution of the predicted phosphodiesterase activity. To do this I mutated residues with predicted essentiality and delivered the mutant construct to the *T brucei* RNAi cell line. As with previous experiments, I assessed the impact of the mutations by the ability of the construct to rescue RNAi-induced growth inhibition.

As established in chapter 3, residues 450-1040 of *Tc*PanK are predicted to form a phosphodiesterase belonging to the endo/exonuclease/phosphatase (EEP) enzyme family. Proteins of this family include DNA repair enzymes and sphingomyelinases and share a common β-sandwich architecture made up of central β-sheets surrounded by α-helices. They also share a requirement for divalent metal ion cofactors for catalysis. One characterised EEP-family member is neutral sphingomyelinase of *B. cereus* (uniprot ID: P11889 [283]). Crystal structure analysis of this protein in complex with cobalt revealed that two cobalt ions (Co^{2+}) and coordinated water molecules bind to conserved residues within a central cleft (refer to Chapter 3; Fig. 3.3.5).

The *B. cereus* structure was aligned with the predicted phosphodiesterase of *Tc*PanK to visualise residues likely to be involved in metal-binding (Fig. 5.6.1). Four highly conserved motifs characterise metal-dependent phosphodiesterases: S/TxN, xQ/NE, GDxN and SDH (refer to Chapter 3; Fig. 3.3.4) and in *Tc*PanK these conserved residues appear to form a metal-binding site at the opening of two central β-sheets, which are surrounded by two α-helices on either side (Fig. 5.6.1). In the *B. cereus* enzyme-Co^{2+} complex, residues H296 and E53 each bind directly to a Co^{2+} and residues N197, D195 and D295 form hydrogen bonds with ligand water molecules (Chapter 3; Fig. 3.3.5 and Fig. 5.6.1C). Based on structural alignment, *Tc*PanK residues H332 and E77 are Co^{2+} ligands and D230 and D331 are water ligands. The predicted ligands D331 and H332 (of the SDH motif) were chosen for mutagenesis based on the extent of conservation, being conserved in human, bacteria and plant phosphodiesterases (Fig. 5.6.2A). Bond formation by these amino acids depends on charged side-chain atoms: two charged oxygens of D331 and an imidazole nitrogen of H332. Therefore, we hypothesised that substituting both residues with alanine, which has a small, neutral, hydrophobic side chain, would impede stabilisation of the metal cofactors and consequently disrupt catalysis.

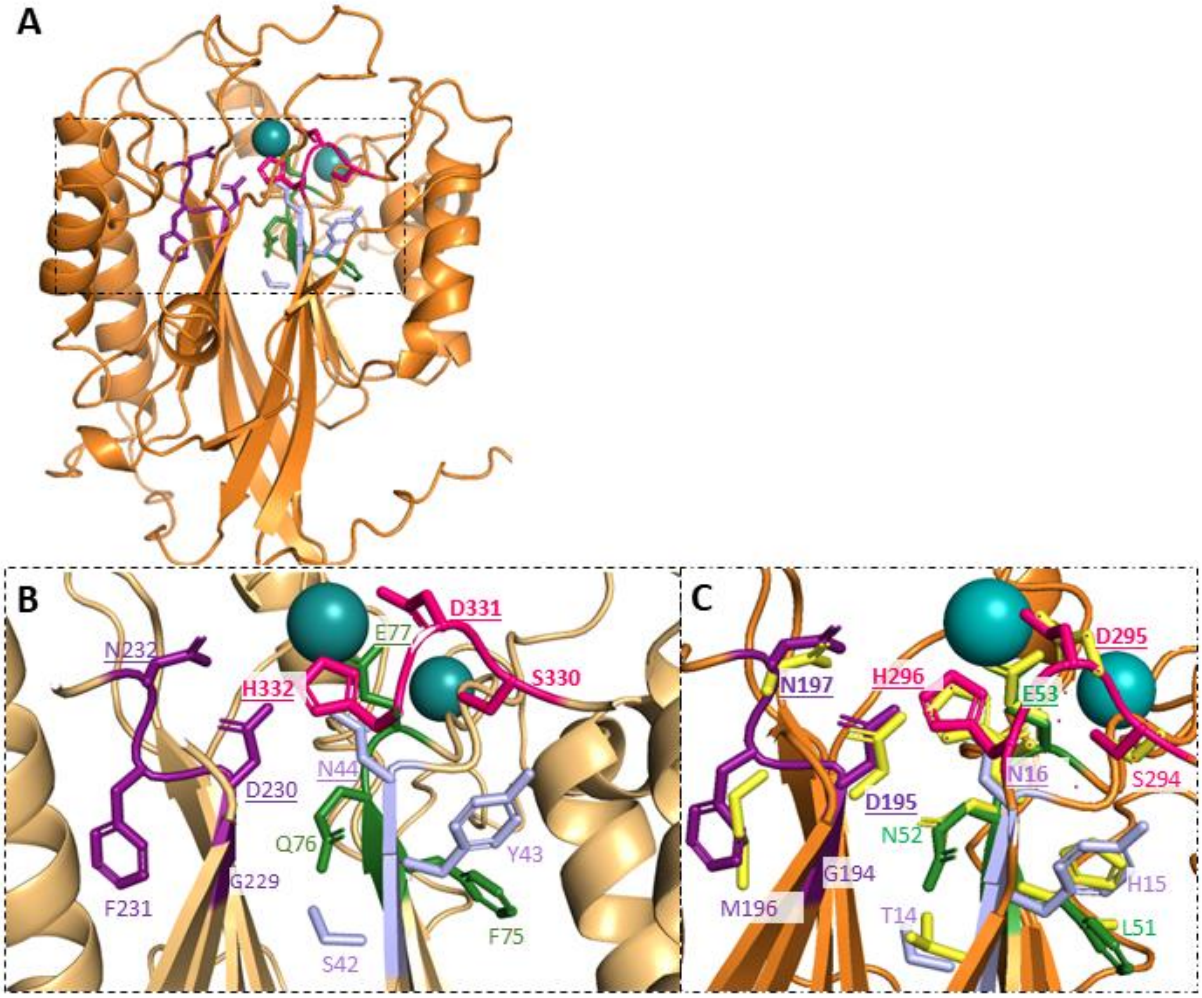


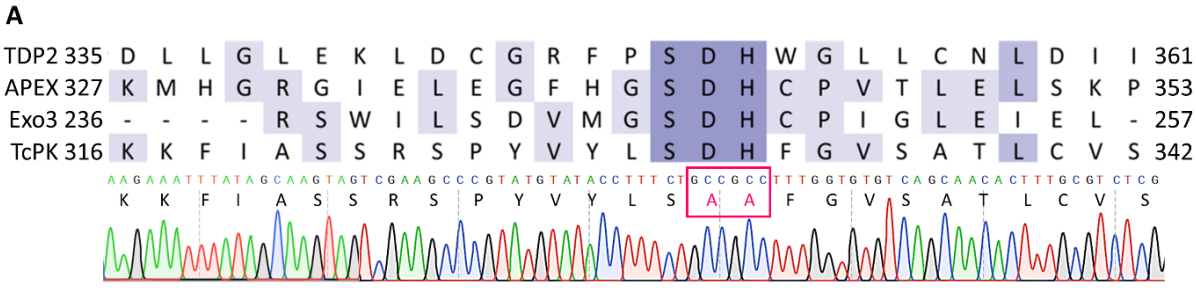
Fig. 5.6.1 Active site interactions of the predicted phosphodiesterase structure of TcPanK: *TcPank contains a predicted phosphodiesterase (residues 450-1040)**.*** *The AlphaFold predicted model of this domain was aligned with the structure of B. cereus neutral sphingomyelinase (nSMase) complexed with cobalt (P11889) to visualise interactions within the metal binding site.* ***(A)*** *The predicted phosphodiesterase domain of TcPanK showing the location of the metal binding site. Cobalt ions are represented as blue spheres.* ***(B)*** *A close up view of the putative metal binding site of TcPanK showing conserved motifs SYN, FQE, GDFH and SDH coloured light purple, green, dark purple and pink respectively and cobalt ions (blue) of the aligned B cereus structure. Underlined residues N44, E77, D230, N232, D331 and H332 are predicted to interact directly with the metal substrate or metal-coordinated water molecules, based on alignment with the B. cereus enzyme* ***(C)*** *B. cereus nSMase residues (yellow) aligned with TcPanK SYN, FQE, GDFH and SDH motifs (light purple, green, dark purple and pink respectively). Residue labels indicate B. cereus nSMase residues. Crystal structure analysis (Ago et al., 2006* [283]*) indicated that underlined residues interact with cobalt (blue spheres) or co-ordinated water molecules (not shown). Figure generated using Pymol.*

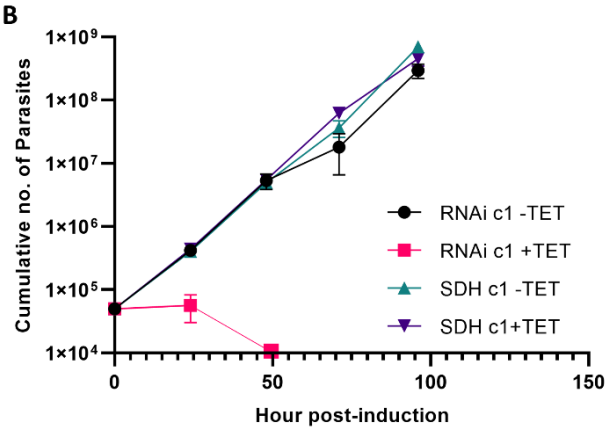
We incorporated D331A and H332A substitutions into the *Tc*PanK gene by PCR mutagenesis (Materials and Methods 2.3.5; appendix Table A.3). This time the pTubEX plasmid containing WT *Tc*PanK was used as a template with primers which amplified the entire plasmid and incorporated alanine substitutions in place of D331 and H332 in a single reaction. The PCR product was treated with Dpn1, phosphorylated and ligated and directly used for transformation. Successful cloning was confirmed by sequencing (Fig. 5.6.2 A) before the

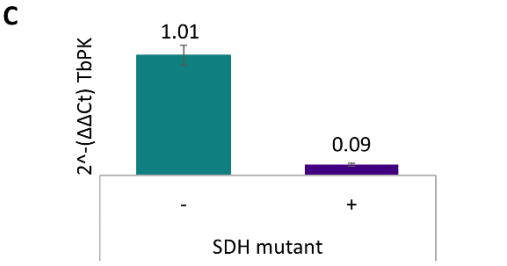
construct was used to transfect *T. brucei* RNAi cell lines containing the *Tb*PanK RNAi construct as described previously (section 5.2). The resulting cloned cell lines were designated SDH c1 and c2.

Transformed bloodstream form *T. brucei* were then assessed for growth inhibition in response to tetracycline-induced *Tb*PanK depletion (Fig.5.6.2 B). SDH clone 1 (SDH c1), which expressed the mutated *T. cruzi* gene, and the parental clone (RNAi c1) were seeded in triplicate wells and cumulative cell density (CCD) was measured as before at 24 hour intervals for 96 hours, in the presence or absence of tetracycline. The induced SDH group grew at the same rate as the un-induced group, indicating that the mutated *Tc*PanK was able to rescue the growth inhibition caused by *Tb*PanK depletion. Tetracycline induction of RNAi resulted in death of the parental cell line with transcript depletion confirmed by qPCR quantification (Fig.5.6.2 C). Induced clones SDH c1 and 2 contained 90-92% less *Tb*Pank mRNA than uninduced groups, respectively. Prior to induction, the SDH clones contained 3 to 8-fold more *Tc*PanK than *Tb*PanK transcript, (Fig. 5.6.2 D).

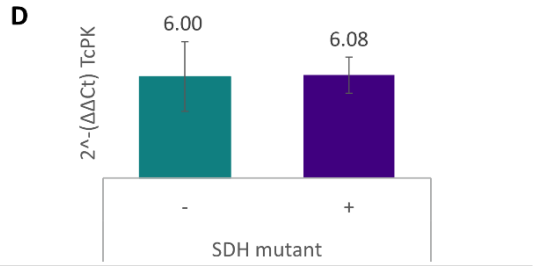
These results therefore show that the SDH motif does not contribute to the essential function of multi-domain *Tc*PanK and suggest that the activity of the predicted phosphodiesterase is not fundamental to *Tc*PanK activity, at least in the bloodstream stage of *T. brucei*.







Group	TbPk Ct	TeRT Ct	ΔCT TbPK	ΔΔCT TbPK	$2^{-(\Delta\Delta Ct)}$ TbPK
SDH c1 -TET	24.87	20.76	4.11	0.17	0.89
SDH c2 -TET	24.80	21.03	3.77	-0.17	1.13
SDH c1 +TET	28.89	21.15	7.74	3.80	0.07
SDH c2 +TET	28.24	21.10	7.14	3.21	0.11



Group	TcPk Ct	TeRT Ct	ΔCT TcPK	ΔCT TbPK	ΔΔCT TcPK	$2^{-(\Delta\Delta Ct)}$ TcPK
SDH c1 -TET	23.08	20.77	2.31	4.11	-1.63	3.09
SDH c2 -TET	21.85	21.07	0.78	3.77	-3.15	8.90
SDH c1 +TET	22.94	21.19	1.74		-2.19	4.57
SDH c2 +TET	22.12	21.11	1.01		-2.93	7.60

Fig. 5.6.2 Complementation with the TcPanK phosphodiesterase mutant rescues TbPanK depletion: (A) *The SDH metal-binding motif is conserved in human tyrosyl-DNA phosphodiesterase 2 (TDP2, O95551), APEX1 of A. thaliana (APEX, P27695), thermophilic exonuclease III of Methanothermobacter thermautotrophicus (Exo3, 3FZI) and the predicted phosphodiesterase domain of TcPanK (TcPK). As the chromatogram shows, residues D331 and H332 of TcPanK were substituted for alanine in the phosphodiesterase mutant complementation construct* ***(B)*** *Growth curve showing the effect of tetracycline (TET, 1 µg/ml)-induced TbPanK depletion on cell lines containing RNAi hairpin constructs, without (RNAi c1) or in combination with (SDH c1), a TcPanK complementation construct containing mutations (D331A and H332A) in the phosphodiesterase domain. For comparison, groups were included which received no tetracycline (-TET). Each data point is an average of three wells, counted in triplicate. (**C-D**) qPCR data for RNA extracted from SDH mutant clones (SDH c1 and c2) at 16 hours post-induction.* ***(C)*** *Depletion of TbPanK (TbPK) transcript with tetracycline treatment (+, +TET) relative to untreated groups (-, -TET), using telomerase reverse transcriptase (TeRT) as a house-keeping gene.* ***(D)*** *Levels of TcPanK (TcPK) in add-back cells, relative to TbPK levels in uninduced parasites. Bar charts indicate mean 2^-(ΔΔCt) and standard error for each experimental group (2 clones per group) and tables provide mean Ct values per clone. Calculations are based on triplicate cDNA samples from each experimental clone.*

5.7 Summary

The use of our *T. brucei* RNAi model with constitutively expressed add-back constructs revealed the importance of specific residues and whole domains within the multifunctional *T. cruzi* pantothenate kinase. These experiments demonstrated that:

- Substitution of the conserved PanK active site residue R1270 by a non-polar alanine, disrupted rescue of RNAi lethality, suggesting that this amino acid is necessary for normal proliferation of *T. brucei*.

- The R1270A substitution ablates PanK activity in assays of recombinant *Tc*PanK.

- *T. brucei* add-back strains which lack PanK activity due to the R1270A substitution, are partially rescued from RNAi-mediated growth deficiency in the presence of CoA.

- Residues 1074-1480 of *Tc*PanK encode a functional pantothenate kinase which retains catalytic activity in the absence of the N-terminally fused domains that form part of the expressed polypeptide in kinetoplastids.

- Expression of the active PanK domain alone (residues 1074-1480) was not sufficient to sustain normal growth in the RNAi model, indicating that the predicted phosphodiesterase and adenylation functions may have important roles in these parasites.

- Add-back of a mutated *Tc*PanK in which alanines were substituted for residues D331 and H332, which are highly conserved among metal-dependent phosphodiesterases, suggested that this activity is non-essential when the PanK domain is intact and active.

The importance of the central putative adenylation domain was not assessed during this study. Whether lack of this activity explains the insufficiency of the isolated PanK domain to sustain parasite growth remains to be determined.

6. Discussion and Conclusions

Key findings of this study

In this work, I have identified putative components of the CoA biosynthesis pathway in the genomes of clinically important trypanosomatids *T. cruzi, T. brucei* and *L. major*. Unlike bacteria, plants, certain fungi and coccidia, **kinetoplastids appear to lack the capacity for *de novo* pantothenate synthesis**, but their genomes seem to encode the required enzymes to synthesise CoA from the vitamin pantothenate, which is **acquired from hosts, potentially via a proton-dependent major facilitator superfamily transporter**. A homologous transporter of this class is absent from humans, as mammals utilise a sodium-dependent multivitamin transporter (SLC5A6) for pantothenate uptake. Another component of CoA biosynthesis that differs from mammalian counterparts is pantothenate kinase (PanK). This enzyme has gained substantial interest in the context of drug discovery, having been shown to be the rate-limiting enzyme of CoA synthesis in bacterial and eukaryotic pathogens. Sequence and structure alignments revealed that **kinetoplastids have two unique domains at the N-terminus of PanK**, predicted to have phosphodiesterase and adenylation activities. All three domains are highly conserved amongst all kinetoplastids, which suggests that they play important roles.

Features of the multi-domain PanK in trypanosomatids

The most similar proteins to the unique N-terminal *Tc*PanK domain (residues ~38-340), are neutral sphingomyelinases (nSMases) based on conservation of amino acid sequences and predicted structures. This raises the possibility that this domain catalyses the hydrolysis of sphingomyelin (SM) or inositol phosphorylceramide (IPC). This would be a redundant activity as far as sphingolipid metabolism is concerned, as trypanosomes possess functional SMases encoded elsewhere in the genome. Several DNA-binding enzymes such as tyrosyl-DNA phosphodiesterase 2 (TDP2), deoxyribonuclease 1 (DNase1) and the DNA endonuclease APEX1 also share conserved sequence motifs and structural folds with this domain. Based on these similarities, it is highly probable that the activity involves hydrolysis of a nucleotide or other phosphodiester-bond-containing substrate, using a divalent metal ion co-factor, such as magnesium.

The central domain of *Tc*PanK (residues 450-1040) shows very limited conservation with other proteins at the sequence level. In spite of this, sequences and predicted structures for this region align to adenylating enzymes with a high degree of confidence, based on multiple scoring methods. One such enzyme is long-chain fatty acid CoA synthetase (LC-FACS), which

activates fatty acids for their use in lipid metabolism including the β-oxidation pathway [295]. Another is the adenylation domain of non-ribosomal peptide synthases. Both catalyse the activation of a carboxylate with AMP for transfer to the pantetheine group of coenzyme A or ACP.

By tagging PanK in *T. cruzi* epimastigotes using CRISPR-spCas9, I demonstrated that the enzyme is expressed as a multi-domain polypeptide in this life cycle stage, based on its mass of ~170kDa, which was consistently observed in three different tagging events (Fig. 4.2.3). Microscopy analysis of epimastigotes tagged at the PanK carboxyl terminus by insertion of Ty and PTP peptides, revealed punctate staining throughout the cytoplasm (Fig. 4.2.4). The fact that signal was obtained for PTP-Ty-tagged *Tc*PanK (tagged at C terminus) but not for mNG-Ty tagging of the N- terminus, using anti-Ty antibody in both cases, suggests that the 3'-UTRs may impact expression levels. The native 3'-UTR which is retained in the mNG-Ty tagged cell line may sustain lower *Tc*PanK expression levels compared with the paraflagellar rod (PFR2) 3'UTR inserted with the C PTP-Ty tag. While the outcomes of endogenous *Tc*PanK tagging suggest expression levels are low in epimastigotes, other life-stages may be better suited to localisation studies. Rick Tarleton's study on differential protein expression throughout the *T. cruzi* life cycle revealed that PanK RNA levels were higher in amastigotes compared to epimastigotes (http://trypsnetdb.org/interactions.aspx) [296]. The study also found PanK expression to be higher in *T. brucei* procyclic forms than in BSF, which was associated with motifs found in 3' regulatory elements.

The signal pattern observed from *Tc*PanK tagging is consistent with previously documented tagging outcomes in *T. brucei*, which suggested ER or cytoplasmic localisation, regardless of whether tagging was performed at N or C termini. Consistent with an ER or other membrane-bound location, *Tc*PanK is predicted to contain transmembrane (TM) domains by the deep learning algorithm DeepTMHMM (https://dtu.biolib.com/DeepTMHMM) [297]. This uses deep neural networks to predict TM topology within eukaryotic and prokaryotic protein sequences. Two TM alpha helices are predicted, encompassing residues 374-394 and 397-412, placing residues 395-396 outside a membrane and the remaining protein inside a membrane (Fig. 6.1). The putative TM regions sit between the predicted phosphodiesterase (residues ~38-340) and adenylation (residues 450-1040) domains of *Tc*PanK.

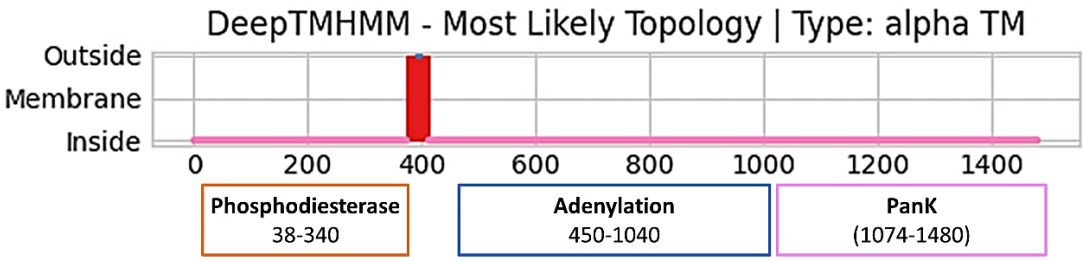


Fig. 6.1. Predicted transmembrane topology of TcPanK: *DeepTMHMM predicts 2 TM helices (shown in red) located between the putative phosphodiesterase and adenylation domains. Protein regions located inside or outside the membrane are shown in pink and blue respectively.*

The signal peptide prediction software PSORT II [270,298] does not predict any localisation signals including ER membrane retention signals, instead predicting a cytoplasmic location. However, the reliability of this prediction is questionable as the characterised ER retention motifs on which the model is trained are not a prerequisite for ER localisation [298,299]. Another predictive tool, available as a web service (http://134.36.66.166:8070/home) [300], reveals putative N-glycosylation sites within *Tb*PanK. These were predicted by a machine-learning algorithm trained on mass spectrometry data for *T. brucei* glycosylated proteins. Using *Tb*PanK (Tb11.v5.0254, TREU927 strain) as the query sequence, three putative N- glycosylation sites were predicted at residues 1180, 1363 and 1396. This may support an ER localisation of kinetoplastid PanKs, being a common site for protein glycosylation, although some proteins are glycosylated in the cytosol [301]. Most glycosylated proteins are trafficked to the plasma membrane or secretory pathways, however this would be in contrast with the results of *Tc*PanK and *Tb*PanK localisation and the essential requirement for CoA in intracellular pathways.

Most characterised PanKs are reported to be cytosolic [197], consistent with this being a major site of CoA metabolism (refer to Chapter 1.6.2). Some PanKs, including human isoforms, have been localised to mitochondria, which is metabolically convenient as the TCA cycle is mitochondrial in many eukaryotes and requires CoA derivatives [128,302]. Human mitochondria have been reported to contain the largest pool of intracellular CoA [303].

We did not rule out the possibility that the tag itself caused incorrect processing of PanK transcripts and thereby artefactual persistence of the three domains and/or mis-localisation of the protein. However, the fact that separately tagging each terminus produced a protein of the same size reduces this likelihood. Furthermore, experiments in *T. brucei* indicated that tagging of amino and carboxyl termini preserved *Tc*PanK function, based on normal proliferation of RNAi cell lines depleted in *Tb*PanK and constitutively expressing His or mNeonGreen-tagged *Tc*PanK constructs (Fig. 4.4.6).

Functional studies

The inability to replace *TcPanK* with a drug resistance cassette in epimastigotes suggested that the corresponding activity was essential for parasite viability (Fig. 4.3.2). While replacement of a single allele was achieved, null mutants could not be generated. This finding is commonly reported when targeting essential genes [304]. Southern blotting indicated that attempted dual allelic replacement selected for a cell line with a re-arranged PanK locus, that may have contained multiple duplicated PanK copies. This suggestion of essentiality was substantiated by RNAi targeting of *Tb*PanK in bloodstream form *T. brucei* (Fig. 4.4.3). The lethal effect of tetracycline on cell lines containing the RNAi construct was reproducible, as three independent clones displayed a comparable sensitivity. Specific depletion of *Tb*PanK mRNA was validated in two cell lines using qPCR, which indicated more than a 90% reduction by 16 hours-post induction, a reduced level that that was sufficient to have a lethal impact.

Next, I demonstrated that *Tc*PanK is functionally homologous to the *T. brucei* orthologue and rescues the lethality of *Tb*PanK depletion (Fig. 4.4.3). This allowed me to further characterise the *T. cruzi* protein by addback analyses, providing results relevant to Chagas disease drug development, where the current need is greater than for HAT.

Targeting of the CoA biosynthesis pathway in RNAi experiments was confirmed by the rescue of RNAi lethality by supplementation with 1 mM CoA (Fig. 4.5). In comparison to uninduced parasites, RNAi cells supplemented with CoA grew at a slower rate. This contrasts with the results of Fletcher and colleagues, who observed 70-95% rescue of growth inhibition by CoA pathway inhibitors using 1mM CoA[219]. However, the compounds used in this study appeared to target the final pathway enzymes PPAT and/or DPCK rather than acting on PanK. The slower growth rate observed in our rescue experiments raised the possibility of fused activities contributing to the essential role of the multifunctional PanK on proliferation. To investigate whether the incomplete rescue effect is due to absence of the fused activities, we performed CoA rescue using cell lines possessing the putative Pde and A, domains but lacking an active PanK. The rescued cells in this case still grew slower than uninduced parasites. In light of this, it is possible that CoA is not sufficiently stable to feed energetic requirements for exponential growth, or perhaps conversely there is an overabundance of CoA due to the absence of PanK, which we know is highly responsive to feedback inhibition.

Using the RNAi model, we showed that arginine 1270, within the broadly conserved PanK domain is essential to proliferation of *T. brucei* (Fig. 5.2.3). In contrast, serine 1254 is a non-essential residue. Both were chosen based on their predicted interactions with substrates

within the PanK active site. Based on alignment of the *Tc*PanK structure model with the human PanK3 crystal structure, we predict S1254 to form a single hydrogen bond with ATP, whilst R1270 forms two hydrogen bonds with pantothenate. The *in vitro* results indicate that the S1254 side-chain is non-essential to the orientation and stabilisation of the ATP·Mg^{2+} complex, therefore other active site residues such as E1200, which aligns with E138 in *hs*PanK3, and pantothenate itself, must provide a sufficient source of hydrogen bonds in its absence (refer to Fig 5.2.1, Chapter 5).

Biochemical studies

We confirmed that residues 1044-1480 of *Tc*PanK function as an active pantothenate kinase by expressing the partial 51 kDa domain in bacteria, purifying it using nickel affinity chromatography and then coupling its ATP-dependent phosphorylation of pantothenate to downstream pyruvate kinase and lactate dehydrogenase reactions (Fig. 5.3.5). The measured K_m and K_{cat} values are low compared with characterised PanKs of bacterial [235] and eukaryotic homologues [223,305] (Table 6). The presence of a small number of contaminant proteins in recombinant PanK samples after NAC purification could impact slightly on the calculated K_{cat} values, but not enough to account for these differences. The comparatively low activity may reflect sub-optimal expression of the eukaryotic protein using prokaryotes, which have a limited capacity for post-translational modification. A fungal PanK, belonging to *Aspergillus nidulans* was expressed using the same bacterial strain (BL21 DE3) and similar expression conditions to this study and was highly active [223]. Therefore, we would not expect the expression system to be limiting *TcPanK* activity to a large extent.

Difficulties in expressing the entire multi-functional PanK as a recombinant protein restricted interpretation of whether the absence of the fused Pde and A domains may impact PanK activity. However, this is certainly a possibility. The higher K_m for ATP (990 µM) compared to pantothenate (240 µM) is consistent with other characterised PanKs. The results of feedback inhibition, which showed acetyl-CoA to be around 25x more inhibitory than CoA, are in contrast to studies on *E. coli* Type 1 PanKs, which are inhibited more strongly by CoA [231], but are similar to findings for PanK II of *A. nidulans.* K_i for acetyl CoA was reported as 9 mM for the *A. nidulans* enzyme [223], which is >1000x higher than that measured for *Tc*PanK, suggesting that the partial kinetoplastid enzyme is hypersensitive to feedback inhibition. The inactivity of the mutant *Tc*PanK domain containing the R1270A substitution is consistent with findings from other studies, as the conserved arginine was shown to be essential in human PanKs [233,239].

**Table 6: Reported kinetic parameters of PanKs ***SD not reported

Isoform	Enzyme	Kcat (s^{-1})	Substrate	K_m (μM)	Source
Prokaryotic (Type I)	*Ec*CoaA	0.30 ± 0.13	Pantothenate	36 ± 4	Song and Jackowski, 1994 [235]
			ATP	136 ± 15	
Eukaryotic (Type II)	*An*PanK	1.95*	Pantothenate	60	Calder et al., 1999 [223]
			ATP	145	
	*Sa*CoaA	1.73* 1.56*	Pantothenate	23	Leonardi et al., 2005 [305]
			ATP	34	
	*Tc*PanK	0.10 0.20	Pantothenate	240	This study
			ATP	990	
CoaX-like (Type III)	*Bs*CoaX	2.12 ± 0.17	Pantothenate	168 ± 27	Brand and Strauss, 2005 [231]
			ATP	3050 ± 520	
	*Hp*CoaX	2.09 ± 0.26	Pantothenate	101 ± 26	
			ATP	9590 ± 2140	

Despite the measured activity of the recombinant *Tc*PanK domain and its proven essentiality to growth of BSF *T. brucei*, it was not sufficient for growth of these parasites in the absence of the additional N-terminally fused domains. This may indicate the importance of their activities or structural contributions, suggesting that characterisation of the functions of these domains may contribute to drug development against this protein, as well as informing the biology of the kinetoplastids. We did not rule out the possibility that the truncated protein mis-localises, which could be determined by complementation with a tagged *Tc*PanK domain.

Our observation that add-back of *Tc*PanK with substituted D331 and H332 residues complemented loss of *Tb*PanK, suggested that the predicted phosphodiesterase function of this enzyme is unlikely to play a fundamental role in growth of the bloodstream form parasite, unlike the PanK domain itself. The possibility that the central putative adenylation activity contributes to *Tc*PanK's role in parasite proliferation is worth investigation; however, a lack of conserved residues within this domain is a limiting factor.

Importance of the fused domains

The relationship between the predicted fused activities and the CoA pathway still remains to be established. Based on the bioinformatics predictions, it would be interesting to determine whether sphingolipid hydrolysis is linked to CoA biosynthesis by the multifunctional kinetoplastid PanK (kPanK). If the N-terminal domain functions as a nSMase or ISC and generates ceramide (N-acylsphingosine), then this could potentially be a source of acyl groups for CoA, which can occur via the sphingosine N-acyltransferase (EC 2.3.1.24) catalysed reaction:

CoA + N-acylsphingosine ⇌ acyl-CoA + sphingosine.

Ceramide can alternatively act as a source of free fatty acid, via its degradation by N-acylsphingosine amidohydrolase (ASAH, EC 3.5.1.23). As we have shown, the central kPanK domain is similar to acyl-CoA synthetases (ACS), including fatty acyl-CoA synthetases (FACS), based on BLAST searches and predicted structure alignments. These enzymes typically activate fatty acids by forming a fatty acyl-AMP intermediate and then transferring the fatty acyl group to the pantetheine arm of coenzyme A, to be subsequently used in β-oxidation and other lipid pathways [295]. A tentative explanation then for the fusion of putative sphingomyelinase and adenylation domains is that it allows ceramide generated by the former to be used as a source of acyl groups (or carboxylate if ASAH activity is provided), for transfer to CoA.

Sphingolipids are abundant in trypanosomes, due to the extensive membrane re-modelling involved in cell division and differentiation, and, in the case of *T. brucei,* VSG turnover. SMase activity would therefore make the most of an abundant acyl source. The C-terminal PanK may come into the equation by placing CoA in proximity to the fused domains, since CoA is an allosteric regulator which binds to the ATP binding site of PanKs. In this work we demonstrated that CoA and acetyl-CoA inhibit *Tc*PanK activity, with acetyl-CoA being the more potent inhibitor. The central adenylation domain, were it to transfer an acyl-AMP intermediate to CoA within the PanK domain active site, would potentiate (or reduce) PanK inhibition, finely-tuning a negative feedback cycle (Fig. 6.2). Experimental validation for this hypothesis has been hampered by inability to obtain recombinant *Tc*PanK in its entirety for biochemical assessment, which would involve radiolabelling of sphingolipid substrates and demonstration of acyl group incorporation into acyl-CoA in purified kinetoplastid PanK preparations.

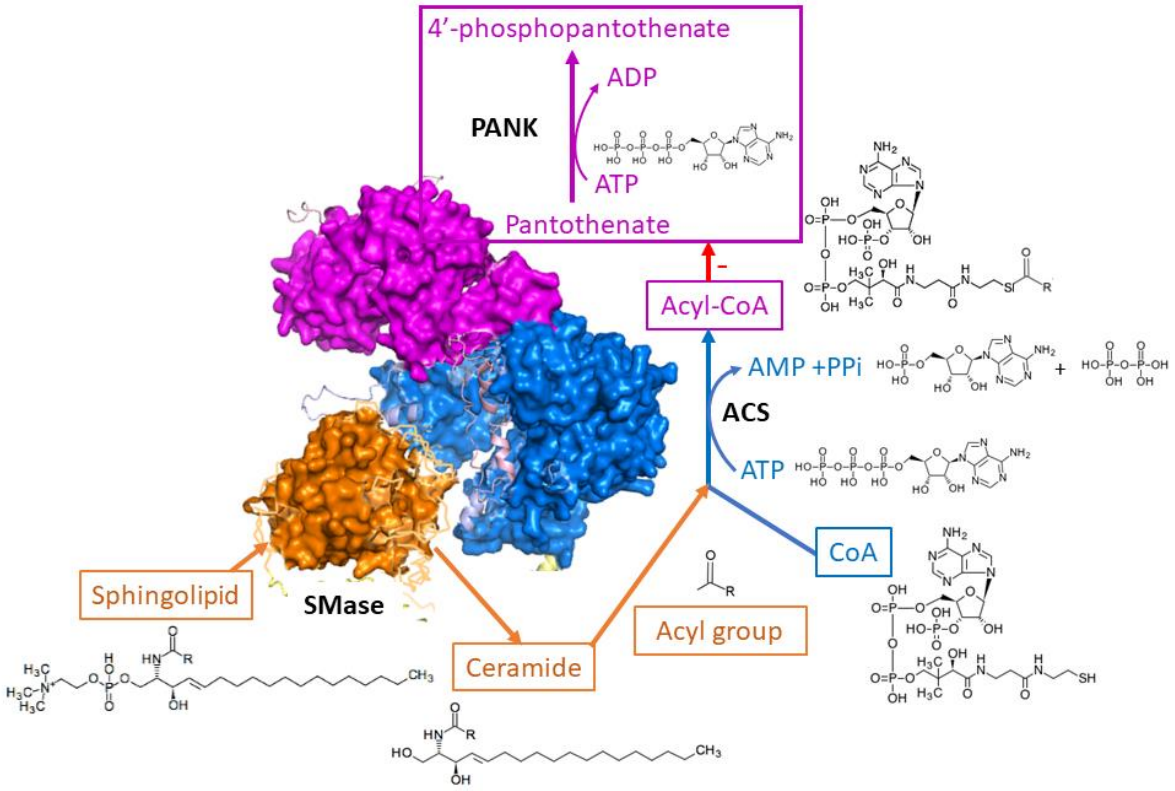


Fig. 6.2. Hypothetical kPanK regulation by fused domain activities: *The kPanK SMase activity (orange) generates ceramide from sphingolipid. The R group represents the alkyl portion of a fatty acid ($R=(CH_2)_n$-CH_3, where n=0-15). The ceramide acyl group (RCO) is transferred to CoA by acyl-CoA synthetase (ACS) activity of the central adenylation domain (blue), which activates the acyl group with AMP for transfer to CoA. Acylation of CoA at the PanK active site (magenta) potentiates competitive inhibition of ATP binding.*

As well as SMases, several DNA-binding enzymes share a similar structural fold to that predicted for the N-terminal kPanK domain. Since an extra-nuclear location of *T. cruzi* and *T. brucei* PanKs was observed in this study, consistent with a lack of identified nuclear-localisation signals, it seems improbable that kPanK interacts with DNA. The similarity between these enzymes may result from the fusion of an ancestral nuclease to the PanK locus. This could involve duplication of the nuclease followed by fusion of one copy to PanK and persistence of the second copy at the original locus, or fusion of a non-duplicated nuclease and loss of the original locus.

Perhaps a potential association between kPanKs and DNA should not be entirely disregarded. In contrast to our observations, nuclear PanKs have been described in some organisms: episomal PanK expression throughout the nucleus and cytosol of *P. falciparum* was reported [238] and the mammalian isoform PanK1α is a nuclear enzyme [306]. Furthermore, in bacteria, a link between DNA homeostasis and the CoA pathway has been described. YacG, a zinc-finger protein that regulates activity of DNA gyrase (a type IIA topoisomerase), is frequently encoded

downstream from dephospho-CoA kinase (DPCK/CoaE), which catalyses the penultimate step of CoA biosynthesis. In one bacterial species, *Actinobacillus succinogenes*, YacG is fused to CoaE as part of a single polypeptide(A6VLY6) [307]. Topoisomerases control chromosomal densities by altering the level of supercoiling, which influences DNA replication, expression and repair. DNA gyrase adds supercoils to a stretch of DNA, making it denser and easier to move around the chromosome [308]. This action is inhibited by YacG, which occludes the required DNA and ATP binding sites. It was suggested that the association between YacG and CoaE facilitates epigenetic regulation of cell growth in response to metabolic status [307].

During the cell cycle of yeast, there are distinct metabolic phases, whereby oxidative/ATP-generating and reductive/glycolytic metabolism occur separately and DNA replication, transcription and the process of cell division occur during the latter, which minimises the effects of reactive oxygen species (ROS)-mediated DNA-damage [309,310]. With this in mind, it was proposed that co-expression of CoaE and YacG during oxidative growth of bacteria might act to repress gyrase-facilitated DNA replication and transcription, to avoid ATP-associated DNA damage. It is interesting that TDP2, which also acts on topoisomerases, shares common motifs with the kPanK N-terminus. When topoisomerases re-arrange DNA strands, they form temporary covalent bonds with DNA and sometimes these DNA-topoisomerase complexes are not effectively cleaved, preventing release of the enzyme. TDP2 acts to release trapped topoisomerases from DNA, which maintains genomic integrity [265]. It is tempting to consider that a DNA-regulatory PanK may have existed, or may still exist in some kinetoplastids, although this seems not to be the case in *T. cruzi* epimastigotes and BSF *T. brucei*.

An intriguing possibility is that the predicted phosphodiesterase acts on CoA itself, which contains a type of diphosphate linkage separating the ADP and 4'phosphopantetheine moieties (Fig 6.3). CoA hydrolysis is known to occur in bacteria [311] and several eukaryotic species [312] and importantly, enzymes which catalyse this reaction are dependent on metal ion cofactors, in common with EEPs. In mammals, serum ectonucleotide pyrophosphatases [293] and intracellular Nudix-type (NudT) hydrolases catalyse hydrolysis of CoA and sometimes CoA esters and oxidized CoA [313]. These NudT hydrolases are localised to hotspots for CoA metabolism: NudT7 and NudT19 are found in peroxisomes and Nudt8 is mitochondrial [312]. It is proposed that they eliminate oxidized CoA from highly oxidizing environments, such as peroxisomes. CoA-metabolising NudT hydrolases, possess a particular motif 'LLTXR(SA)X_3RX_3GX_3FPGG' that is thought to determine substrate specificity for CoA and its derivatives. The Pde domain of kPanK shares 27% identity with NudT hydrolase PCD1 of *S. cerevisiae,* but only 4/12 conserved residues of the CoA-associated motif (Fig. 6.4).

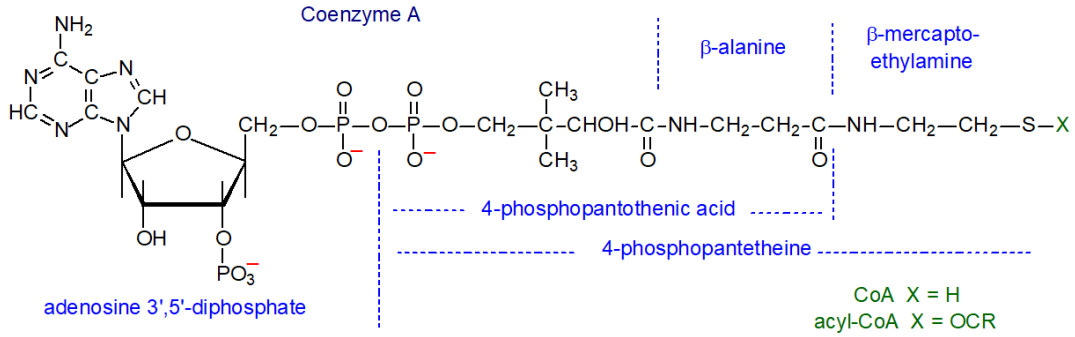


Fig. 6.3. Structure of CoA

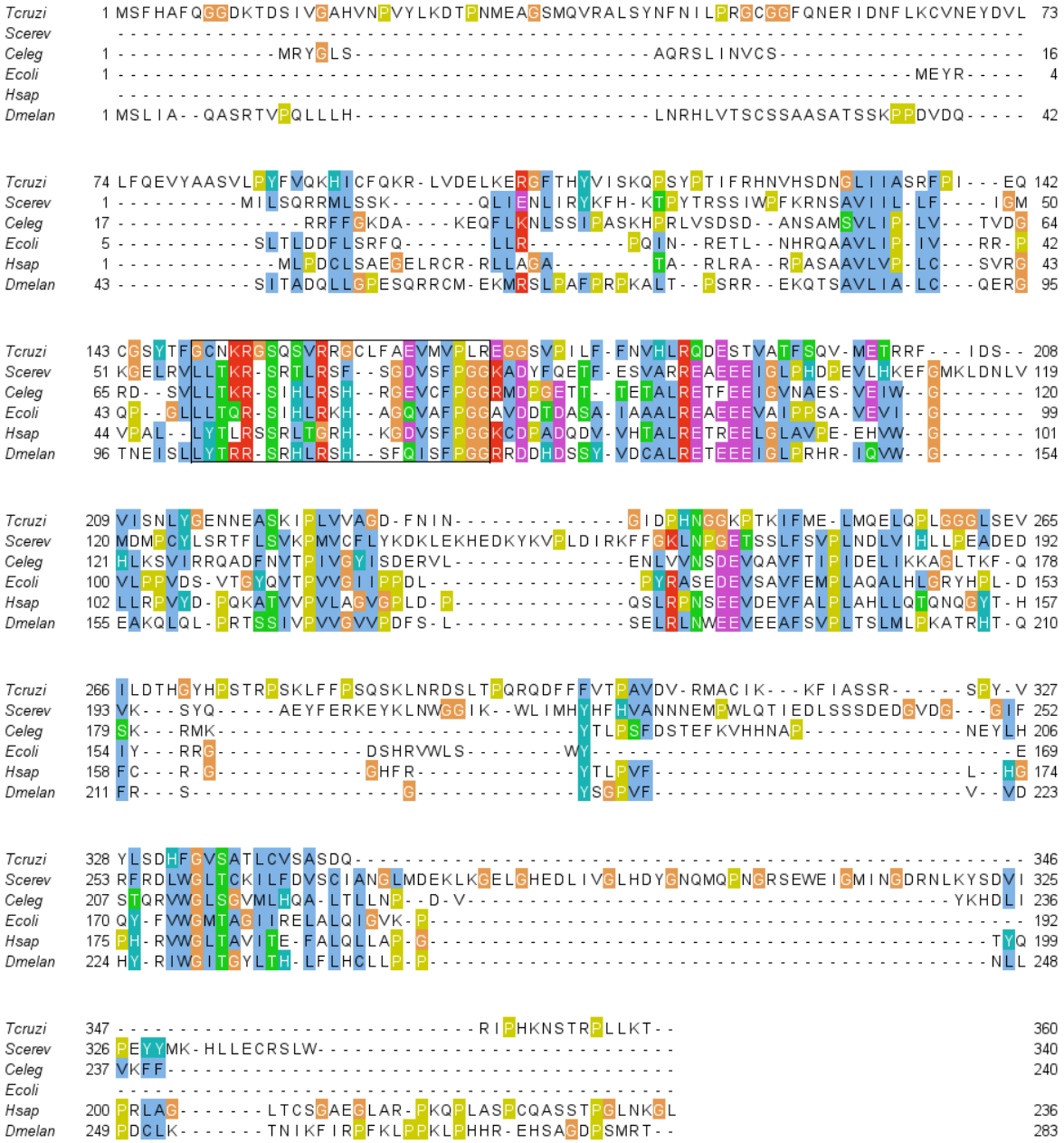


Fig. 6.4 TcPanK aligned to NudT hydrolases. *TcPanK was aligned (ClustalO) to CoA-specific Nudix-type hydrolases of S. cerevisiae (Q12524) C. elegans (Q23236), E. coli (P43337), H. sapiens (Q8WV74) and D. melanogaster (Q9VY79).The NudT motif associated with CoA substrate specificity is outlined in black. Colouring: yellow, proline; orange, glycine; red, positive charge; magenta, negative charge; blue, hydrophobic; green, polar; cyan, aromatic; white, unconserved.*

The CoA pathway metabolite 4 ′-phosphopantetheine (4’-PP) should also be considered a potential substrate of this domain. The precursor to CoA is formed from phosphopantetheine-cysteine in the third step of CoA synthesis. It can alternatively be produced from pantetheine, which is catalysed by some PanKs, and is a breakdown product of CoA and ACPs. Studies have shown that under certain metabolic conditions, 4’-PP and its oxidised or sulfonated forms can accumulate with potentially toxic outcomes. For instance, oxidised 4 ′-PP can be converted to inactive forms of CoA and acyl carrier protein, thereby competitively inhibiting CoA synthesis. It appears that some eukaryotic PanKs, named PanK4 in mammals and PanK2 in plants, possesses a phosphatase domain that can hydrolyse 4’-P and its sulfonated/oxidised forms[233]. PanK4 isoforms consist of an amino-terminal PanK and a carboxy-terminal Domain of Unknown Function 89 (DUF89) which is a ubiquitous, metal-dependent phosphatase. DUF89 has been broadly implicated in the prevention of DNA damage. For example, the stand-alone DUF89 protein of *Arabidopsis* (At2g17340) acts on sugar phosphates, which are extremely potent glycating agents. Similarly, a yeast DUF89 (YMR027W) limits damage by the DNA toxin methyl methane sulfonate. To determine the likelihood of kPanK having such a regulatory function, DUF89 domains of PanKs from *H. sapiens* (PanK4), *Danio rerio* (PanK4) *and A. thaliana* (PanK2) were aligned with *Tc*PanK residues 1-1020 (Fig. 6.5). Domains 1 and 2 of *Tc*PanK, which share 36% identity with *A. thaliana* PanK2, possess limited regions of similarity with DUF89 proteins and therefore do not clearly resemble phosphopantetheineases based on sequence homology.

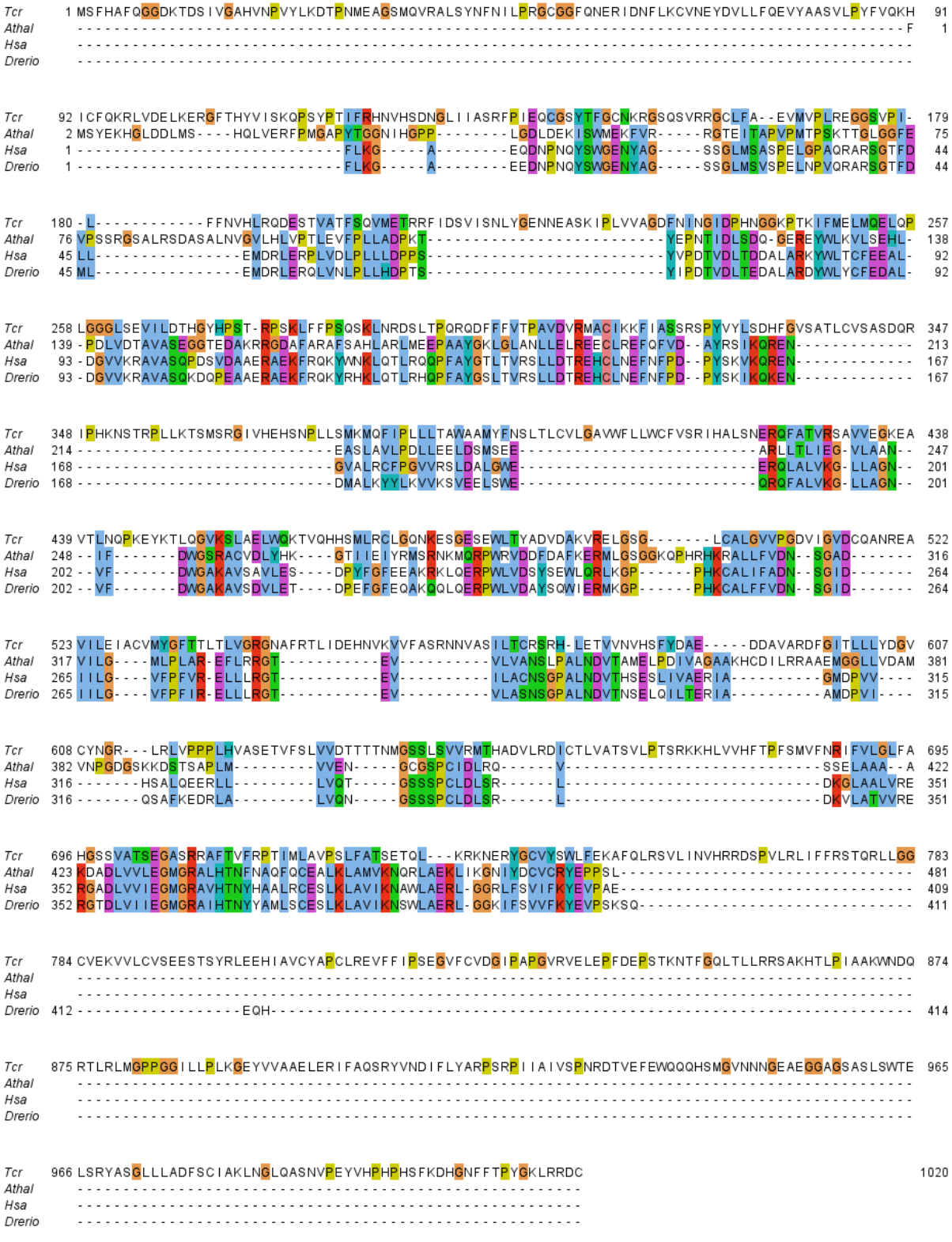


Fig. 6.5. DUF89 proteins aligned with TcPanKs: *Phosphopantetheineases of H. sapiens (Hsa) and Danio rerio (Drerio) PanK4 and A. thaliana (Athal) PanK2 were aligned with TcPanK residues 1-1020 (inclusive of domains 1 and 2). Colouring as in Fig. 6.3.*

Future work

To obtain conclusive insights into the essentiality of PanK in *T. cruzi,* conditional deletion would be an informative addition to these analyses. A system that is currently available is pTcINDEX, which can be used to deliver an ectopically expressed gene under the control of tetracycline-regulatable promoter to the transcriptionally silent ribosomal non-transcribed spacer [314]. By providing an additional copy of *Tc*PanK and inducing its expression with tetracycline, then both genomic loci can be targeted using CRISPR. Removal of tetracycline from the conditional null mutants should arrest expression of the introduced copy causing a gradual depletion of the remaining enzyme pool, allowing the essential nature of the gene to be confirmed, or otherwise. The main limitation of this system is that expression of the construct may still occur at a low level in the absence of tetracycline. Furthermore, due to the requirement for cloning and multiple transfections on top of careful molecular validations of generated cell lines, this is a time-consuming method.

An alternative approach that is gaining interest as a gene characterisation tool, is the use of stabilisation/destabilisation proteins such as small molecule–assisted shutoff (SMASh). The SMASh technique involves fusion of the target protein to a degron tag that self-cleaves due to the presence of a viral protease and intergenic cleavage site [315]. Asunepravir and other inhibitors of the viral protease occlude the intergenic cleavage site leading to persistence of the tag and targeting of the fusion protein to the cytoplasmic proteasome. So far this method has been applied successfully in mammalian cells and yeast and is in development for use with *T. cruzi*. Delivery of a SMASh tag-PanK fusion construct with concomitant disruption of endogenous *PanK* loci, followed by treatment of transgenic lines with asunepravir could be an effective and specific method of inducible PanK depletion.

Inducible CRISPR depletion methods are another option for highly specific and efficient targeting of essential enzymes. Targeted mRNA degradation can be achieved using RNase activity of CRISPR Cas13. Successful knockdown of specific mRNAs has been achieved using CRISPR Cas13 in mammalian and plant cells, but has yet to be applied to trypanosomes. The feasibility of stably expressing Cas13 in *T. cruzi* is being evaluated and pending success, PanK mRNA will be targeted in cell lines expressing tagged protein (fluorescent or PTP-tagged) so that epitope signals can be monitored to validate knockdown. Alternatively, the combined use of Cas9 and Di-CRE recombinase systems to efficiently induce knockdown of essential genes has been successfully adopted for *Leishmania* strains and therefore might be adaptable to *T. cruzi* [316]. This approach requires integration of LoxP sequences by CRISPR-Cas9 followed by Cre-recombinase catalysis of LoxP site recombination to excise the target gene. The LoxP sequences should be inserted in minimally transcribed regions upstream or

downstream of the target gene to avoid disrupting its function and localisation. Insertion of these sites can be achieved by PCR generation of a donor cassette containing the target site flanked by LoxP sequences, which is delivered to the cell line with sgRNAs for Cas9 insertion. As well as requiring T7 polymerase and Cas9 expression, compatibility with this method requires stable expression of Di-CRE- a CRE recombinase split into two inactive fragments, one fused to a FKBP12 domain and the other to a FRB domain. The fragments dimerise and hence become active upon exposure to rapamycin, resulting in excision of the LoxP-flanked region.

To predict the functions of the fused domains, immunoprecipitation was attempted using PTP-tagged *Tc*PanK, based on a method by Günzl et al, in the hope that co-purified protein interactors could be identified by mass spectrometry [273]. The PTP tag consists of an inner Protein C epitope (ProtC) and an outer Protein A domain (ProtA), separated by a TEV cleavage site. The protein is initially immobilised on an IgG-coated matrix and then released by TEV cleavage, followed by a second immobilisation on anti-ProtC beads and release using EGTA. The construct that we targeted to the C-terminus of *Tc*PanK contained 3 Ty repeats internal to the PTP tag. Despite western blots indicating successful tagging (using anti-Ty primary antibody), attempted immobilisation of tagged *Tc*PanK on IgG resin was unsuccessful, potentially due to inaccessibility of the PTP tag. Another limiting factor was the seemingly low expression level of *Tc*PanK, which indicates that immunoprecipitation of overexpressed protein may be required to obtain a sufficient quantity of protein for analysis.

An alternative approach to immunoprecipitation which can identify physiologically relevant protein interactions is proximity labelling, such as use of the BioID method [317]. This would require engineering a *T. cruzi* cell line that expresses a PANK fused to a biotin ligase such as BirA or APEX2 [318,319], which should promiscuously biotinylate nearby proteins, including those which interact weakly or transiently. This captures interactions over time as proteins dissociate and reassociate with the enzyme and potentially accumulate biotin. The biotinylated interactors can then be purified from lysed cells for subsequent identification using well-established biotin-affinity capture methods, such as streptavidin-coated beads or columns. A number of studies have successfully used BioID to study protein complexes in kinetoplastids, particularly in procyclic forms of *T. brucei* but also in *L. mexicana* promastigotes (reviewed in Geoghegan et al., 2022) [320]. In several of these studies the endogenous enzyme was tagged with a biotin ligase [319,321–323]. However, a majority of studies overexpressed the fusion enzyme under a tetracycline-regulatable promoter [324–330]. Based on reported *Tc*PanK expression levels [296] and failure to detect fluorescence in several *Tc*PanK tagged cell lines (Chapter 4.2), overexpression of the fusion protein may be required for epimastigote studies investigating

PanK interactions, but endogenous tagging in amastigotes would be preferable to avoid mis-localisation and any resulting artefactual interactions.

Concluding remarks

In summary, we provide the first description of an active kinetoplastid pantothenate kinase and evidence for essentiality of PanK to proliferation of BSF *T. brucei*. We predict that the enzyme is also necessary for viability of *T. cruzi* epimastigotes. PanKs are already being studied in *P. falciparum, M. tuberculosis* and other pathogens due to their demonstrated importance to growth and infection and therefore their potential to be effective drug targets. The results of this work suggest that kinetoplastid PanKs may be even more desirable as drug targets, due to their possession of two N-terminally fused domains, predicted to be phosphodiesterase and adenylation activities, which are highly conserved amongst kinetoplastids, including the clinically important pathogens *T. cruzi, T. brucei* and *Leishmania,* while being absent from the PanKs of other organisms, including humans. Their presence raises the possibility of highly selective targeting of this important enzyme.

While it is clear from our *T. brucei* experiments that the PanK domain is essential for proliferation, including the active site residue, arginine 1270, it is not in itself sufficient for *in vitro* growth and the fused domains therefore contribute to the essential role of the multifunctional enzyme. We offer a tentative hypothesis that the fused activities have regulatory roles, either modulating PanK by negative feedback, or regulating levels of CoA metabolites to prevent toxicity associated with their accumulation. Selective chemotherapies are urgently required to combat Chagas disease, Human African Trypanosomiasis and Leishmaniases as current drugs pose the risk of toxic side-effects. Further characterisation of the fused activities will determine whether selective inhibition may provide a means to potentiate the effect of PanK inhibitors whilst minimising host toxicity.

References

1. Murray, H. W. The kinetoplast in the diagnosis of visceral leishmaniasis. *IDCases* **29**, e01565 (2022).
2. Aphasizhev, R. & Aphasizheva, I. Mitochondrial RNA editing in trypanosomes: small RNAs in control. *Biochimie* **100**, 125–131 (2014).
3. Alvar, J. *et al.* Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE* vol. 7 e35671 Preprint at https://doi.org/10.1371/journal.pone.0035671 (2012).
4. Stanaway, J. D. & Roth, G. The burden of Chagas disease: estimates and challenges. *Glob. Heart* **10**, 139–144 (2015).
5. Chagas disease. *Drugs for Neglected Diseases initiative (DNDi)* https://dndi.org/diseases/chagas/facts/ (2020).
6. El-Sayed, N. M. Comparative Genomics of Trypanosomatid Parasitic Protozoa. *Science* **309**, 404–409 (2005).
7. De Pablos, L. M. & Osuna, A. Multigene families in Trypanosoma cruzi and their role in infectivity. *Infect. Immun.* **80**, 2258–2264 (2012).
8. Melville, S. E., Leech, V., Navarro, M. & Cross, G. A. The molecular karyotype of the megabase chromosomes of Trypanosoma brucei stock 427. *Mol. Biochem. Parasitol.* **111**, 261–273 (2000).
9. Weatherly, D. B., Boehlke, C. & Tarleton, R. L. Chromosome level assembly of the hybrid Trypanosoma cruzi genome. *BMC Genomics* **10**, 255 (2009).
10. Ivens, A. C. *et al.* The genome of the kinetoplastid parasite, Leishmania major. *Science* **309**, 436–442 (2005).
11. Clayton, C. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol.* **9**, 190072 (2019).
12. Hamilton, P. B., Teixeira, M. M. G. & Stevens, J. R. The evolution of Trypanosoma cruzi: The “bat seeding” hypothesis. *Trends in Parasitology* vol. 28 136–141 Preprint at https://doi.org/10.1016/j.pt.2012.01.006 (2012).

13. Pérez-Molina, J. A. & Molina, I. Chagas disease. *Lancet* **391**, 82–94 (2018).
14. Chagas disease (also known as American trypanosomiasis). https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis).
15. Xavier, S. C. das C. *et al.* Lower richness of small wild mammal species and chagas disease risk. *PLoS Negl. Trop. Dis.* **6**, e1647 (2012).
16. Bender, A., Python, A., Lindsay, S. W., Golding, N. & Moyes, C. L. Modelling geospatial distributions of the triatomine vectors of Trypanosoma cruzi in Latin America. *bioRxiv* 738310 (2019) doi:10.1101/738310.
17. Gaspe, M. S. *et al.* Urbanisation, risk stratification and house infestation with a major vector of Chagas disease in an endemic municipality of the Argentine Chaco. *Parasit. Vectors* **13**, 316 (2020).
18. Lima, L. M. de *et al.* Prevalence of Chagas disease in blood donors at the Uberaba Regional Blood Center, Brazil, from 1995 to 2009. *Rev. Soc. Bras. Med. Trop.* **45**, 723–726 (2012).
19. Cevallos, A. M. & Hernández, R. Chagas' disease: pregnancy and congenital transmission. *Biomed Res. Int.* **2014**, 401864 (2014).
20. Apt, W. *et al.* Congenital infection by Trypanosoma cruzi in an endemic area of Chile: a multidisciplinary study. *Trans. R. Soc. Trop. Med. Hyg.* **107**, 98–104 (2013).
21. Pinto, A. Y. das N., Valente, S. A., Valente, V. da C., Ferreira Junior, A. G. & Coura, J. R. [Acute phase of Chagas disease in the Brazilian Amazon region: study of 233 cases from Pará, Amapá and Maranhão observed between 1988 and 2005]. *Rev. Soc. Bras. Med. Trop.* **41**, 602–614 (2008).
22. Souza-Lima, R. de C. de *et al.* Outbreak of acute Chagas disease associated with oral transmission in the Rio Negro region, Brazilian Amazon. *Rev. Soc. Bras. Med. Trop.* **46**, 510–514 (2013).
23. Antinori, S. *et al.* Chagas disease in Europe: A review for the internist in the globalized world. *Eur. J. Intern. Med.* **43**, 6–15 (2017).

24. Imai, K. *et al.* Chronic Chagas disease with advanced cardiac complications in Japan: Case report and literature review. *Parasitol. Int.* **64**, 240–242 (2015).
25. Bern, C., Kjos, S., Yabsley, M. J. & Montgomery, S. P. Trypanosoma cruzi and chagas' disease in the united states. *Clinical Microbiology Reviews* Preprint at https://doi.org/10.1128/CMR.00005-11 (2011).
26. Jackson, Y., Pinto, A. & Pett, S. Chagas disease in Australia and New Zealand: risks and needs for public health interventions. *Trop. Med. Int. Health* **19**, 212–218 (2014).
27. Sangenito, L. S., Branquinha, M. H. & Santos, A. L. S. Funding for Chagas Disease: A 10-Year (2009-2018) Survey. *Trop Med Infect Dis* **5**, (2020).
28. WHO. *Control of Chagas Disease: Second Report of the WHO Expert Committee. WHO Technical Report Series*. https://apps.who.int/iris/bitstream/handle/10665/42443/WHO_TRS_905.pdf?sequence=1&isAllowed=y (2002).
29. Bruneto, E. G. *et al.* Case-fatality From Orally-transmitted Acute Chagas Disease: A Systematic Review and Meta-analysis. *Clin. Infect. Dis.* **72**, 1084–1092 (2021).
30. Coura, J. R. The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions--a comprehensive review. *Mem. Inst. Oswaldo Cruz* **110**, 277–282 (2015).
31. Sabino, E. C. *et al.* Ten-year incidence of Chagas cardiomyopathy among asymptomatic Trypanosoma cruzi-seropositive former blood donors. *Circulation* **127**, 1105–1115 (2013).
32. Nunes, M. C. P. *et al.* Chagas Cardiomyopathy: An Update of Current Clinical Knowledge and Management: A Scientific Statement From the American Heart Association. *Circulation* **138**, e169–e209 (2018).
33. Pech-Aguilar, A. G., Haro-Álvarez, A. P. & Rosado-Vallado, M. E. Updated review on the pathophysiology of Chagas cardiomyopathy. *Rev. Med. Inst. Mex. Seguro Soc.* **58**, 328–334 (2020).

34. Matsuda, N. M., Miller, S. M. & Evora, P. R. B. The chronic gastrointestinal manifestations of Chagas disease. *Clinics* **64**, 1219–1224 (2009).
35. Jabari, S., de Oliveira, E. C., Brehmer, A. & da Silveira, A. B. M. Chagasic megacolon: enteric neurons and related structures. *Histochem. Cell Biol.* **142**, 235–244 (2014).
36. Martins, M. C. L. *et al.* Chagasic Megaesophagus-Associated Carcinoma: Clinical Pattern and Outcomes. *J Glob Oncol* **5**, 1–5 (2019).
37. Morillo, C. A. *et al.* Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy. *N. Engl. J. Med.* **373**, 1295–1306 (2015).
38. Bocchi, E. A. Heart transplants for patients with Chagas' heart disease. *Sao Paulo Med. J.* **113**, 873–879 (1995).
39. Alonso-Padilla, J. *et al.* Strategies to enhance access to diagnosis and treatment for Chagas disease patients in Latin America. *Expert Rev. Anti. Infect. Ther.* **17**, 145–157 (2019).
40. Lee, B. Y., Bacon, K. M., Bottazzi, M. E. & Hotez, P. J. Global economic burden of Chagas disease: a computational simulation model. *Lancet Infect. Dis.* **13**, 342–348 (2013).
41. Steverding, D. The history of Chagas disease. *Parasites and Vectors* vol. 7 317 Preprint at https://doi.org/10.1186/1756-3305-7-317 (2014).
42. Zingales, B. *et al.* The revised Trypanosoma cruzi subspecific nomenclature: Rationale, epidemiological relevance and research applications. *Infect. Genet. Evol.* **12**, 240–253 (2012).
43. Zingales, B. *et al.* A new consensus for Trypanosoma cruzi intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem. Inst. Oswaldo Cruz* **104**, 1051–1054 (2009).
44. Lima, L. *et al.* Genetic diversity of Trypanosoma cruzi in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). *Acta Tropica* **151**, 166–177 (2015).
45. El-Sayed, N. M. The Genome Sequence of Trypanosoma cruzi, Etiologic Agent of Chagas Disease. *Science* **309**, 409–415 (2005).

46. Sales Junior, P. A. *et al.* Experimental and Clinical Treatment of Chagas Disease: A Review. *Am. J. Trop. Med. Hyg.* **97**, 1289–1303 (2017).
47. Fernandes, M. C. & Andrews, N. W. Host cell invasion by Trypanosoma cruzi: a unique strategy that promotes persistence. *FEMS Microbiol. Rev.* **36**, 734–747 (2012).
48. Maeda, F. Y., Cortez, C. & Yoshida, N. Cell signaling during Trypanosoma cruzi invasion. *Front. Immunol.* **3**, 361 (2012).
49. Epting, C. L., Coates, B. M. & Engman, D. M. Molecular mechanisms of host cell invasion by Trypanosoma cruzi. *Exp. Parasitol.* **126**, 283–291 (2010).
50. Ward, A. I., Lewis, M. D. & Khan, A. In Vivo Analysis of Trypanosoma cruzi Persistence Foci at Single-Cell Resolution. *mBio* **11**, (2020).
51. Trypanosomiasis, human African (sleeping sickness). https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness).
52. Schultzberg, M., Ambatsis, M., Samuelsson, E. B., Kristensson, K. & van Meirvenne, N. Spread of Trypanosoma brucei to the nervous system: early attack on circumventricular organs and sensory ganglia. *J. Neurosci. Res.* **21**, 56–61 (1988).
53. Kristensson, K., Masocha, W. & Bentivoglio, M. Mechanisms of CNS invasion and damage by parasites. *Handb. Clin. Neurol.* **114**, 11–22 (2013).
54. Laperchia, C. *et al.* Trypanosoma brucei Invasion and T-Cell Infiltration of the Brain Parenchyma in Experimental Sleeping Sickness: Timing and Correlation with Functional Changes. *PLoS Negl. Trop. Dis.* **10**, e0005242 (2016).
55. Kennedy, P. G. The pathogenesis and modulation of the post-treatment reactive encephalopathy in a mouse model of Human African Trypanosomiasis. *J. Neuroimmunol.* **100**, 36–41 (1999).
56. Rodgers, J. Trypanosomiasis and the brain. *Parasitology* **137**, 1995–2006 (2010).
57. Kennedy, P. G. E. Human African trypanosomiasis-neurological aspects. *J. Neurol.* **253**, 411–416 (2006).
58. Mugnier, M. R., Stebbins, C. E. & Papavasiliou, F. N. Masters of Disguise: Antigenic Variation and the VSG Coat in Trypanosoma brucei. *PLoS Pathog.* **12**, e1005784 (2016).

59. Horn, D. Antigenic variation in African trypanosomes. *Mol. Biochem. Parasitol.* **195**, 123–129 (2014).
60. Leishmaniasis. https://www.who.int/news-room/fact-sheets/detail/leishmaniasis.
61. Sasidharan, S. & Saudagar, P. Leishmaniasis: where are we and where are we heading? *Parasitol. Res.* **120**, 1541–1554 (2021).
62. Akhoundi, M. *et al.* A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies. *PLoS Neglected Tropical Diseases* vol. 10 e0004349 Preprint at https://doi.org/10.1371/journal.pntd.0004349 (2016).
63. Alvar, J., Cañavate, C., Molina, R., Moreno, J. & Nieto, J. Canine leishmaniasis. *Advances in Parasitology* vol. 57 1–88 Preprint at https://doi.org/10.1016/S0065-308X(04)57001-X (2004).
64. Maroli, M., Feliciangeli, M. D., Bichaud, L., Charrel, R. N. & Gradoni, L. Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Med. Vet. Entomol.* **27**, 123–147 (2013).
65. Gutiérrez-Kobeh, L. *et al.* Inhibition of dendritic cell apoptosis by Leishmania mexicana amastigotes. *Parasitol. Res.* **112**, 1755–1762 (2013).
66. Wenzel, U. A. *et al.* Leishmania major parasite stage-dependent host cell invasion and immune evasion. *FASEB J.* **26**, 29–39 (2012).
67. McConville, M. J. & Naderer, T. Metabolic pathways required for the intracellular survival of Leishmania. *Annu. Rev. Microbiol.* **65**, 543–561 (2011).
68. Bhardwaj, S., Srivastava, N., Sudan, R. & Saha, B. Leishmania interferes with host cell signaling to devise a survival strategy. *J. Biomed. Biotechnol.* **2010**, 109189 (2010).
69. Rodrigues, A. *et al.* Leishmania infantum antigens modulate memory cell subsets of liver resident T lymphocyte. *Immunobiology* **222**, 409–422 (2017).
70. De Oliveira, C. I. & Brodskyn, C. I. The immunobiology of Leishmania braziliensis infection. *Frontiers in Immunology* vol. 3 Preprint at https://doi.org/10.3389/fimmu.2012.00145 (2012).

71. Bennis, I., De Brouwere, V., Belrhiti, Z., Sahibi, H. & Boelaert, M. Psychosocial burden of localised cutaneous Leishmaniasis: a scoping review. *BMC Public Health* **18**, 358 (2018).
72. David, C. V. & Craft, N. Cutaneous and mucocutaneous leishmaniasis. *Dermatol. Ther.* **22**, 491–502 (2009).
73. WHO. *Control of the Leishmaniasis: Report of a Meeting of the World Health Organization Expert Committee on the Control of Leishmaniases, WHO Technical Report Series*. https://apps.who.int/iris/bitstream/handle/10665/44412/WHO_TRS_949_eng.pdf?sequence=1&isAllowed=y (WHO; Geneva: 2010).
74. Ronet, C., Beverley, S. M. & Fasel, N. Muco-cutaneous leishmaniasis in the New World: the ultimate subversion. *Virulence* **2**, 547–552 (2011).
75. Herwaldt, B. L. Leishmaniasis. *Lancet* **354**, 1191–1199 (1999).
76. Blum, J., Nkunku, S. & Burri, C. Clinical description of encephalopathic syndromes and risk factors for their occurrence and outcome during melarsoprol treatment of human African trypanosomiasis. *Trop. Med. Int. Health* **6**, 390–400 (2001).
77. Fairlamb, A. H. & Horn, D. Melarsoprol Resistance in African Trypanosomiasis. *Trends Parasitol.* **34**, 481–492 (2018).
78. Phillips, M. A. & Wang, C. C. A Trypanosoma brucei mutant resistant to alpha-difluoromethylornithine. *Mol. Biochem. Parasitol.* **22**, 9–17 (1987).
79. Vincent, I. M. *et al.* A molecular mechanism for eflornithine resistance in African trypanosomes. *PLoS Pathog.* **6**, e1001204 (2010).
80. Deeks, E. D. Fexinidazole: First Global Approval. *Drugs* **79**, 215–220 (2019).
81. Betu Kumeso, V. K. *et al.* Efficacy and safety of acoziborole in patients with human African trypanosomiasis caused by Trypanosoma brucei gambiense: a multicentre, open-label, single-arm, phase 2/3 trial. *Lancet Infect. Dis.* (2022) doi:10.1016/S1473-3099(22)00660-0.
82. Berman, J. D., Waddell, D. & Hanson, B. D. Biochemical mechanisms of the antileishmanial activity of sodium stibogluconate. *Antimicrob. Agents Chemother.* **27**, 916–920 (1985).

83. Mosimann, V., Neumayr, A., Paris, D. H. & Blum, J. Liposomal amphotericin B treatment of Old World cutaneous and mucosal leishmaniasis: A literature review. *Acta Trop.* **182**, 246–250 (2018).
84. Kumari, S. *et al.* Amphotericin B: A drug of choice for Visceral Leishmaniasis. *Acta Trop.* **235**, 106661 (2022).
85. Piccica, M., Lagi, F., Bartoloni, A. & Zammarchi, L. Efficacy and safety of pentamidine isethionate for tegumentary and visceral human leishmaniasis: a systematic review. *J. Travel Med.* **28**, (2021).
86. Maia, A. C. R. G. *et al.* Leishmania infantum amastigote nucleoside triphosphate diphosphohydrolase 1 (NTPDase 1): Its inhibition as a new insight into mode of action of pentamidine. *Exp. Parasitol.* **200**, 1–6 (2019).
87. Kaur, G. & Rajput, B. Comparative analysis of the omics technologies used to study antimonial, amphotericin B, and pentamidine resistance in leishmania. *J. Parasitol. Res.* **2014**, 726328 (2014).
88. Sundar, S. *et al.* Short-course of oral miltefosine for treatment of visceral leishmaniasis. *Clin. Infect. Dis.* **31**, 1110–1113 (2000).
89. Sindermann, H. & Engel, J. Development of miltefosine as an oral treatment for leishmaniasis. *Trans. R. Soc. Trop. Med. Hyg.* **100 Suppl 1**, S17-20 (2006).
90. Javier Pérez-Victoria, F., Gamarro, F., Ouellette, M. & Castanys, S. Functional Cloning of the Miltefosine Transporter: A NOVEL P-TYPE PHOSPHOLIPID TRANSLOCASE FROM LEISHMANIA INVOLVED IN DRUG RESISTANCE *. *J. Biol. Chem.* **278**, 49965–49971 (2003).
91. Sundar, S., Jha, T. K., Thakur, C. P., Sinha, P. K. & Bhattacharya, S. K. Injectable paromomycin for Visceral leishmaniasis in India. *N. Engl. J. Med.* **356**, 2571–2581 (2007).
92. Maarouf, M., Adeline, M. T., Solignac, M., Vautrin, D. & Robert-Gero, M. Development and characterization of paromomycin-resistant Leishmania donovani promastigotes. *Parasite* **5**, 167–173 (1998).

93. Hall, B. S., Bot, C. & Wilkinson, S. R. Nifurtimox activation by trypanosomal type I nitroreductases generates cytotoxic nitrile metabolites. *J. Biol. Chem.* **286**, 13088–13095 (2011).
94. Docampo, R. Sensitivity of parasites to free radical damage by antiparasitic drugs. *Chem. Biol. Interact.* **73**, 1–27 (1990).
95. Wilkinson, S. R., Taylor, M. C., Horn, D., Kelly, J. M. & Cheeseman, I. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5022–5027 (2008).
96. Mejia, A. M. *et al.* Benznidazole-resistance in Trypanosoma cruzi is a readily acquired trait that can arise independently in a single population. *J. Infect. Dis.* **206**, 220–228 (2012).
97. Mejía-Jaramillo, A. M., Fernández, G. J., Palacio, L. & Triana-Chávez, O. Gene expression study using real-time PCR identifies an NTR gene as a major marker of resistance to benzonidazole in Trypanosoma cruzi. *Parasit. Vectors* **4**, 169 (2011).
98. Baker, N., Alsford, S. & Horn, D. Genome-wide RNAi screens in African trypanosomes identify the nifurtimox activator NTR and the eflornithine transporter AAT6. *Mol. Biochem. Parasitol.* **176**, 55–57 (2011).
99. de Castro, S. L. The challenge of Chagas' disease chemotherapy: an update of drugs assayed against Trypanosoma cruzi. *Acta Trop.* **53**, 83–98 (1993).
100. Andrade, S. G., Magalhães, J. B. & Pontes, A. L. Evaluation of chemotherapy with benznidazole and nifurtimox in mice infected with Trypanosoma cruzi strains of different types. *Bull. World Health Organ.* **63**, 721–726 (1985).
101. Machado-de-Assis, G. F. *et al.* A serological, parasitological and clinical evaluation of untreated Chagas disease patients and those treated with benznidazole before and thirteen years after intervention. *Mem. Inst. Oswaldo Cruz* **108**, 873–880 (2013).
102. Stass, H., Ince, I., Grossmann, U., Weimann, B. & Willmann, S. Nifurtimox for Treatment of Chagas Disease in Pediatric Patients: the Challenges of Applying Pharmacokinetic-Pharmacodynamic Principles to Dose Finding. *AAPS J.* **24**, 92 (2022).

103. Forsyth, C. J. *et al.* Safety Profile of Nifurtimox for Treatment of Chagas Disease in the United States. *Clin. Infect. Dis.* **63**, 1056–1062 (2016).
104. Molina, I. *et al.* Toxic Profile of Benznidazole in Patients with Chronic Chagas Disease: Risk Factors and Comparison of the Product from Two Different Manufacturers. *Antimicrob. Agents Chemother.* **59**, 6125–6131 (2015).
105. Perez-Molina, J. A. *et al.* Use of benznidazole to treat chronic Chagas' disease: a systematic review with a meta-analysis. *J. Antimicrob. Chemother.* **64**, 1139–1147 (2009).
106. Francisco, A. F. *et al.* Limited Ability of Posaconazole To Cure both Acute and Chronic Trypanosoma cruzi Infections Revealed by Highly Sensitive In Vivo Imaging. *Antimicrob. Agents Chemother.* **59**, 4653–4661 (2015).
107. Morillo, C. A. *et al.* Benznidazole and Posaconazole in Eliminating Parasites in Asymptomatic T. Cruzi Carriers: The STOP-CHAGAS Trial. *J. Am. Coll. Cardiol.* **69**, 939–947 (2017).
108. Bahia, M. T. *et al.* Antitrypanosomal activity of fexinidazole metabolites, potential new drug candidates for Chagas disease. *Antimicrob. Agents Chemother.* **58**, 4362–4370 (2014).
109. Bahia, M. T. *et al.* Fexinidazole: a potential new drug candidate for Chagas disease. *PLoS Negl. Trop. Dis.* **6**, e1870 (2012).
110. Francisco, A. F. *et al.* Nitroheterocyclic drugs cure experimental Trypanosoma cruzi infections more effectively in the chronic stage than in the acute stage. *Sci. Rep.* **6**, 35351 (2016).
111. Padilla, A. M. *et al.* Discovery of an orally active benzoxaborole prodrug effective in the treatment of Chagas disease in non-human primates. *Nat Microbiol* **7**, 1536–1546 (2022).
112. Lee, C.-Y. G. *Co-purification of coenzyme-dependent enzymes by affinity chromatography**. vol. 57 27–40 (1983).

113. Leonardi, R., Zhang, Y. M., Rock, C. O. & Jackowski, S. Coenzyme A: Back in action. *Progress in Lipid Research* Preprint at https://doi.org/10.1016/j.plipres.2005.04.001 (2005).

114. Shi, L. & Tu, B. P. Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. *Curr. Opin. Cell Biol.* **33**, 125–131 (2015).

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

116. Gilbert, R. J., Klein, R. A. & Miller, P. G. The role of threonine in the metabolism of acetyl coenzyme a by Trypanosoma brucei brucei. *Comp. Biochem. Physiol. B* **74**, 277–281 (1983).

117. Ginger, M. L., Chance, M. L. & Goad, L. J. Elucidation of carbon sources used for the biosynthesis of fatty acids and sterols in the trypanosomatid Leishmania mexicana. *Biochem. J* **342 (Pt 2)**, 397–405 (1999).

118. Trindade, S. *et al.* Trypanosoma brucei Parasites Occupy and Functionally Adapt to the Adipose Tissue in Mice. *Cell Host Microbe* **19**, 837–848 (2016).

119. Book: Biochemistry Free & Easy (Ahern and Rajagopal). *Biology LibreTexts* https://bio.libretexts.org/Bookshelves/Biochemistry/Book%3A_Biochemistry_Free_and_Easy_(Ahern_and_Rajagopal) (2014).

120. Michels, P. A. M., Bringaud, F., Herman, M. & Hannaert, V. Metabolic functions of glycosomes in trypanosomatids. *Biochim. Biophys. Acta* **1763**, 1463–1477 (2006).

121. Buscaglia, C. A. *et al.* A putative pyruvate dehydrogenase α subunit gene fromTrypanosoma cruzi. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1309**, 53–57 (1996).

122. Roberts, C. W. *et al.* Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. *Mol. Biochem. Parasitol.* **126**, 129–142 (2003).

123. Herman, J. D., Gallalee, J. V. & Best, J. M. Sodium stibogluconate (pentostam) inhibition of glucose catabolism via the glycolytic pathway, and fatty acid β-oxidation in leishmania mexicana amastigotes. *Biochem. Pharmacol.* **36**, 197–201 (1987).

124. Patino, L. H., Muskus, C. & Ramírez, J. D. Transcriptional responses of Leishmania (Leishmania) amazonensis in the presence of trivalent sodium stibogluconate. *Parasit. Vectors* **12**, 348 (2019).

125. Lee, S. H., Stephens, J. L. & Englund, P. T. A fatty-acid synthesis mechanism specialized for parasitism. *Nat. Rev. Microbiol.* **5**, 287–297 (2007).

126. van Hellemond, J. J., Opperdoes, F. R. & Tielens, A. G. M. The extraordinary mitochondrion and unusual citric acid cycle in Trypanosoma brucei. *Biochem. Soc. Trans.* **33**, 967–971 (2005).

127. Saunders, E. C. *et al.* Isotopomer profiling of Leishmania mexicana promastigotes reveals important roles for succinate fermentation and aspartate uptake in tricarboxylic acid cycle (TCA) anaplerosis, glutamate synthesis, and growth. *J. Biol. Chem.* **286**, 27706–27717 (2011).

128. Saunders, E. C. *et al.* Induction of a stringent metabolic response in intracellular stages of Leishmania mexicana leads to increased dependence on mitochondrial metabolism. *PLoS Pathog.* **10**, e1003888 (2014).

129. Docampo, R., de Boiso, J. F. & Stoppani, A. O. Tricarboxylic acid cycle operation at the kinetoplast-mitochondrion complex of Trypanosoma cruzi. *Biochim. Biophys. Acta* **502**, 466–476 (1978).

130. Morita, Y. S., Paul, K. S. & Englund, P. T. Specialized fatty acid synthesis in African trypanosomes: Myristate for GPI anchors. *Science* **288**, 140–143 (2000).

131. Doering, T. L., Pessin, M. S., Hart, G. W., Raben, D. M. & Englund, P. T. The fatty acids in unremodelled trypanosome glycosyl-phosphatidylinositols. *Biochem. J* **299 (Pt 3)**, 741–746 (1994).

132. Flugel, R. S., Hwangbo, Y., Lambalot, R. H., Cronan, J. E., Jr & Walsh, C. T. Holo-(acyl carrier protein) synthase and phosphopantetheinyl transfer in Escherichia coli. *J. Biol. Chem.* **275**, 959–968 (2000).

133. Riedel, S. L., Lu, J., Stahl, U. & Brigham, C. J. Lipid and fatty acid metabolism in Ralstonia eutropha: relevance for the biotechnological production of value-added products. *Appl. Microbiol. Biotechnol.* **98**, 1469–1483 (2014).

134. Zhu, G., Marchewka, M. J., Woods, K. M., Upton, S. J. & Keithly, J. S. Molecular analysis of a Type I fatty acid synthase in Cryptosporidium parvum. *Mol. Biochem. Parasitol.* **105**, 253–260 (2000).

135. Vigueira, P. A. & Paul, K. S. Trypanosoma brucei: inhibition of acetyl-CoA carboxylase by haloxyfop. *Exp. Parasitol.* **130**, 159–165 (2012).

136. Goad, L. J., Holz, G. G., Jr & Beach, D. H. Sterols of ketoconazole-inhibited Leishmania mexicana mexicana promastigotes. *Mol. Biochem. Parasitol.* **15**, 257–279 (1985).

137. Haughan, P. A., Chance, M. L. & Goad, L. J. Effects of an azasterol inhibitor of sterol 24-transmethylation on sterol biosynthesis and growth of Leishmania donovani promastigotes. *Biochem. J* **308 (Pt 1)**, 31–38 (1995).

138. Coppens, I. & Courtoy, P. J. Exogenous and endogenous sources of sterols in the culture-adapted procyclic trypomastigotes of Trypanosoma brucei. *Mol. Biochem. Parasitol.* **73**, 179–188 (1995).

139. De Cicco, N. N. T. *et al.* LDL uptake by Leishmania amazonensis: involvement of membrane lipid microdomains. *Exp. Parasitol.* **130**, 330–340 (2012).

140. Andrade-Neto, V. V. *et al.* The pharmacological inhibition of sterol biosynthesis in Leishmania is counteracted by enhancement of LDL endocytosis. *Acta Trop.* **119**, 194–198 (2011).

141. Pietrocola, F., Galluzzi, L., Bravo-San Pedro, J. M., Madeo, F. & Kroemer, G. Acetyl coenzyme A: a central metabolite and second messenger. *Cell Metab.* **21**, 805–821 (2015).

142. Hollebeke, J., Van Damme, P. & Gevaert, K. N-terminal acetylation and other functions of Nα-acetyltransferases. *Biol. Chem.* **393**, 291–298 (2012).

143. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat. Rev. Mol. Cell Biol.* **15**, 536–550 (2014).

144. Weinert, B. T. *et al.* Acetylation dynamics and stoichiometry in Saccharomyces cerevisiae. *Mol. Syst. Biol.* **10**, 716 (2014).

145. Cai, L., Sutter, B. M., Li, B. & Tu, B. P. Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol. Cell* **42**, 426–437 (2011).

146. Donohoe, D. R. *et al.* The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Mol. Cell* **48**, 612–626 (2012).

147. Lee, J. V. *et al.* Akt-dependent metabolic reprogramming regulates tumor cell histone acetylation. *Cell Metab.* **20**, 306–319 (2014).

148. Wellen, K. E. *et al.* ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* **324**, 1076–1080 (2009).

149. Shimazu, T. *et al.* Suppression of oxidative stress by β-hydroxybutyrate, an endogenous histone deacetylase inhibitor. *Science* **339**, 211–214 (2013).

150. Vallari, D. S. & Rock, C. O. Isolation and characterization of temperature-sensitive pantothenate kinase (coaA) mutants of Escherichia coli. *J. Bacteriol.* **169**, 5795–5800 (1987).

151. Gerdes, S. Y. *et al.* From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. *J. Bacteriol.* **184**, 4555–4572 (2002).

152. Vallari, D. S. & Rock, C. O. Isolation and characterization of Escherichia coli pantothenate permease (panF) mutants. *J. Bacteriol.* **164**, 136–142 (1985).

153. Jackowski, S. & Alix, J. H. Cloning, sequence, and expression of the pantothenate permease (panF) gene of Escherichia coli. *J. Bacteriol.* **172**, 3842–3848 (1990).

154. Reizer, J., Reizer, A. & Saier, M. H., Jr. The Na+/pantothenate symporter (PanF) of Escherichia coli is homologous to the Na+/proline symporter (PutP) of E. coli and the Na+/glucose symporters of mammals. *Res. Microbiol.* **141**, 1069–1072 (1990).

155. Sambandamurthy, V. K. *et al.* A pantothenate auxotroph of Mycobacterium tuberculosis is highly attenuated and protects mice against tuberculosis. *Nat. Med.* **8**, 1171–1174 (2002).

156. Hung, A. W. *et al.* Application of fragment growing and fragment linking to the discovery of inhibitors of Mycobacterium tuberculosis pantothenate synthetase. *Angew. Chem. Int. Ed Engl.* **48**, 8452–8456 (2009).

157. White, E. L. *et al.* A novel inhibitor of Mycobacterium tuberculosis pantothenate synthetase. *J. Biomol. Screen.* **12**, 100–105 (2007).

158. Freiberg, C. *et al.* Identification of novel essential Escherichia coli genes conserved among pathogenic bacteria. *J. Mol. Microbiol. Biotechnol.* **3**, 483–489 (2001).

159. Dunn, S. D. & Snell, E. E. Isolation of temperature-sensitive pantothenate kinase mutants of Salmonella typhimurium and mapping of the coaA gene. *J. Bacteriol.* **140**, 805–808 (1979).

160. Paige, C., Reid, S. D., Hanna, P. C. & Claiborne, A. The type III pantothenate kinase encoded by coaX is essential for growth of Bacillus anthracis. *J. Bacteriol.* **190**, 6271–6275 (2008).

161. Clifton, G., Bryant, S. R. & Skinner, C. G. N'-(substituted) pantothenamides, antimetabolites of pantothenic acid. *Arch. Biochem. Biophys.* **137**, 523–528 (1970).

162. Choudhry, A. E. *et al.* Inhibitors of pantothenate kinase: novel antibiotics for staphylococcal infections. *Antimicrob. Agents Chemother.* **47**, 2051–2055 (2003).

163. Zhang, Y.-M. *et al.* Acyl carrier protein is a cellular target for the antibacterial action of the pantothenamide class of pantothenate antimetabolites. *J. Biol. Chem.* **279**, 50969–50975 (2004).

164. Strauss, E. & Begley, T. P. The antibiotic activity of N-pentylpantothenamide results from its conversion to ethyldethia-coenzyme a, a coenzyme a antimetabolite. *J. Biol. Chem.* **277**, 48205–48209 (2002).

165. Abiko, Y., Tomikawa, M. & Shimizu, M. Enzymatic Conversion of Pantothenylalcohol to Pantothenic Acid. *THE JOURNAL OF VITAMINOLOGY* (1968).

166. Snell, E. E. & Shive, W. GROWTH INHIBITION BY ANALOGUES OF PANTOTHENIC ACID. PANTOTHENYL ALCOHOL AND RELATED COMPOUNDS. *J. Biol. Chem.* **158**, 551–559 (1945).

167. Chohnan, S., Murase, M., Kurikawa, K., Higashi, K. & Ogata, Y. Antimicrobial activity of pantothenol against staphylococci possessing a prokaryotic type II pantothenate kinase. *Microbes Environ.* **29**, 224–226 (2014).

168. Kumar, P., Chhibber, M. & Surolia, A. How pantothenol intervenes in Coenzyme-A biosynthesis of Mycobacterium tuberculosis. *Biochem. Biophys. Res. Commun.* **361**, 903–909 (2007).

169. Gupta, A., Sharma, P., Singh, T. P. & Sharma, S. Phosphopantetheine Adenylyltransferase: A promising drug target to combat antibiotic resistance. *Biochim. Biophys. Acta: Proteins Proteomics* **1869**, 140566 (2021).

170. Gupta, A. *et al.* Structural and binding studies of phosphopantetheine adenylyl transferase from Acinetobacter baumannii. *Biochim. Biophys. Acta: Proteins Proteomics* **1867**, 537–547 (2019).

171. Mishra, P. K. & Drueckhammer, D. G. Coenzyme A Analogues and Derivatives: Synthesis and Applications as Mechanistic Probes of Coenzyme A Ester-Utilizing Enzymes. *Chem. Rev.* **100**, 3283–3310 (2000).

172. Genschel, U. Coenzyme A biosynthesis: reconstruction of the pathway in archaea and an evolutionary scenario based on comparative genomics. *Mol. Biol. Evol.* **21**, 1242–1251 (2004).

173. Kurtov, D., Kinghorn, J. R. & Unkles, S. E. The Aspergillus nidulans panB gene encodes ketopantoate hydroxymethyltransferase, required for biosynthesis of pantothenate and Coenzyme A. *Mol. Gen. Genet.* **262**, 115–120 (1999).

174. Genschel, U., Powell, C. A., Abell, C. & Smith, A. G. The final step of pantothenate biosynthesis in higher plants: cloning and characterization of pantothenate synthetase from Lotus japonicus and Oryza sativum (rice). *Biochem. J* **341 (Pt 3)**, 669–678 (1999).

175. White, W. H., Gunyuzlu, P. L. & Toyn, J. H. Saccharomyces cerevisiae is capable of de Novo pantothenic acid biosynthesis involving a novel pathway of beta-alanine production from spermine. *J. Biol. Chem.* **276**, 10794–10800 (2001).

176. Gojkovic, Z., Jahnke, K., Schnackerz, K. D. & Piskur, J. PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. *J. Mol. Biol.* **295**, 1073–1087 (2000).

177. Gojković, Z., Sandrini, M. P. & Piskur, J. Eukaryotic beta-alanine synthases are functionally related but have a high degree of structural diversity. *Genetics* **158**, 999–1011 (2001).

178. Stolz, J., Caspari, T., Carr, A. M. & Sauer, N. Cell division defects of Schizosaccharomyces pombe liz1- mutants are caused by defects in pantothenate uptake. *Eukaryot. Cell* **3**, 406–412 (2004).

179. Stolz, J. & Sauer, N. The fenpropimorph resistance gene FEN2 from Saccharomyces cerevisiae encodes a plasma membrane H+-pantothenate symporter. *J. Biol. Chem.* **274**, 18747–18752 (1999).

180. Moynihan, E. B. & Enoch, T. Liz1p, a novel fission yeast membrane protein, is required for normal cell division when ribonucleotide reductase is inhibited. *Mol. Biol. Cell* **10**, 245–257 (1999).

181. Marcireau, C., Joets, J., Pousset, D., Guilloton, M. & Karst, F. FEN2: a gene implicated in the catabolite repression-mediated regulation of ergosterol biosynthesis in yeast. *Yeast* **12**, 531–539 (1996).

182. Prasad, P. D. *et al.* Cloning and functional expression of a cDNA encoding a mammalian sodium-dependent vitamin transporter mediating the uptake of pantothenate, biotin, and lipoate. *J. Biol. Chem.* **273**, 7501–7506 (1998).

183. Giaever, G. *et al.* Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387–391 (2002).

184. Spry, C., Kirk, K. & Saliba, K. J. Coenzyme A biosynthesis: An antimicrobial drug target. *FEMS Microbiology Reviews* vol. 32 56–106 Preprint at https://doi.org/10.1111/j.1574-6976.2007.00093.x (2008).

185. Müller, S. & Kappes, B. Vitamin and cofactor biosynthesis pathways in Plasmodium and other apicomplexan parasites. *Trends Parasitol.* **23**, 112–121 (2007).

186. Divo, A. A., Geary, T. G., Davis, N. L. & Jensen, J. B. Nutritional requirements of Plasmodium falciparum in culture. I. Exogenously supplied dialyzable components necessary for continuous growth. *J. Protozool.* **32**, 59–64 (1985).

187. Saliba, K. J., Ferru, I. & Kirk, K. Provitamin B5 (pantothenol) inhibits growth of the intraerythrocytic malaria parasite. *Antimicrob. Agents Chemother.* **49**, 632–637 (2005).

188. Trager, W. FURTHER STUDIES ON THE SURVIVAL AND DEVELOPMENT IN VITRO OF A MALARIAL PARASITE. *J. Exp. Med.* **77**, 411–420 (1943).

189. Brackett, S., Waletzky, E. & Baker, M. The Relation between Panthothenic Acid and Plasmodium gallinaceum Infections in the Chicken and the Antimalarial Activity of Analogues of Pantothenic Acid. *J. Parasitol.* **32**, 453–462 (1946).

190. Saliba, K. J., Horner, H. A. & Kirk, K. Transport and metabolism of the essential vitamin pantothenic acid in human erythrocytes infected with the malaria parasite Plasmodium falciparum. *J. Biol. Chem.* **273**, 10190–10195 (1998).

191. Desai, S. A., Krogstad, D. J. & McCleskey, E. W. A nutrient-permeable channel on the intraerythrocytic malaria parasite. *Nature* **362**, 643–646 (1993).

192. Augagneur, Y. *et al.* Identification and functional analysis of the primary pantothenate transporter, PfPAT, of the human malaria parasite Plasmodium falciparum. *J. Biol. Chem.* **288**, 20558–20567 (2013).

193. Saliba, K. J. & Kirk, K. H+-coupled pantothenate transport in the intracellular malaria parasite. *J. Biol. Chem.* **276**, 18115–18121 (2001).

194. Hart, R. J., Lawres, L., Fritzen, E., Ben Mamoun, C. & Aly, A. S. I. Plasmodium yoelii vitamin B5 pantothenate transporter candidate is essential for parasite transmission to the mosquito. *Sci. Rep.* **4**, 5665 (2014).

195. Nurkanto, A. *et al.* Characterization of Plasmodium falciparum Pantothenate Kinase and Identification of Its Inhibitors From Natural Products. *Front. Cell. Infect. Microbiol.* **11**, 639065 (2021).

196. Hart, R. J. *et al.* Genetic Characterization of Plasmodium Putative Pantothenate Kinase Genes Reveals Their Essential Role in Malaria Parasite Transmission to the Mosquito. *Sci. Rep.* **6**, 33518 (2016).

197. de Vries, L. E., Lunghi, M., Krishnan, A., Kooij, T. W. A. & Soldati-Favre, D. Pantothenate and CoA biosynthesis in Apicomplexa and their promise as antiparasitic drug targets. *PLoS Pathog.* **17**, e1010124 (2021).

198. Ralph, S. A. *et al.* Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast. *Nat. Rev. Microbiol.* **2**, 203–216 (2004).

199. Zhyvoloup, A. *et al.* Subcellular localization and regulation of coenzyme A synthase. *J. Biol. Chem.* **278**, 50316–50321 (2003).

200. Trager, W. & Brohn, F. H. Coezyme A requirement of malaria parasites: effects of coenzyme A precursors on extracellular development in vitro of Plasmodium lophurae. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 1834–1837 (1975).

201. Trager, W. Coenzyme A and the antimalarial action in vitro of antipantothenate against Plasmodium lophurae, P. coatneyi and P. falciparum. *Trans. N. Y. Acad. Sci.* **28**, 1094–1108 (1966).

202. Brohn, F. H. & Trager, W. Coenzyme A requirement of malaria parasites: enzymes of coenzyme A biosynthesis in normal duck erythrocytes and erythrocytes infected with Plasmodium lophurae. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 2456–2458 (1975).

203. Winterbottom, R. & Clapp, J. W. Studies in chemotherapy; amides of pantoyltaurine. *J. Am. Chem. Soc.* **69**, 1393–1401 (1947).
204. Trager, W. Further studies on the effects of antipantothenates on malaria parasites (Plasmodium coatneyi and P. falciparum) in vitro. *J. Protozool.* **18**, 232–239 (1971).
205. Senear, A. E., Rapport, M. M. & Koepfli, J. B. The synthesis of potential antimalarials; derivatives of alpha, gamma-dihydroxy-beta, beta-dimethyl-N-ethylbutyramide. *J. Biol. Chem.* **167**, 229–234 (1947).
206. Schalkwijk, J. *et al. Antimalarial pantothenamide metabolites target acetyl-coenzyme A biosynthesis in Plasmodium falciparum*. vol. 11 9917 http://stm.sciencemag.org/ (2019).
207. Spry, C. *et al.* Pantothenamides are potent, on-target inhibitors of Plasmodium falciparum growth when serum pantetheinase is inactivated. *PLoS One* **8**, e54974 (2013).
208. Macuamule, C. J. *et al.* A pantetheinase-resistant pantothenamide with potent, on-target, and selective antiplasmodial activity. *Antimicrob. Agents Chemother.* **59**, 3666–3668 (2015).
209. de Villiers, M. *et al.* Structural modification of pantothenamides counteracts degradation by pantetheinase and improves antiplasmodial activity. *ACS Med. Chem. Lett.* **4**, 784–789 (2013).
210. Spry, C. *et al.* Toward a Stable and Potent Coenzyme A-Targeting Antiplasmodial Agent: Structure-Activity Relationship Studies of N-Phenethyl-α-methyl-pantothenamide. *ACS Infect Dis* **6**, 1844–1854 (2020).
211. de Vries, L. E. *et al.* Preclinical characterization and target validation of the antimalarial pantothenamide MMV693183. *Nat. Commun.* **13**, 2158 (2022).
212. Lunghi, M. *et al.* Pantothenate biosynthesis is critical for chronic infection by the neurotropic parasite Toxoplasma gondii. *Nat. Commun.* **13**, 345 (2022).
213. Mageed, S. N. *et al.* Pantothenic acid biosynthesis in the parasite Toxoplasma gondii: a target for chemotherapy. *Antimicrob. Agents Chemother.* **58**, 6345–6353 (2014).
214. Kidder, G. W. & Dutta, B. N. The growth and nutrition of Crithidia fasciculata. *J. Gen. Microbiol.* **18**, 621–638 (1958).

215. Cowperthwaite, J., Weber, M. M., Packer, L. & Hutner, S. H. Nutrition of Herpetomonas (Strigomonas) culicidarum. *Ann. N. Y. Acad. Sci.* **56**, 972–981 (1953).

216. Newton, B. A. A synthetic growth medium for the trypanosomid flagellate Strigomonas (Herpetomonas) oncopelti. *Nature* **177**, 279–280 (1956).

217. Mundim, M. H., Roitman, I., Hermans, M. A. & Kitajima, E. W. Simple nutrition of Crithidia deanei, a reduviid trypanosomatid with an endosymbiont. *J. Protozool.* **21**, 518–521 (1974).

218. Klein, C. C. *et al.* Biosynthesis of vitamins and cofactors in bacterium-harbouring trypanosomatids depends on the symbiotic association as revealed by genomic analyses. *PLoS One* **8**, e79786 (2013).

219. Fletcher, S. *et al.* Biological characterization of chemically diverse compounds targeting the Plasmodium falciparum coenzyme A synthesis pathway. *Parasites and Vectors* **9**, 1–13 (2016).

220. Lux, H., Heise, N., Klenner, T., Hart, D. & Opperdoes, F. R. Ether--lipid (alkyl-phospholipid) metabolism and the mechanism of action of ether--lipid analogues in Leishmania. *Mol. Biochem. Parasitol.* **111**, 1–14 (2000).

221. Dorlo, T. P. C., Balasegaram, M., Beijnen, J. H. & de Vries, P. J. Miltefosine: a review of its pharmacology and therapeutic efficacy in the treatment of leishmaniasis. *J. Antimicrob. Chemother.* **67**, 2576–2597 (2012).

222. Croft, S. L., Seifert, K. & Duchêne, M. Antiprotozoal activities of phospholipid analogues. *Mol. Biochem. Parasitol.* **126**, 165–172 (2003).

223. Calder, R. B. *et al.* Cloning and characterization of a eukaryotic pantothenate kinase gene (panK) from Aspergillus nidulans. *J. Biol. Chem.* **274**, 2014–2020 (1999).

224. Vallari, D. S., Jackowski, S. & Rock, C. O. Regulation of pantothenate kinase by coenzyme A and its thioesters. *J. Biol. Chem.* **262**, 2468–2471 (1987).

225. Jackowski, S. & Rock, C. O. Regulation of coenzyme A biosynthesis. *J. Bacteriol.* **148**, 926–932 (1981).

226. Robishaw, J. D., Berkich, D. & Neely, J. R. Rate-limiting step and control of coenzyme A synthesis in cardiac muscle. *J. Biol. Chem.* **257**, 10967–10972 (1982).
227. Cronan, J. E., Jr, Littel, K. J. & Jackowski, S. Genetic and biochemical analyses of pantothenate biosynthesis in Escherichia coli and Salmonella typhimurium. *J. Bacteriol.* **149**, 916–922 (1982).
228. Song, W. J. & Jackowski, S. Cloning, sequencing, and expression of the pantothenate kinase (coaA) gene of Escherichia coli. *J. Bacteriol.* **174**, 6411–6417 (1992).
229. Maas, W. K. & Davis, B. D. Pantothenate studies. I. Interference by D-serine and L-aspartic acid with pantothenate synthesis in Escherichia coli. *J. Bacteriol.* **60**, 733–745 (1950).
230. Vadali, R. V., Bennett, G. N. & San, K.-Y. Cofactor engineering of intracellular CoA/acetyl-CoA and its effect on metabolic flux redistribution in Escherichia coli. *Metab. Eng.* **6**, 133–139 (2004).
231. Brand, L. A. & Strauss, E. Characterization of a new pantothenate kinase isoform from Helicobacter pylori. *J. Biol. Chem.* **280**, 20185–20188 (2005).
232. Awasthy, D. *et al.* Essentiality and functional analysis of type I and type III pantothenate kinases of Mycobacterium tuberculosis. *Microbiology* **156**, 2691–2701 (2010).
233. Yao, J., Subramanian, C., Rock, C. O. & Jackowski, S. Human pantothenate kinase 4 is a pseudo-pantothenate kinase. *Protein Sci.* **28**, 1031–1047 (2019).
234. Zhou, B. *et al.* A novel pantothenate kinase gene (PANK2) is defective in Hallervorden-Spatz syndrome. *Nat. Genet.* **28**, 345–349 (2001).
235. Song, W. J. & Jackowski, S. Kinetics and regulation of pantothenate kinase from Escherichia coli. *J. Biol. Chem.* **269**, 27051–27058 (1994).
236. Hong, B. S. *et al.* Crystal Structures of Human Pantothenate Kinases: INSIGHTS INTO ALLOSTERIC REGULATION AND MUTATIONS LINKED TO A NEURODEGENERATION DISORDER*. *J. Biol. Chem.* **282**, 27984–27993 (2007).

237. Yang, K. *et al.* Crystal structure of a type III pantothenate kinase: insight into the mechanism of an essential coenzyme A biosynthetic enzyme universally distributed in bacteria. *J. Bacteriol.* **188**, 5532–5540 (2006).

238. Tjhin, E. T., Howieson, V. M., Spry, C., van Dooren, G. G. & Saliba, K. J. A novel heteromeric pantothenate kinase complex in apicomplexan parasites. *PLoS Pathog.* **17**, e1009797 (2021).

239. Subramanian, C. *et al.* Allosteric regulation of mammalian pantothenate kinase. *J. Biol. Chem.* **291**, 22302–22314 (2016).

240. Souza, R. T. *et al.* Genome size, karyotype polymorphism and chromosomal evolution in Trypanosoma cruzi. *PLoS One* **6**, e23042 (2011).

241. Luís Reis-Cunha, J. *et al.* Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct Trypanosoma cruzi strains. (2011) doi:10.1186/s12864-015-1680-4.

242. Araújo, P. R. & Teixeira, S. M. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in Trypanosoma cruzi: a review. *Mem. Inst. Oswaldo Cruz* **106**, 257–266 (2011).

243. Cribb, P., Esteban, L., Trochine, A., Girardini, J. & Serra, E. Trypanosoma cruzi TBP shows preference for C/G-rich DNA sequences in vitro. *Exp. Parasitol.* **124**, 346–349 (2010).

244. Palenchar, J. B. & Bellofatto, V. Gene transcription in trypanosomes. *Mol. Biochem. Parasitol.* **146**, 135–141 (2006).

245. Teixeira, S. M. R. & daRocha, W. D. Control of gene expression and genetic manipulation in the Trypanosomatidae. *Genet. Mol. Res.* **2**, 148–158 (2003).

246. De Gaudenzi, J. G., D'Orso, I. & Frasch, A. C. C. RNA recognition motif-type RNA-binding proteins in Trypanosoma cruzi form a family involved in the interaction with specific transcripts in vivo. *J. Biol. Chem.* **278**, 18884–18894 (2003).

247. Gingerich, T. J., Feige, J.-J. & LaMarre, J. AU-rich elements and the control of gene expression through regulated mRNA stability. *Anim. Health Res. Rev.* **5**, 49–63 (2004).

248. Kramer, S. & Carrington, M. Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. *Trends Parasitol.* **27**, 23–30 (2011).

249. Batista, J. A., Teixeira, S. M., Donelson, J. E., Kirchhoff, L. V. & de Sá, C. M. Characterization of a Trypanosoma cruzi poly(A)-binding protein and its genes. *Mol. Biochem. Parasitol.* **67**, 301–312 (1994).

250. D'Orso, I. & Frasch, A. C. TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes. *J. Biol. Chem.* **276**, 34801–34809 (2001).

251. Espinosa, J. M. *et al.* Trypanosoma cruzi poly-zinc finger protein: a novel DNA/RNA-binding CCHC-zinc finger protein. *Mol. Biochem. Parasitol.* **131**, 35–44 (2003).

252. Kolev, N. G., Tschudi, C. & Ullu, E. RNA Interference in Protozoan Parasites: Achievements and Challenges. *Eukaryot. Cell* **10**, 1156–1163 (2011).

253. Siomi, H. & Siomi, M. C. On the road to reading the RNA-interference code. *Nature* **457**, 396–404 (2009).

254. Alsford, S., Kawahara, T., Glover, L. & Horn, D. Tagging a T. brucei RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol. Biochem. Parasitol.* **144**, 142–148 (2005).

255. Kelso, A. A., Waldvogel, S. M., Luthman, A. J. & Sehorn, M. G. Homologous Recombination in Protozoan Parasites and Recombinase Inhibitors. *Front. Microbiol.* **8**, 1716 (2017).

256. Costa, F. C. *et al.* Expanding the toolbox for Trypanosoma cruzi: A parasite line incorporating a bioluminescence-fluorescence dual reporter and streamlined CRISPR/Cas9 functionality for rapid in vivo localisation and phenotyping. *PLoS Negl. Trop. Dis.* **12**, e0006388 (2018).

257. Aslett, M. *et al.* TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* **38**, D457-62 (2010).

258. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

259. Procter, J. B. *et al.* Alignment of Biological Sequences with Jalview. *Methods Mol. Biol.* **2231**, 203–224 (2021).

260. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

261. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

262. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC. https://pymol.org/2.

263. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

264. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225-9 (2011).

265. Gao, R., Huang, S.-Y. N., Marchand, C. & Pommier, Y. Biochemical characterization of human tyrosyl-DNA phosphodiesterase 2 (TDP2/TTRAP): a Mg(2+)/Mn(2+)-dependent phosphodiesterase specific for the repair of topoisomerase cleavage complexes. *J. Biol. Chem.* **287**, 30842–30852 (2012).

266. Chou, K.-M. & Cheng, Y.-C. An exonucleolytic activity of human apurinic/apyrimidinic endonuclease on 3' mispaired DNA. *Nature* **415**, 655–659 (2002).

267. Pfeifer, S. & Greiner-Stöffele, T. A recombinant exonuclease III homologue of the thermophilic archaeon Methanothermobacter thermautotrophicus. *DNA Repair* **4**, 433–444 (2005).

268. Zhang, C., Freddolino, P. L. & Zhang, Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Res.* **45**, W291–W299 (2017).

269. Wu, Q., Peng, Z., Zhang, Y. & Yang, J. COACH-D: improved protein-ligand binding sites prediction with refined ligand-binding poses through molecular docking. *Nucleic Acids Res.* **46**, W438–W442 (2018).

270. Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585-7 (2007).

271. Alsford, S. & Horn, D. Single-locus targeting constructs for reliable regulated RNAi and transgene expression in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **161**, 76–79 (2008).

272. Hirumi, H. & Hirumi, K. Continuous cultivation of Trypanosoma brucei blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J. Parasitol.* **75**, 985–989 (1989).

273. Günzl, A. & Schimanski, B. Tandem affinity purification of proteins. *Curr. Protoc. Protein Sci.* **Chapter 19**, Unit 19.19 (2009).

274. Peng, D. & Tarleton, R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. *Microb Genom* **1**, e000033 (2015).

275. Dean, S. *et al.* A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. *Open Biol.* **5**, 140197 (2015).

276. Beneke, T. *et al.* A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *Royal Society open science* **4**, 170095 (2017).

277. Cross, M. *et al.* J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. *Mol. Microbiol.* **46**, 37–47 (2002).

278. Burkard, G., Fragoso, C. M. & Roditi, I. Highly efficient stable transformation of bloodstream forms of Trypanosoma brucei. *Mol. Biochem. Parasitol.* **153**, 220–223 (2007).

279. Kelly, J. M. Isolation of DNA and RNA from Leishmania. *Methods Mol. Biol.* **21**, 123–131 (1993).

280. Kilmer, G., Wolf, B., Starwalt, S., Judd, G. & Webb, B. ELECTROBLOT TRANSFER BUFFER. *Patent* (2014).

281. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108 (2008).

282. Zhang, Y.-M., Rock, C. O. & Jackowski, S. Biochemical properties of human pantothenate kinase 2 isoforms and mutations linked to pantothenate kinase-associated neurodegeneration. *J. Biol. Chem.* **281**, 107–114 (2006).

283. Ago, H. *et al.* Structural basis of the sphingomyelin phosphodiesterase activity in neutral sphingomyelinase from Bacillus cereus. *J. Biol. Chem.* **281**, 16157–16167 (2006).

284. Sawai, H. *et al.* Identification of ISC1 (YER019w) as inositol phosphosphingolipid phospholipase C in Saccharomyces cerevisiae. *J. Biol. Chem.* **275**, 39793–39798 (2000).

285. Zhang, O. *et al.* Degradation of host sphingomyelin is essential for Leishmania virulence. *PLoS Pathog.* **5**, e1000692 (2009).

286. Dickie, E. A., Young, S. A. & Smith, T. K. Substrate specificity of the neutral sphingomyelinase from Trypanosoma brucei. *Parasitology* **146**, 604–616 (2019).

287. Young, S. A. & Smith, T. K. The essential neutral sphingomyelinase is involved in the trafficking of the variant surface glycoprotein in the bloodstream form of Trypanosoma brucei. *Mol. Microbiol.* **76**, 1461–1482 (2010).

288. De Lederkremer, R. M., Agusti, R. & Docampo, R. Inositolphosphoceramide metabolism in Trypanosoma cruzi as compared with other trypanosomatids. *J. Eukaryot. Microbiol.* **58**, 79–87 (2011).

289. Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M. & Escalante-Semerena, J. C. The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry* **42**, 2866–2873 (2003).

290. Gahloth, D. *et al.* Structures of carboxylic acid reductase reveal domain dynamics underlying catalysis. *Nat. Chem. Biol.* **13**, 975–981 (2017).

291. Obado, S. O. *et al.* Centromere-associated topoisomerase activity in bloodstream form Trypanosoma brucei. *Nucleic Acids Res.* **39**, 1023–1033 (2011).

292. Dean, S., Sunter, J. D. & Wheeler, R. J. Tryptag.org: A Trypanosome Genome-wide Protein Localisation Resource. *Trends Parasitol.* **33**, 80–82 (2017).

293. Srinivasan, B. *et al.* Extracellular 4'-phosphopantetheine is a source for intracellular coenzyme A synthesis. *Nat. Chem. Biol.* **11**, 784–792 (2015).

294. Nurkanto, A. *et al.* Characterization and validation of Entamoeba histolytica pantothenate kinase as a novel anti-amebic drug target. *Int. J. Parasitol. Drugs Drug Resist.* **8**, 125–136 (2018).

295. Shockey, J. M., Fulda, M. S. & Browse, J. A. Arabidopsis contains nine long-chain acyl-coenzyme a synthetase genes that participate in fatty acid and glycerolipid metabolism. *Plant Physiol.* **129**, 1710–1722 (2002).

296. Minning, T. A., Weatherly, D. B., Atwood, J., 3rd, Orlando, R. & Tarleton, R. L. The steady-state transcriptome of the four major life-cycle stages of Trypanosoma cruzi. *BMC Genomics* **10**, 370 (2009).

297. Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* 2022.04.08.487609 (2022) doi:10.1101/2022.04.08.487609.

298. Nakai, K. & Horton, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **24**, 34–36 (1999).

299. Teasdale, R. D. & Jackson, M. R. SIGNAL-MEDIATED SORTING OF MEMBRANE PROTEINS BETWEEN THE ENDOPLASMIC RETICULUM AND THE GOLGI APPARATUS. (2003) doi:10.1146/annurev.cellbio.12.1.27.

300. Tinti, M. & Ferguson, M. A. J. Visualisation of experimentally determined and predicted protein N-glycosylation and predicted glycosylphosphatidylinositol anchor addition in Trypanosoma brucei. *Wellcome Open Res.* **7**, 33 (2022).

301. Xu, C. & Ng, D. T. W. Glycosylation-directed quality control of protein folding. *Nat. Rev. Mol. Cell Biol.* **16**, 742–752 (2015).

302. Leonardi, R., Zhang, Y.-M., Lykidis, A., Rock, C. O. & Jackowski, S. Localization and regulation of mouse pantothenate kinase 2. *FEBS Lett.* **581**, 4639–4644 (2007).

303. Naquet, P., Kerr, E. W., Vickers, S. D. & Leonardi, R. Regulation of coenzyme A levels by degradation: the “Ins and Outs.” *Prog. Lipid Res.* **78**, 101028 (2020).

304. Chiurillo, M. A., Jensen, B. C. & Docampo, R. Drug Target Validation of the Protein Kinase AEK1, Essential for Proliferation, Host Cell Invasion, and Intracellular Replication of the Human Pathogen Trypanosoma cruzi. *Microbiol Spectr* **9**, e0073821 (2021).

305. Leonardi, R. *et al.* A pantothenate kinase from Staphylococcus aureus refractory to feedback regulation by coenzyme A. *J. Biol. Chem.* **280**, 3314–3322 (2005).

306. Alfonso-Pecchio, A., Garcia, M., Leonardi, R. & Jackowski, S. Compartmentalization of mammalian pantothenate kinases. *PLoS One* **7**, e49509 (2012).

307. Vos, S. M. *et al.* Direct control of type IIA topoisomerase activity by a chromosomally encoded regulatory protein. *Genes Dev.* **28**, 1485–1497 (2014).

308. Rovinskiy, N., Agbleke, A. A., Chesnokova, O., Pang, Z. & Higgins, N. P. Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome. *PLoS Genet.* **8**, e1002845 (2012).

309. Tu, B. P., Kudlicki, A., Rowicka, M. & McKnight, S. L. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science* **310**, 1152–1158 (2005).

310. Chen, Z., Odstrcil, E. A., Tu, B. P. & McKnight, S. L. Restriction of DNA replication to the reductive phase of the metabolic cycle protects genome integrity. *Sciece* **316**, 1916–1919 (2007).

311. Vallari, D. S. & Jackowski, S. Biosynthesis and degradation both contribute to the regulation of coenzyme A content in Escherichia coli. *J. Bacteriol.* **170**, 3961–3966 (1988).

312. Kerr, E. W., Shumar, S. A. & Leonardi, R. Nudt8 is a novel CoA diphosphohydrolase that resides in the mitochondria. *FEBS Lett.* **593**, 1133–1143 (2019).

313. Shumar, S. A. *et al.* Nudt19 is a renal CoA diphosphohydrolase with biochemical and regulatory properties that are distinct from the hepatic Nudt7 isoform. *J. Biol. Chem.* **293**, 4134–4148 (2018).

314. Taylor, M. C. & Kelly, J. M. pTcINDEX: a stable tetracycline-regulated expression vector for Trypanosoma cruzi. *BMC Biotechnol.* **6**, 32 (2006).

315. Chung, H. K. *et al.* Tunable and reversible drug control of protein production via a self-excising degron. *Nat. Chem. Biol.* **11**, 713–720 (2015).

316. Yagoubat, A. *et al.* Universal highly efficient conditional knockout system in Leishmania, with a focus on untranscribed region preservation. *Cell. Microbiol.* **22**, e13159 (2020).
317. Roux, K. J., Kim, D. I., Burke, B. & May, D. G. BioID: A Screen for Protein-Protein Interactions. *Curr. Protoc. Protein Sci.* **91**, 19.23.1-19.23.15 (2018).
318. Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol.* **1266**, 171–184 (2015).
319. Vélez-Ramírez, D. E. *et al.* APEX2 Proximity Proteomics Resolves Flagellum Subdomains and Identifies Flagellum Tip-Specific Proteins in Trypanosoma brucei. *mSphere* **6**, (2021).
320. Geoghegan, V., Mottram, J. C. & Jones, N. G. Tag Thy Neighbour: Nanometre-Scale Insights Into Kinetoplastid Parasites With Proximity Dependent Biotinylation. *Front. Cell. Infect. Microbiol.* **12**, 894213 (2022).
321. Pyrih, J., Rašková, V., Škodová-Sveráková, I., Pánek, T. & Lukeš, J. ZapE/Afg1 interacts with Oxa1 and its depletion causes a multifaceted phenotype. *PLoS One* **15**, e0234918 (2020).
322. Dong, X., Lim, T. K., Lin, Q. & He, C. Y. Basal Body Protein TbSAF1 Is Required for Microtubule Quartet Anchorage to the Basal Bodies in Trypanosoma brucei. *MBio* **11**, (2020).
323. Geoghegan, V., Jones, N. G., Dowle, A. & Mottram, J. C. Protein kinase signalling at the Leishmania kinetochore captured by XL-BioID. *bioRxiv* 2021.07.08.451598 (2021) doi:10.1101/2021.07.08.451598.
324. Morriswood, B. *et al.* Novel bilobe components in Trypanosoma brucei identified using proximity-dependent biotinylation. *Eukaryot. Cell* **12**, 356–367 (2013).
325. McAllaster, M. R. *et al.* Proteomic identification of novel cytoskeletal proteins associated with TbPLK, an essential regulator of cell morphogenesis in Trypanosoma brucei. *Mol. Biol. Cell* **26**, 3013–3029 (2015).

326. Hu, H., Zhou, Q. & Li, Z. SAS-4 Protein in Trypanosoma brucei Controls Life Cycle Transitions by Modulating the Length of the Flagellum Attachment Zone Filament. *J. Biol. Chem.* **290**, 30453–30463 (2015).

327. Zhou, Q., Hu, H. & Li, Z. An EF-hand-containing Protein in Trypanosoma brucei Regulates Cytokinesis Initiation by Maintaining the Stability of the Cytokinesis Initiation Factor CIF1. *J. Biol. Chem.* **291**, 14395–14409 (2016).

328. Dang, H. Q. *et al.* Proximity Interactions among Basal Body Components in Trypanosoma brucei Identify Novel Regulators of Basal Body Biogenesis and Inheritance. *MBio* **8**, (2017).

329. De Pablos, L. M., Kelly, S., de Freitas Nascimento, J., Sunter, J. & Carrington, M. Characterization of RBP9 and RBP10, two developmentally regulated RNA-binding proteins in Trypanosoma brucei. *Open Biol.* **7**, (2017).

330. Pandey, M., Huang, Y., Lim, T. K., Lin, Q. & He, C. Y. Flagellar targeting of an arginine kinase requires a conserved lipidated protein intraflagellar transport (LIFT) pathway in Trypanosoma brucei. *J. Biol. Chem.* **295**, 11326–11336 (2020).

Appendix

A. Primers used in this study

F HindIII PTP	5'-GGGaagcttGAAGATCAGGTGGATCCTCG-3'
R BsmI PTP	5'-AGAgaatgctTCGGTTCAGGCGATCAGGTTG-3'

Table A.1 Primers used in pPOT-PTP cloning. Lower case letters indicate restriction sites *Hind*III and *Bsm*I respectively

N-F Tag	5'-TGTTGCTGTTGTTGCTGTTGTTCTCACCCAgtataatgcagacctgctgc-3'
N-R Tag	5'-TTTATCACCCCCCTGAAACGCATGGAAAGAactacccgatcctgatccag-3'
C-F Tag	5'-ATGCCGACGGATTCAGAAACCGCATCCAAAggttctggtagtggttccgg-3'
C-R Tag	5'-CTCGCCCATAGAAAACACTTAACATCCCCAccaatttgagagacctgtgc-3'
N-F gRNA Tag	5'-gaaattaatacgactcactataggTCCCCCCTCCCCTTAAATGTgttttagagctagaaatagc-3'
C-F gRNA Tag	5'-gaaattaatacgactcactataggAATAGGGTGGAGAAATGAATgttttagagctagaaatagc-3'
F KO	5'-GATTGTTGCTTTGCGCACAGTGCGCACAACgtataatgcagacctgctgc-3'
R KO	5'-TGGCTCCAAATAATCGGGTTGTCCCGTACAccttcctaaccaacctgcag-3'
N-gRNA KO	5'-gaaattaatacgactcactataggATTCTTCCCAGGGGCTGTGGgttttagagctagaaatagc-3'
C-gRNA KO	5'-gaaattaatacgactcactataggGTACGCCAGTTGCGTGATGTgttttagagctagaaatagc-3'
R-gRNA scaffold	5'-aaaagcaccgactcggtgccacttttcaagttgataacggactagccttattttaacttgctatttctagctctaaaac-3'

Table A.2 Primers used for CRISPR construct generation. Upper case letters indicate PanK-specific annealing sites

RNAi F	5’-GTAGATCgggcccggtaccAGCAAAGACACAAAAGATGG-3’
RNAi R	5’-AAGGATCtctaga**ggatcc**GGCACACCGGCATCGTTCTT-3’
TcPK F	5’-TTGCTAGcctgcaggATGTCTTTCCATGCGTTTCAGGGGGGTGAT-3’
TcPK R	5’-GCGATGTggcgcgccTTATTTGGATGCGGTTTCTGAATCC-3’
NG F1	5’-CGTGCCTGCGCGAGGTGTTT-3’
NG R1	5’-ATAAcctgcaggTTTGGATGCGGTTTCTGAAT-3’
NG F2	5’-GTTAGcctgcaggCGGTTCTGGTAGTGGTTCCGGTTCCGGTTCTGTCTCGAAAGGTGAGG AAGACAATATGGCC-3’
NG R2	5’-TCTAggcgcgccTTACTTATACAATTCGTCCATCCCCATCA-3’
1254 F	5’-CTTGTGAACATTGGTGCCGGTATTTCAATCATC-3’
1270 F	5’-GATGGCTCCCATGTTGCCGTTGGGGGAAGTCCC-3’
PKD F	5’-TTGCTAGcctgcaggATGCCGTGCTACACCAGTAGCGGCAAC-3’
PKD R	5’-GCGATGTggcgcgccTTATTTGGATGCGGTTTCTGAATCC-3’
SDH F	5’-ATACCTTTCTGccgcCTTTGGTGTGTCAGCAAC-3’
SDH R	5’-ACATACGGGCTTCGACTAC-3’
N His F	5’-GGACAGCAAATGGGTCGGGATCTGTACGACGATGACGATAAGTCTTTCCATGCGTCTCAG-3’
N His R	5’-ACCAGTCATGCTAGCCATACCATGATGATGATGATGATGCATGCAGGAAGCTTTATTTTATG-3’
C His F	5’-TTCTCATCATCATCACCACCACTAAGCGCCTAGAAAGTGTGAC-3’
C His R	5’-CCGGAACCGGAACCACTACCAGAACCTTTGGATGCGGTTTCTGAATC-3’

Table A.3 Primers used for RNAi construct preparation

Neon-PK F	5’-TGTCTCGAAAGGTGAGGAAGAC-3’
Neon-PK R	5’-GCGATAATGAGGCCGTTATC-3’
PK-Scarlet F	5’-GATTCTACGAAGCGATGGCGGG-3’
PK-Scarlet R	5’-AGCGCCACCATCCTCAAAGTTC-3’
PK-PTP F	5’-CAGCCTCCCGGTTTCTTTGAG-3’
PK-PTP R	5’-GAATGCTTCGGTTCAGGCGATCAGGTTG-3’

Table A.4 PCRs to confirm tagging of PanK in CLB T7-Cas9

KO F1	5’-CTT ATT CCG GCT TCG TCT TT-3’
KO R1	5’-AAGCCAACTAAATGGGCACC-3’
KO F2 Bla	5’-ATGCCTTTGTCTCAAGAAGA-3’
KO R2 Bla	5’-GCCCTCCCACACATAACCAG-3’
KO F2 Puro	5’-ATGACTGAATACAAGCCAAC-3’
KO R2 Puro	5’-TGTGCCATCAGATTACTCCG-3’

Table A.5 PCRs to confirm PanK gene replacement by antibiotic marker

Blast probe F	5’-ATGCCTTTGTCTCAAGAAGA-3’
Blast probe R	5’-GCCCTCCCACACATAACCAG-3’
PanK probe F	5’-ATGTCTTTCCATGCGTCTCA-3’
PanK probe R	5’-GCGATAATGAGGCCGTTATC-3’

Table A.6 Primers used for Southern Blotting Probe synthesis

TERT F	5’-GAGCGTGTGACTTCCGAAGG-3’
TERT R	5’-AGGAACTGTCACGGAGTTTGC-3’
TbPanK F	5’-GGTAATGTGGACAGGTTGTA-3’
TbPanK R	5’-ACGCCAACGTTGGGGCGACC-3’
TcPanK F	5’-GCGGGTGGATCGTTGAGACG-3’
TcPanK R	5’-TCGTTTTCCCTTTCCAAAGG-3’

Table A.7 Primers used for qPCR

NE Nt TcPK F	5’-TTGCTcatatgCATCATCATCATCATCACGGTTCTGGTAGTGGTTCCGGTTCCG GTTCTTCTTTCCATGCGTCTCAGGGGGGTGAT-3’
NE Nt PK-D F	5’-TTGCTcatatgCATCATCATCATCATCACGGTTCTGGTAGTGGTTCCGGTTCC GGTTCTCCGTGCTACACCAGTAGCGGCAACAGTAGC-3’
NE Nt R	5’-TTGCGTGggatccTTATTTGGATGCGGTTTCTGAATCCGTCGGCATGG-3’
NE Ct TcPK F	5’-TTGCTcat**atg**TCTTTCCATGCGTCTCAGGGGGGTG-3’
NE Ct R	5’- ATTCAGAAACCGCATCCAAAAGCAGCGGCCATCATCATCATCATCACTAAg gatccGCGAT-3’

Table A.8 Primers for recombinant *Tc*PanK expression with NEBExpress system. NdeI and BamHI restriction sites are indicated in lowercase.

HEK TcPK F	5’-TCTggtaccATGGGTCATCATCATCATCATCACGGTTCTGGTAGTGGTTCC GGTTCCGGTTCTTCTTTCCATGCGTCTCAGGGGGGTGAT-3’
HEK PK-D F	5’-TCTggtaccATGGGTCATCATCATCATCATCACGGTTCTGGTAGTGGTTCC GGTTCCGGTTCTCCGTGCTACACCAGTAGCGGCAACAGTAGC-3’
HEK R	5’-ATCGCgcggccgcTTATTTGGATGCGGTTTCTGAATCCG-3’

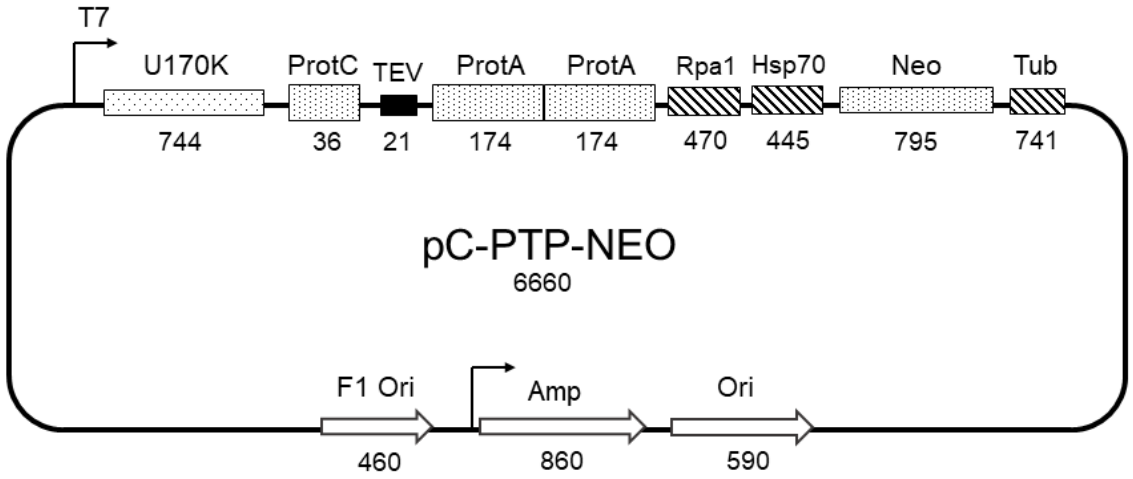
Table A.9 Primers for recombinant *Tc*PanK expression with HEK and pcDNA3.1+. Acc65I and NotI restriction sites are indicated in lowercase.

Trc TcPanK F	5’-TCTggatccGATCTTTCCATGCGTCTCAG-3’
Trc Pk-Domain F	5’-GACCTAggatccCGTGCTACACCAGTAGCGG-3’
pET TcPanK F	5’-TCTggatccTCTTTCCATGCGTCTCAGGGGGGTG-3’
pET Pk-Domain F	5’-TCTggatccCCGTGCTACACCAGTAGCGG-3’
Trc/PET R	5’-CGTTctcgagTTATTTGGATGCGGTTTCTGAAT-3’
Trc Pk-D-Mt F	5’-CTCTCATGTTGCCGTTGGGGGAAGTCCCATTG-3’
Trc Pk-D-Mt R	5’-CCATCCGGGCCGAGGCATT-3’

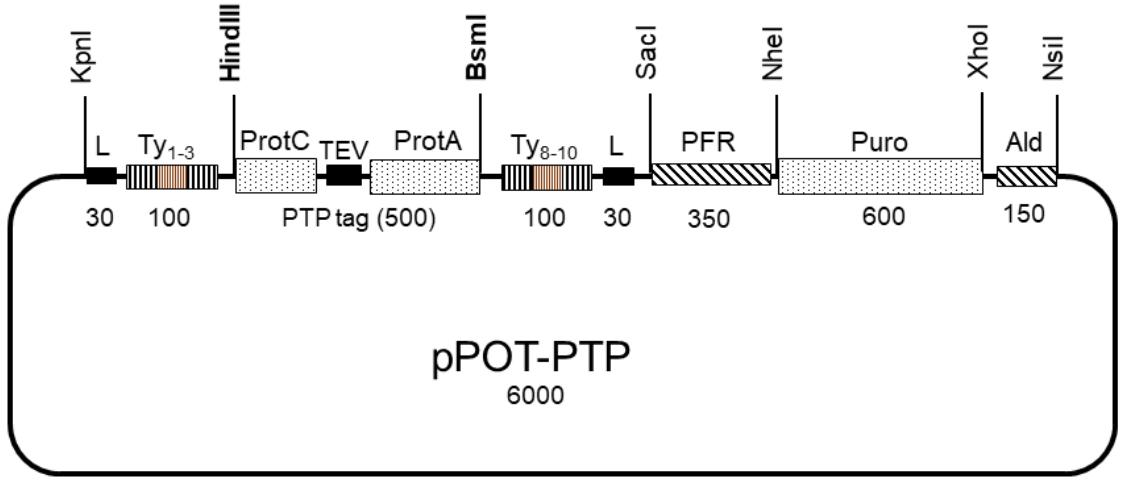
Table A.10 Primers for recombinant *Tc*PanK expression with E.coli/pTrcHisC and pET43.1a Restriction sites BamHI and XhoI are indicated in lowercase. Substituted residues are shown in red.

B. Supplementary Plasmid Maps

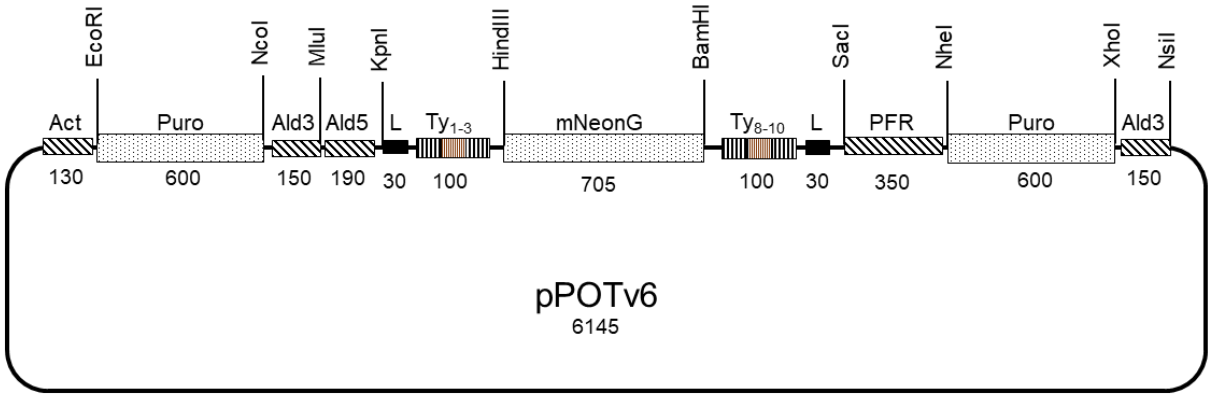
SF1-9: Simplified vector maps showing key regions and restriction sites, not to scale. Approximate sizes are indicated in base pairs.



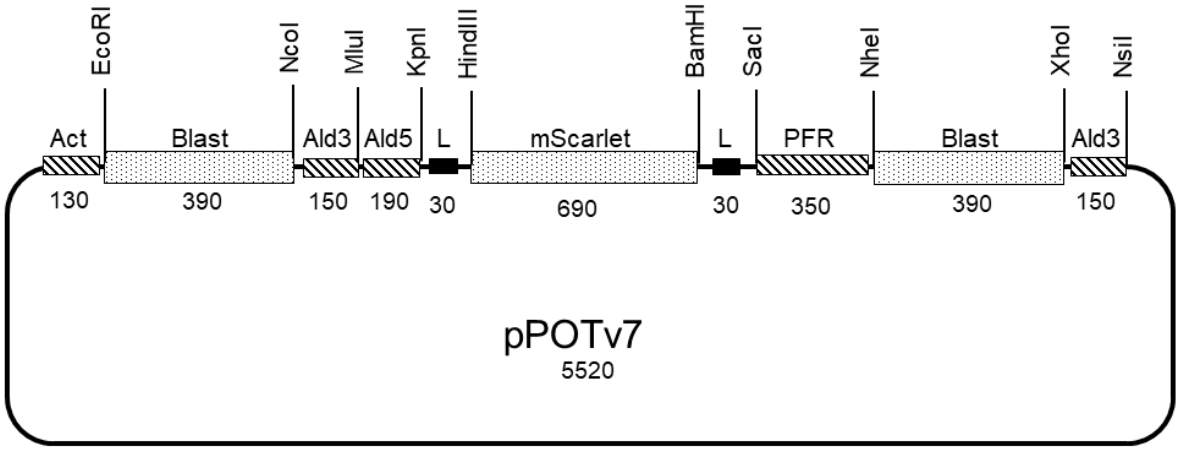
SF1 pC-PTP-NEO map: *T. brucei* PTP integration vector used in this project for pPOT-PTP cloning. T7: T7 promoter, U170K: C-terminal coding region of ribonucleoprotein *Tb*U1-70K (Stuffer), ProtC: protein C epitope (human hepatic plasma protein), TEV: tobacco etch virus (TEV) protease cleavage site, ProtA: IgG-binding unit of *Staphylococcus aureus* protein A, Rpa1: 3' flank from *Tb*RPA1 (the largest subunit of RNA polymerase I), Hsp70: intergenic region of heat shock protein 70 genes 2 and 3, Neo: Neomycin phosphotransferase gene, Tub: intergenic region of β- and α- tubulin genes, Ori: Origin of replication, Amp: Ampicillin resistance gene, F1 Ori: Origin of replication from f1 bacteriophage.



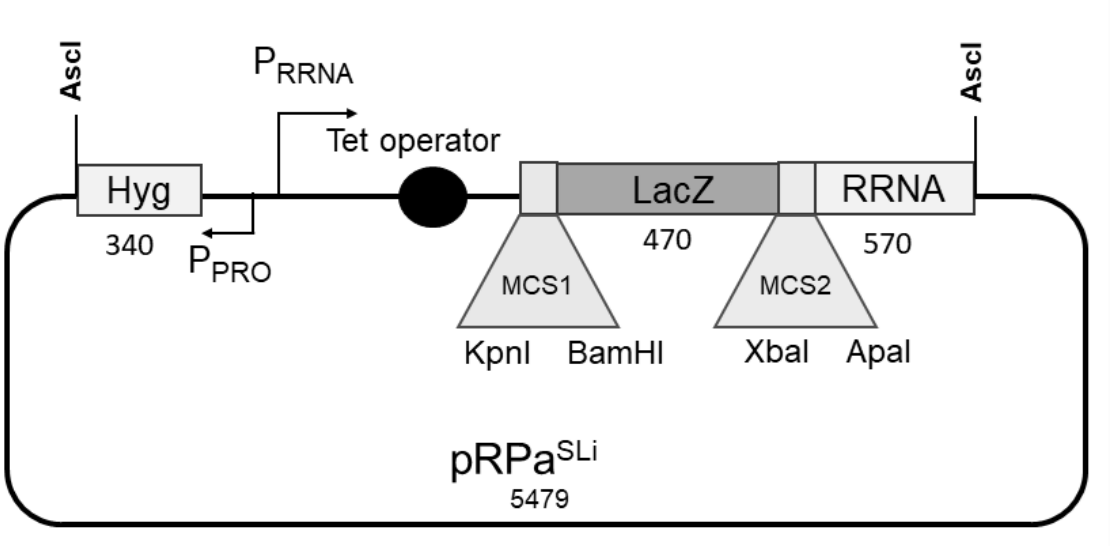
SF2 pPOT-PTP map: For C-terminal tagging, the illustrated region of pPOT-PTP was targeted to the TcPanK 3' terminus. L: Glycine-serine linker, Ty: Ty epitope repeats, ProtC: protein C epitope (human hepatic plasma protein). TEV: tobacco etch virus (TEV) protease cleavage site, ProtA: IgG-binding unit of *Staphylococcus aureus* protein A, PFR2: Paraflagellar rod intergenic region, Puro: puromycin-N-acetyltransferase ORF, Ald: Aldolase 3' UTR.



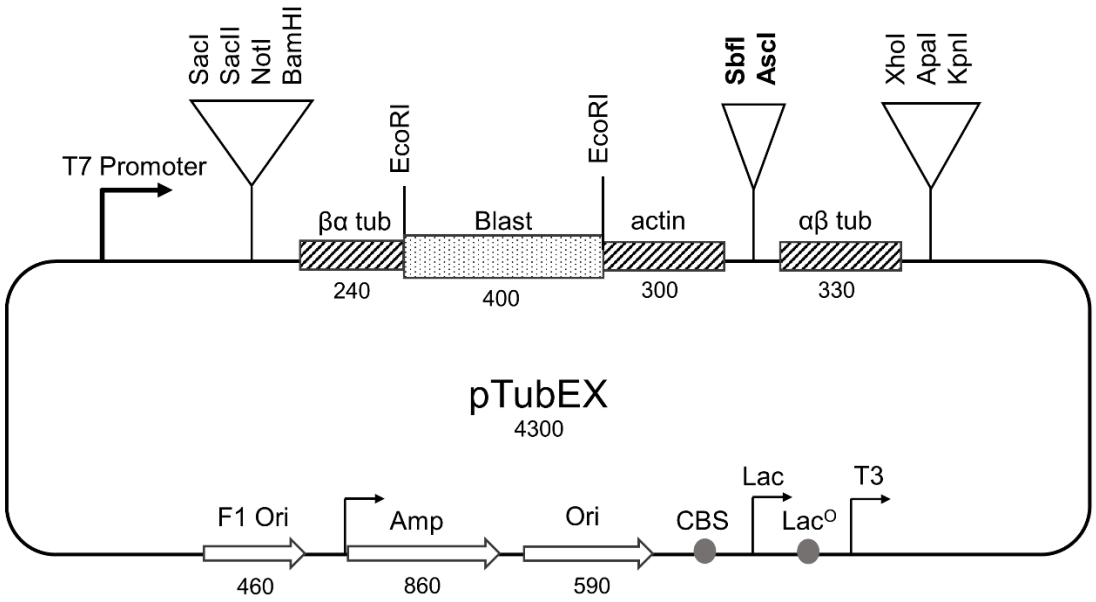
SF3 pPOTv6 puro-puro mNG map: A plasmid for N or C-terminal tagging of mNeonGreen. Act: Actin 5'UTR, Puro: puromycin-N-acetyltransferase ORF, Ald3/5: Aldolase 3'/5' UTR, L: Glycine-serine linker, Ty: Ty epitope repeats, mNeonG: mNeonGreen fluorescence ORF, PFR2: Paraflagellar rod intergenic region.



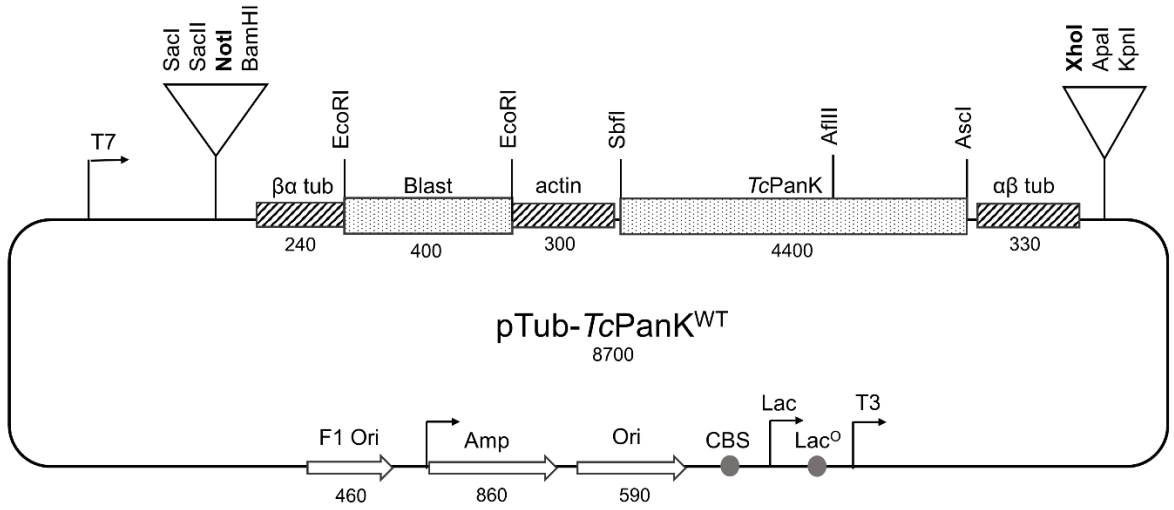
SF4 pPOTv7 blast-blast mScarlet map: A plasmid for N or C-terminal tagging of mNeonGreen. Act: Actin 5'UTR, Blast: blasticidin deaminase ORF, Ald3/5: Aldolase 3'/5' UTR, L: Glycine-serine linker, Ty: Ty epitope repeats, mNeonG: mNeonGreen fluorescence ORF, PFR2: Paraflagellar rod intergenic region.



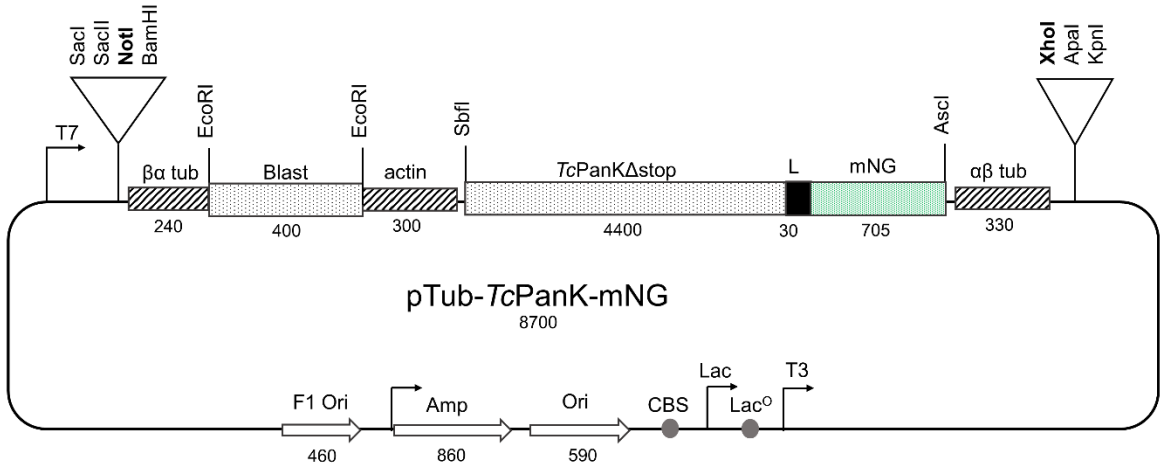
SF5 pRPa^{SLi} map: Plasmid for RNAi in *T. brucei* 2T1. Hyg: Partial hygromycin resistance gene fragment for locus integration, P_{PRO}: procyclin promoter, P_{RRNA}: ribosomal RNA promoter, LacZ: LacZ stuffer, RRNA: ribosomal RNA spacer. AscI was used for linearisation.



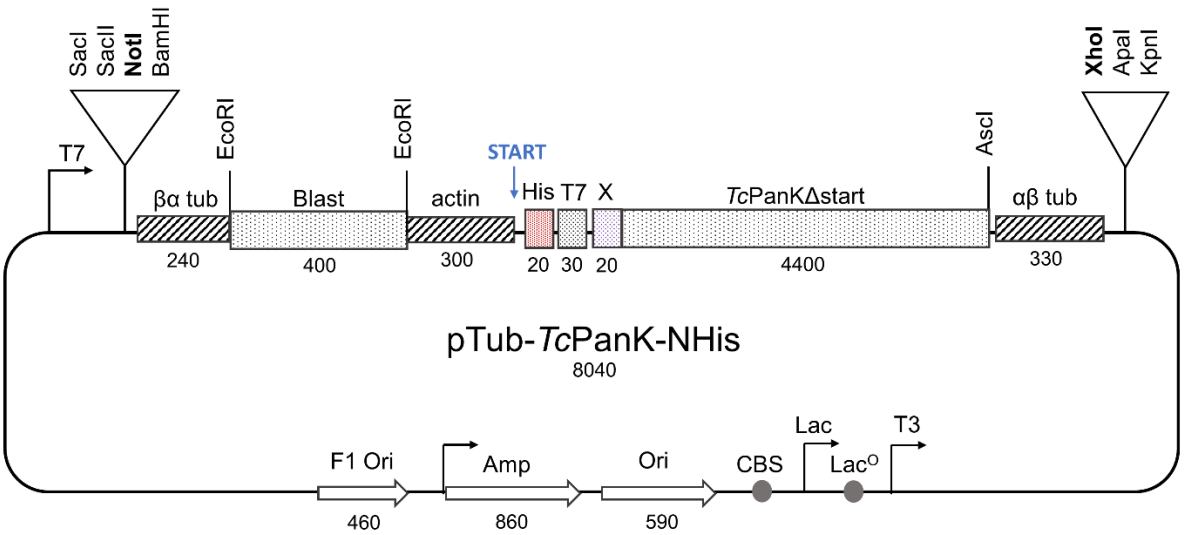
SF6 pTubEX map: For ectopic gene expression by integration in the tubulin locus. Two intergenic β-α tubulin (βα tub, 240bp) and α-β tubulin (αβ tub, 330bp) sequences flank the target gene, guiding the replacement of one copy of the native α tubulin gene with *Tc*PanK. The blasticidin deaminase gene (Blast) allows selection of transformants and an actin intergenic region (actin) provides intergenic RNA processing signals. Additional features include a T7 promoter, Ampicillin resistance components (Origin of replication: F1 Ori, Ori; Ampicillin resistance gene/ promoter: Amp; Lac promoter/ operator: Lac/ Lac^{O}) and transcription components (CAP binding site: CBS; T3 promoter: T3). Sbf and AscI sites were used to insert add-back ORFs.



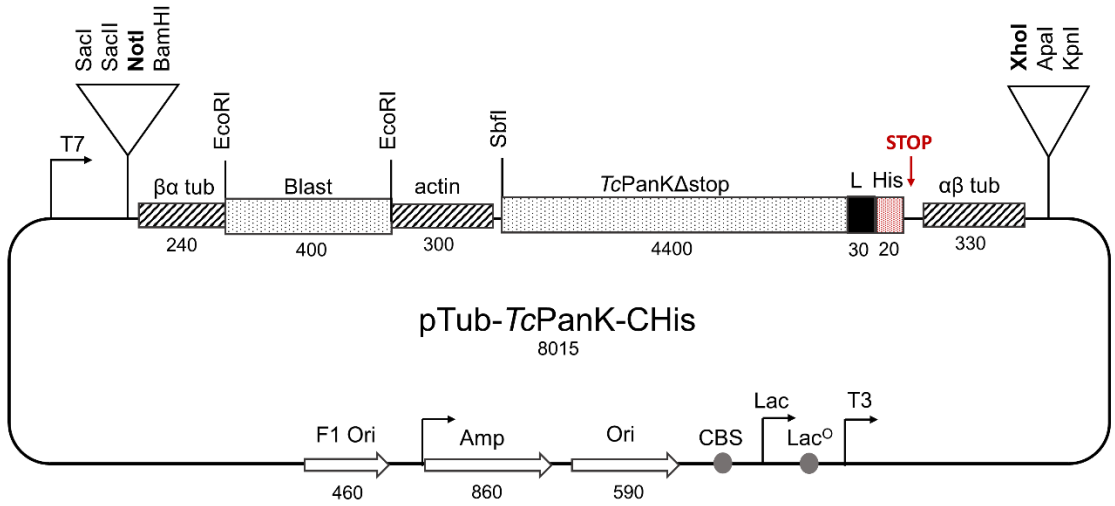
SF7 pTub-*Tc*PanK^{WT} map: For expression of WT *Tc*PanK in *T. brucei* RNAi cell lines (Materials and Methods, section 2.3.5). See SF6 for vector feature annotations.



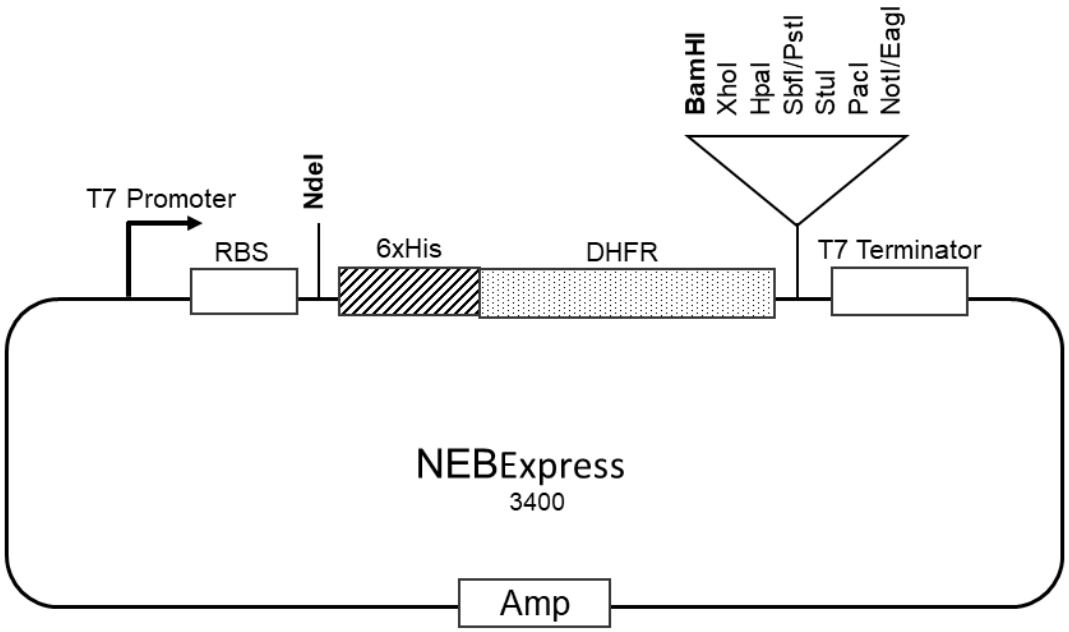
SF8 pTub-*Tc*PanK-mNG map: For expression of WT *Tc*PanK with a C-terminal mNeonGreen (mNG) fluorescence tag in *T. brucei* RNAi cell lines (Materials and Methods, section 2.3.5). L: Glycine-Serine linker peptide (30 bp). See SF6 for additional feature annotations.



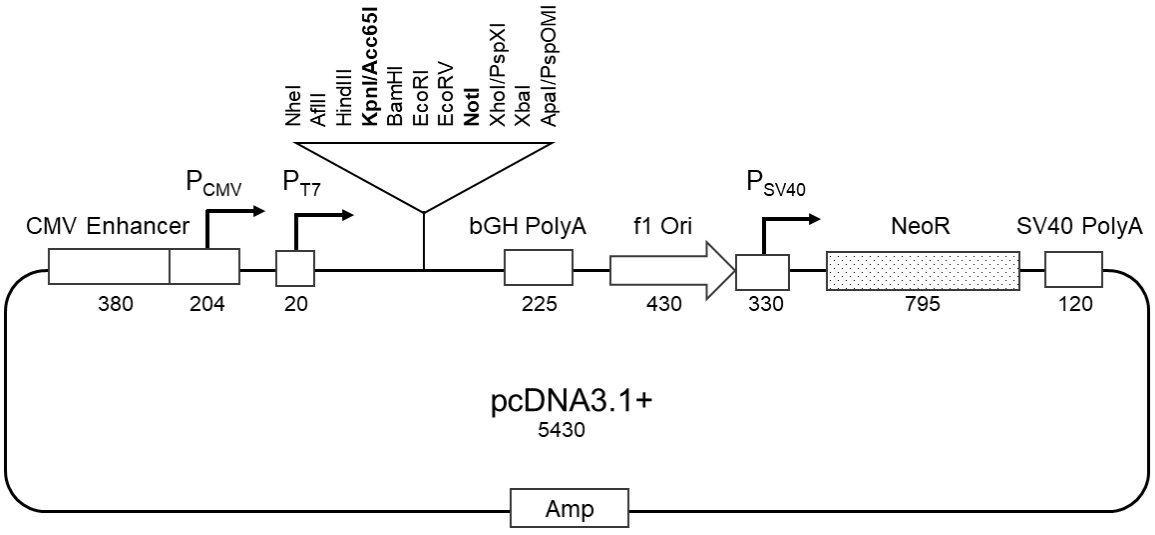
SF9 pTub-*Tc*PanK-NHis map: For expression of WT *Tc*PanK with an N-terminal hexa-histidine tag (His) in *T. brucei* RNAi cell lines (Materials and Methods, section 2.3.5). T7: leader peptide from bacteriophage T7 gene 10; X: Xpress epitope tag. Arrow indicates transcription start site. See SF6 for additional feature annotations.



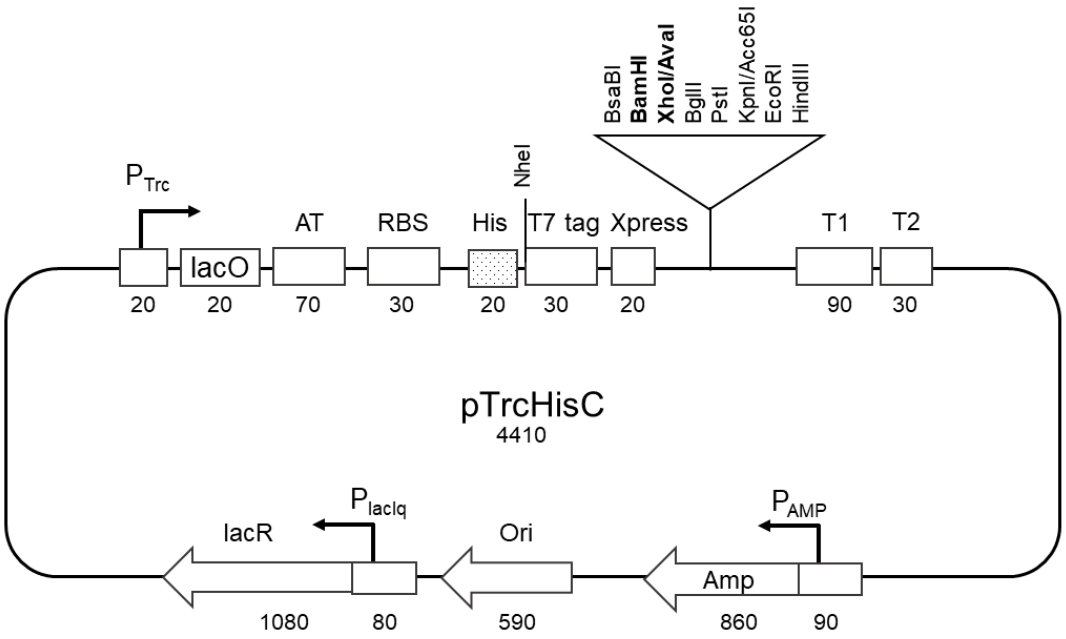
SF10 pTub-*Tc*PanK-CHis map: For expression of WT *Tc*PanK with a C-terminal hexa-histidine tag (His) in *T. brucei* RNAi cell lines (Materials and Methods, section 2.3.5). L: Glycine-Serine linker peptide. Arrow indicated stop codon. Refer to SF6 for additional feature annotations.



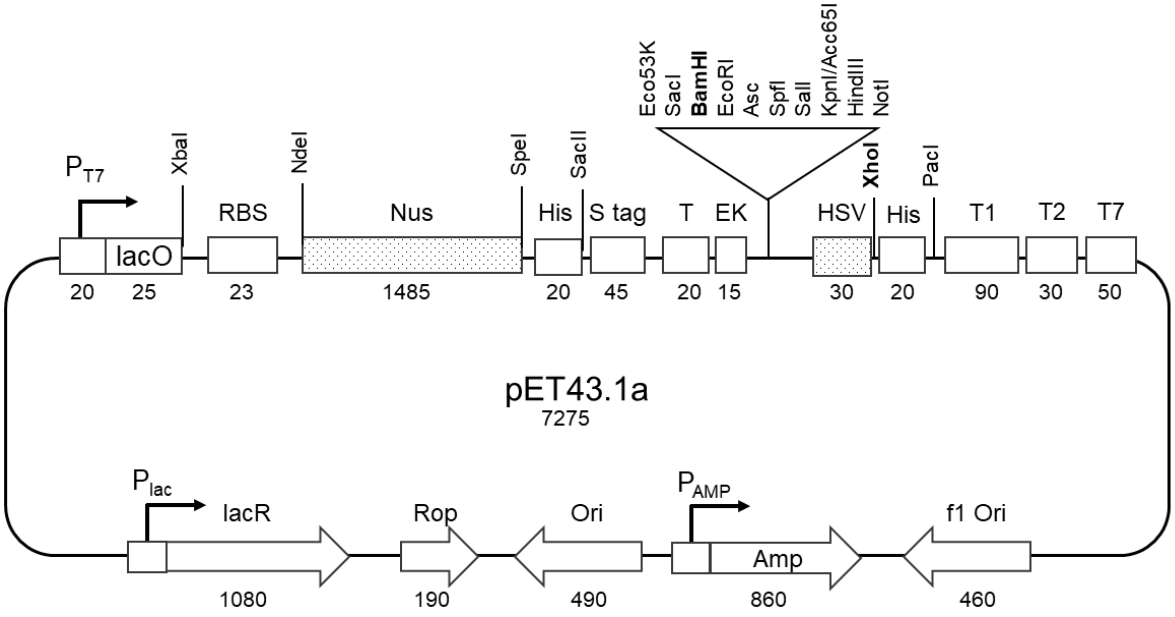
SF11 NEBExpress plasmid map: For cell-free expression using the NEBExpress Kit. RBS: Ribosomal binding site, 6xHis:Hexa-histidine tag, DHFR: Dihydrofolate reductase stuffer gene.



SF12 pcDNA3.1+ plasmid map: For protein expression in mammalian cells. CMV enhancer: human cytomegalovirus (CMV) immediate early enhancer, P: promoter, bGH PolyA: bovine growth hormone polyadenylation signal, f1 Ori: Origin of replication from f1 bacteriophage, NeoR: aminoglycoside phosphotransferase for resistance to neomycin/kanamycin.

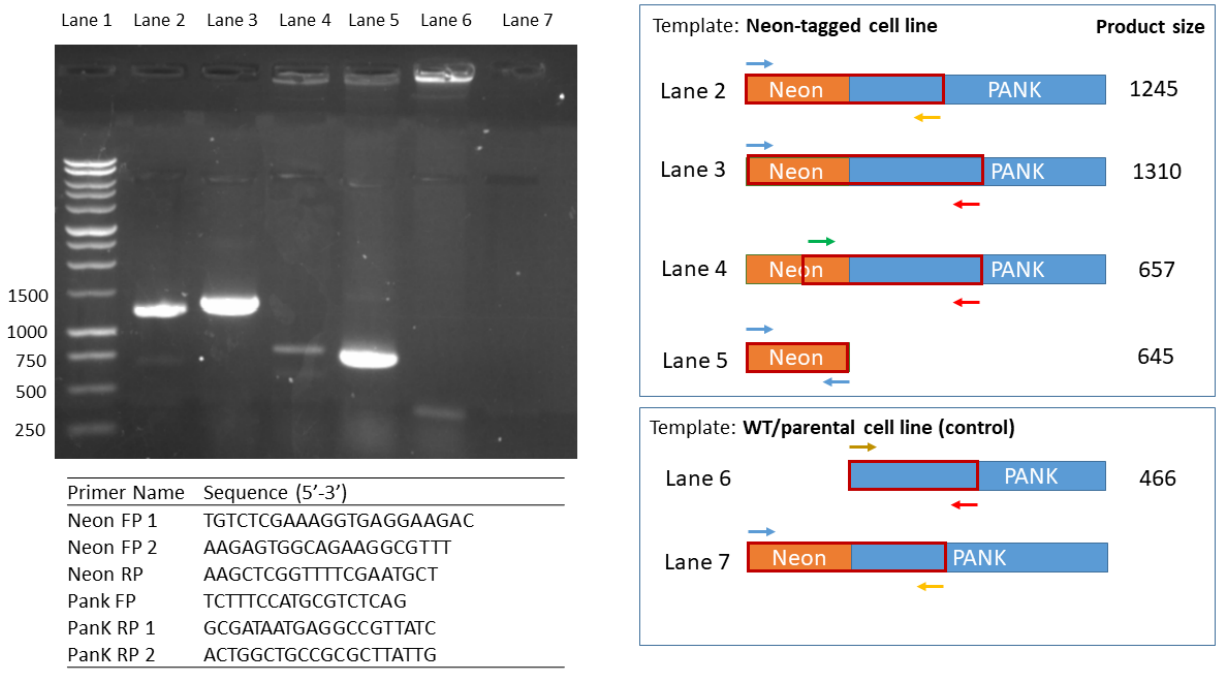


SF13 pTrcHis plasmid map: For bacterial expression. P: promoter, lacO: lac operator, AT: Anti-terminator from the *E. coli* rrnG leader region, RBS: ribosomal binding site, His: Hexa-histidine tag, T7 tag: leader peptide from bacteriophage T7 gene, Xpress: Xpress™ epitope tag containing an enterokinase cleavage site, T1/2: transcription terminators from the E. coli rrnB gene, Amp: β lactamase gene, Ori: high-copy-number ColE1/pMB1/pBR322/pUC origin of replication, lacR: lac repressor gene.



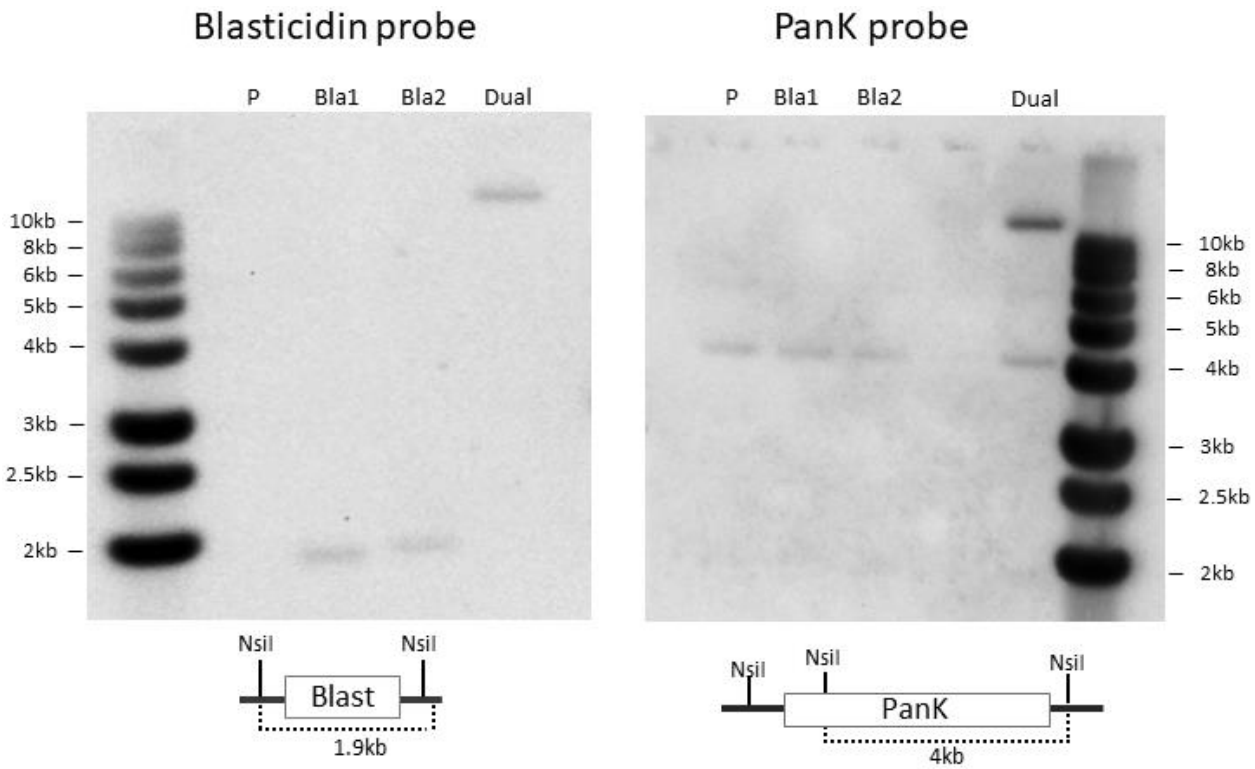
SF14 pET43.1a plasmid map: For bacterial expression. P: promoter, lacO: lac operator, RBS: ribosomal binding site, Nus: N utilization substance protein A for transcription elongation, His: Hexa-histidine tag, S tag: affinity tag derived from pancreatic ribonuclease A, T: thrombin cleavage site, EK: enterokinase cleavage site, HSV: herpes simplex virus epitope tag, T1/2: transcription terminators from the *E. coli rrnB* gene, T7: transcription terminator for bacteriophage T7 RNA polymerase, f1 Ori: Origin of replication from f1 bacteriophage, Amp: β lactamase gene, Ori: high-copy-number ColE1/pMB1/pBR322/pUC origin of replication, Rop: Rop protein, which maintains plasmids at low copy number, lacR: lac repressor gene.

C. Supplementary gel and blot images



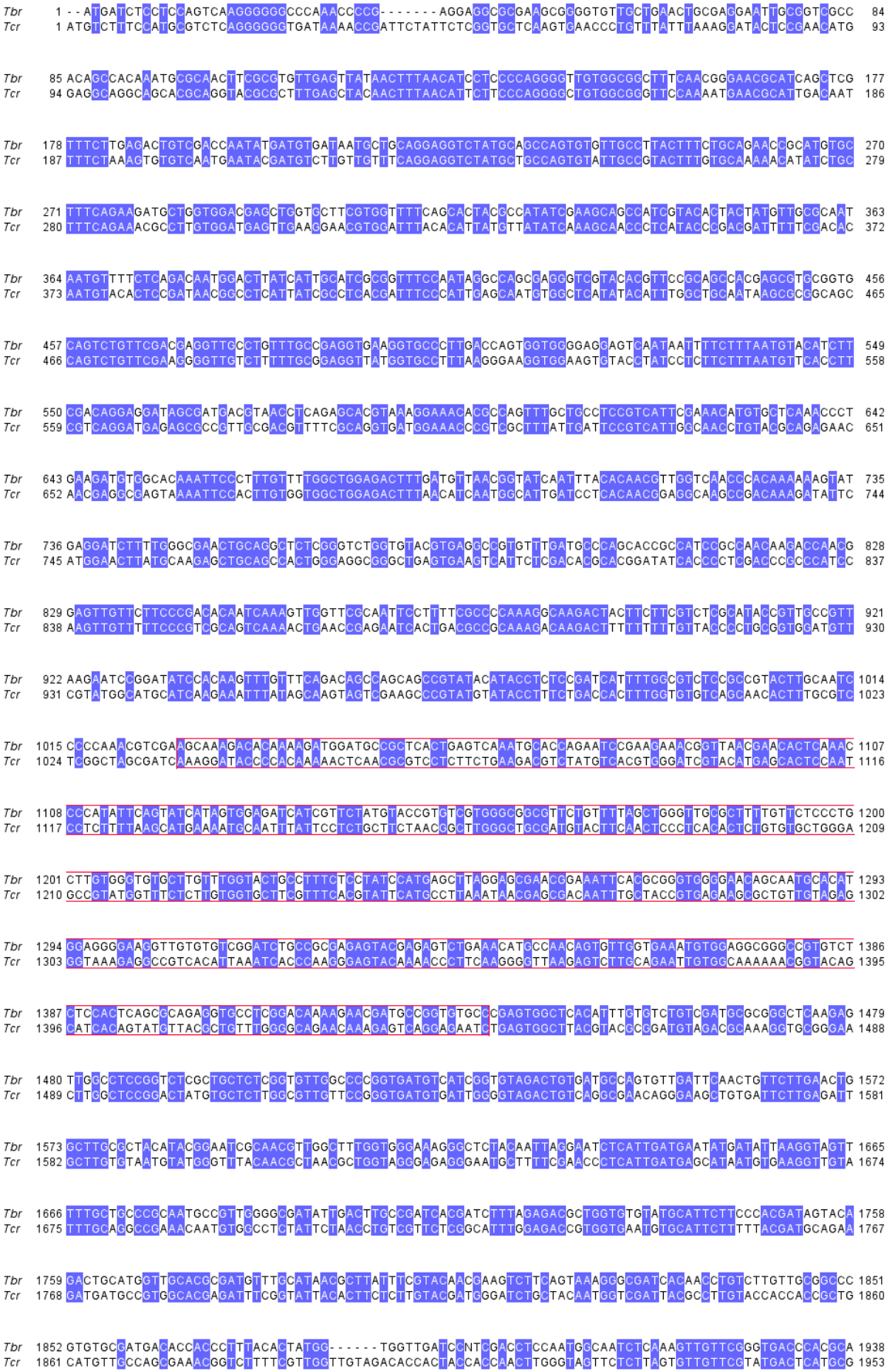
Primer Name	Sequence (5'-3')
Neon FP 1	TGTCTCGAAAGGTGAGGAAGAC
Neon FP 2	AAGAGTGGCAGAAGGCGTTT
Neon RP	AAGCTCGGTTTTCGAATGCT
Pank FP	TCTTTCCATGCGTCTCAG
PanK RP 1	GCGATAATGAGGCCGTTATC
PanK RP 2	ACTGGCTGCCGCGCTTATTG

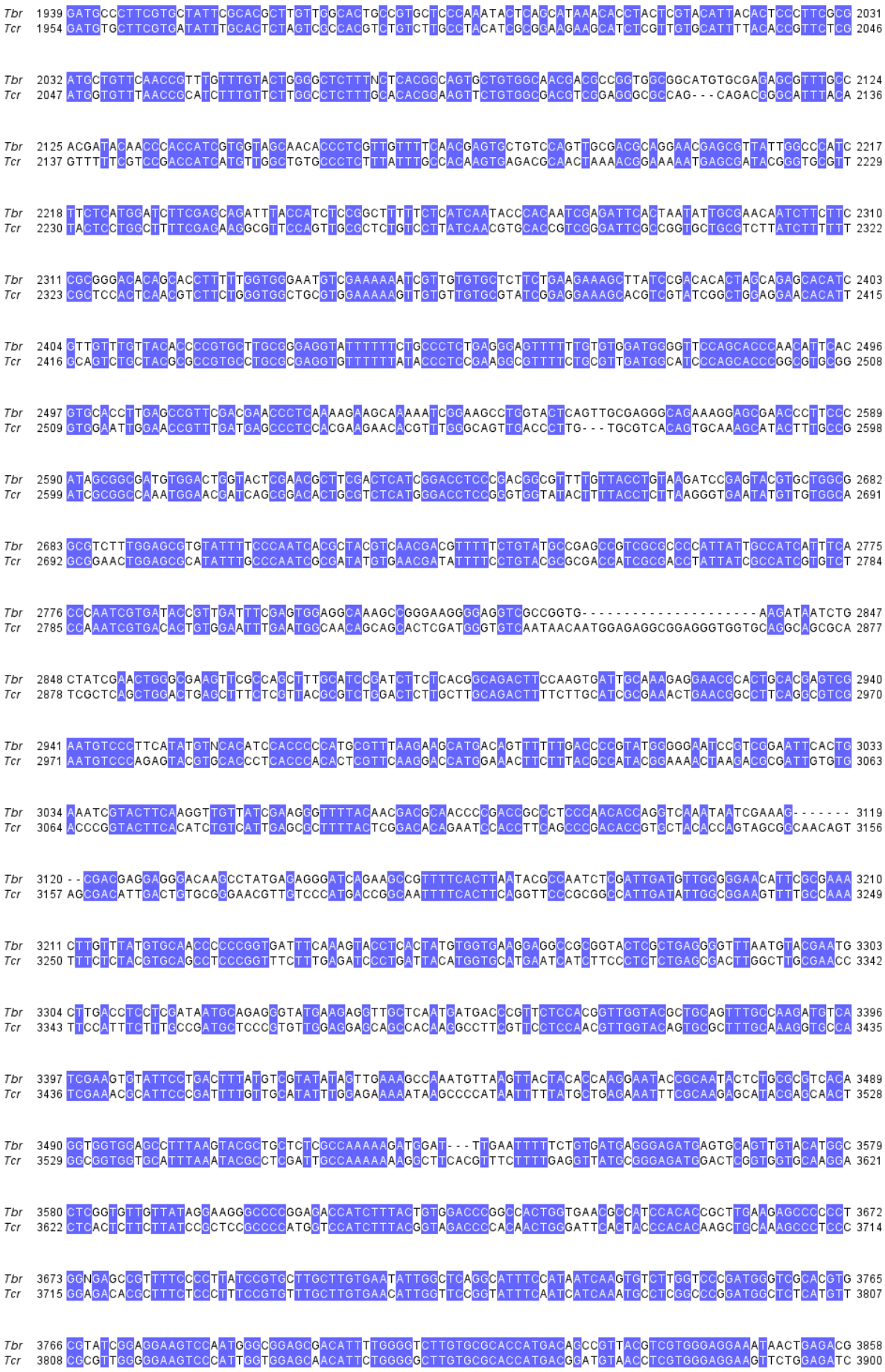
C.1. PCR across the Neon-*Tc*PanK junction: Original gel image for Fig. 4.2.2A showing all PCR products generated using the indicated primer pairs targeting Neon or *Tc*PanK ORFs. Lane 1: 1kbp DNA ladder; Lanes 2-5: Amplification products using template gDNA of Neon-tagged cell lines and the following primers: Lane 2: Neon FP1 and PanK RP1; Lane 3: Neon FP1 and PanK RP2; Lane 4: Neon FP2 and PanK RP2; Lane 5: Neon FP1 and Neon RP; Lanes 6-7: Products using template gDNA of parental (un-tagged) cell lines: Lane 6: PanK FP1 and PanK RP2; Lane 7: Neon FP1 and PanK RP1. Maps indicate primer annealing sites and expected product sizes for each reaction. Ladder sizes and expected product sizes are indicated in bp.

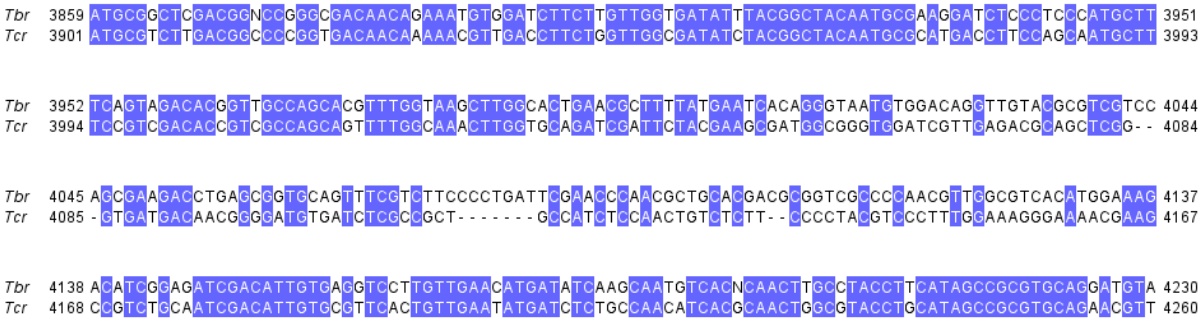


C.2. Southern blotting of *Tc*PanK replacement cell lines: Origin Southern blots for Fig. 4.3.2. Southern Produced by NsiI-digestion and Bla- or PanK-probing of gDNA from Bla- or dual drug-resistance gene-replacement (Bla1 and 2, Dual) and parental (P) cell lines. Expected product sizes are indicated on the restriction maps. Bla: blasticidin deaminase; Puro: puromycin N-acetyltransferase.

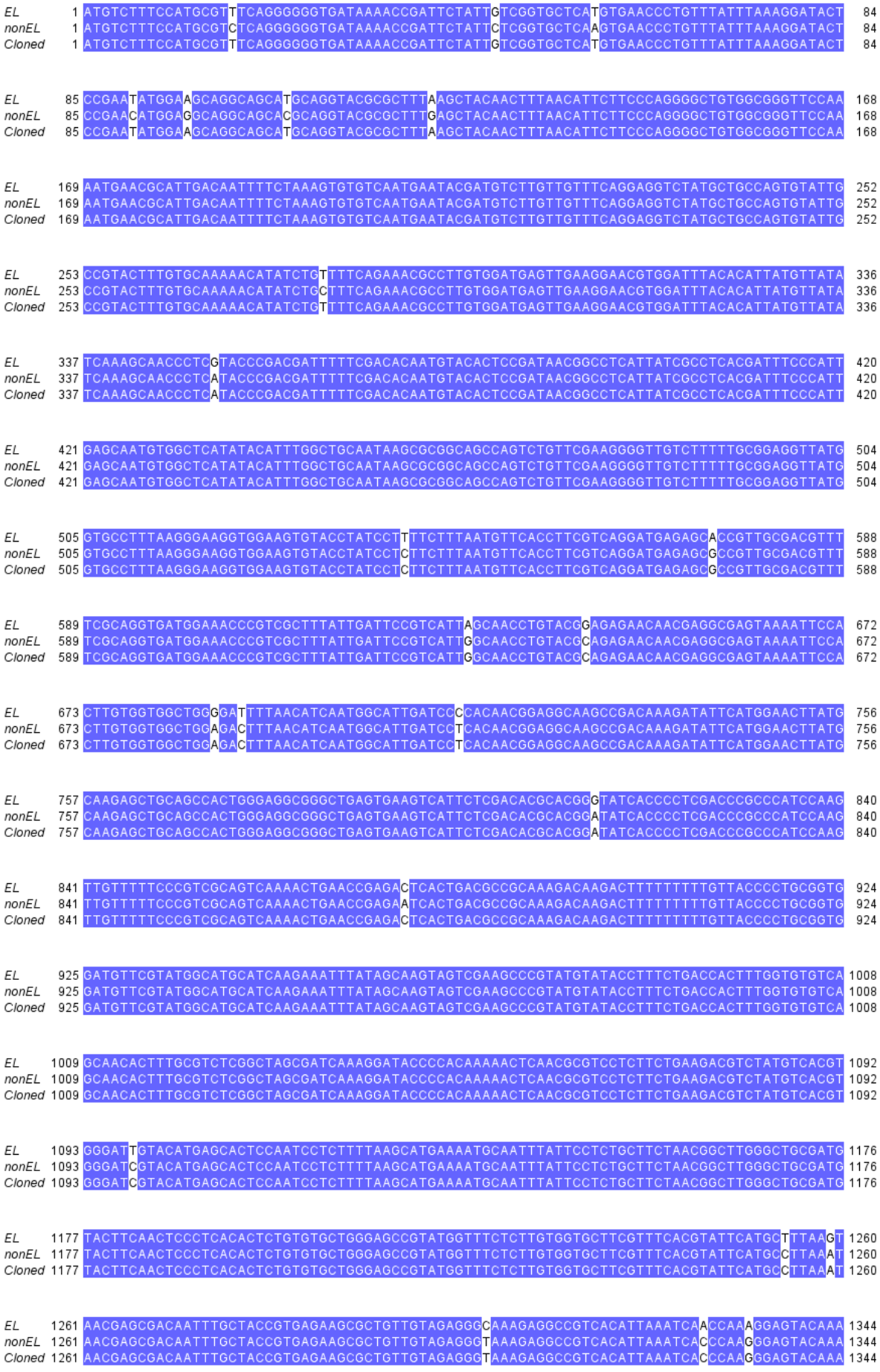
D. Supplementary sequence alignments

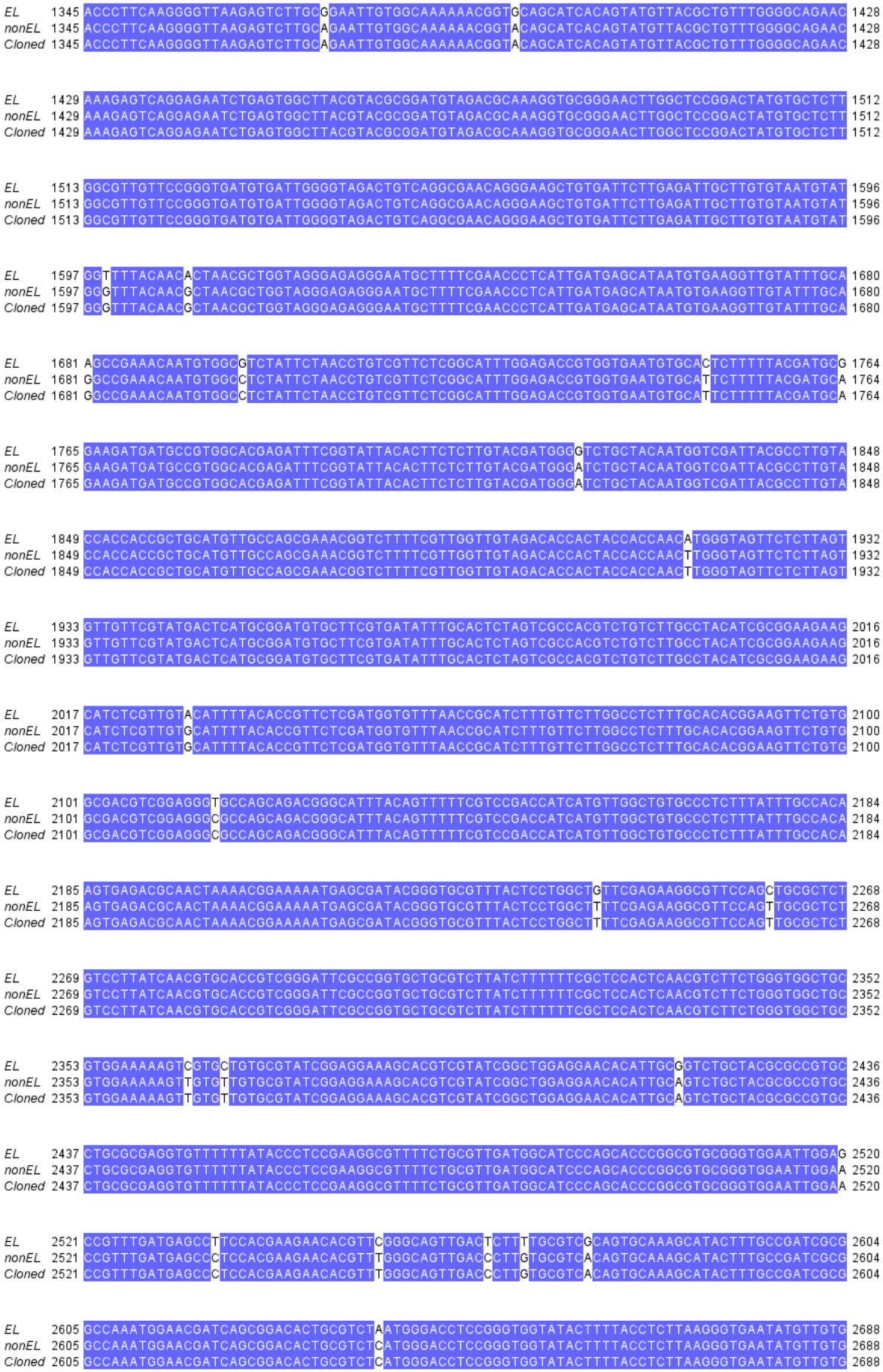


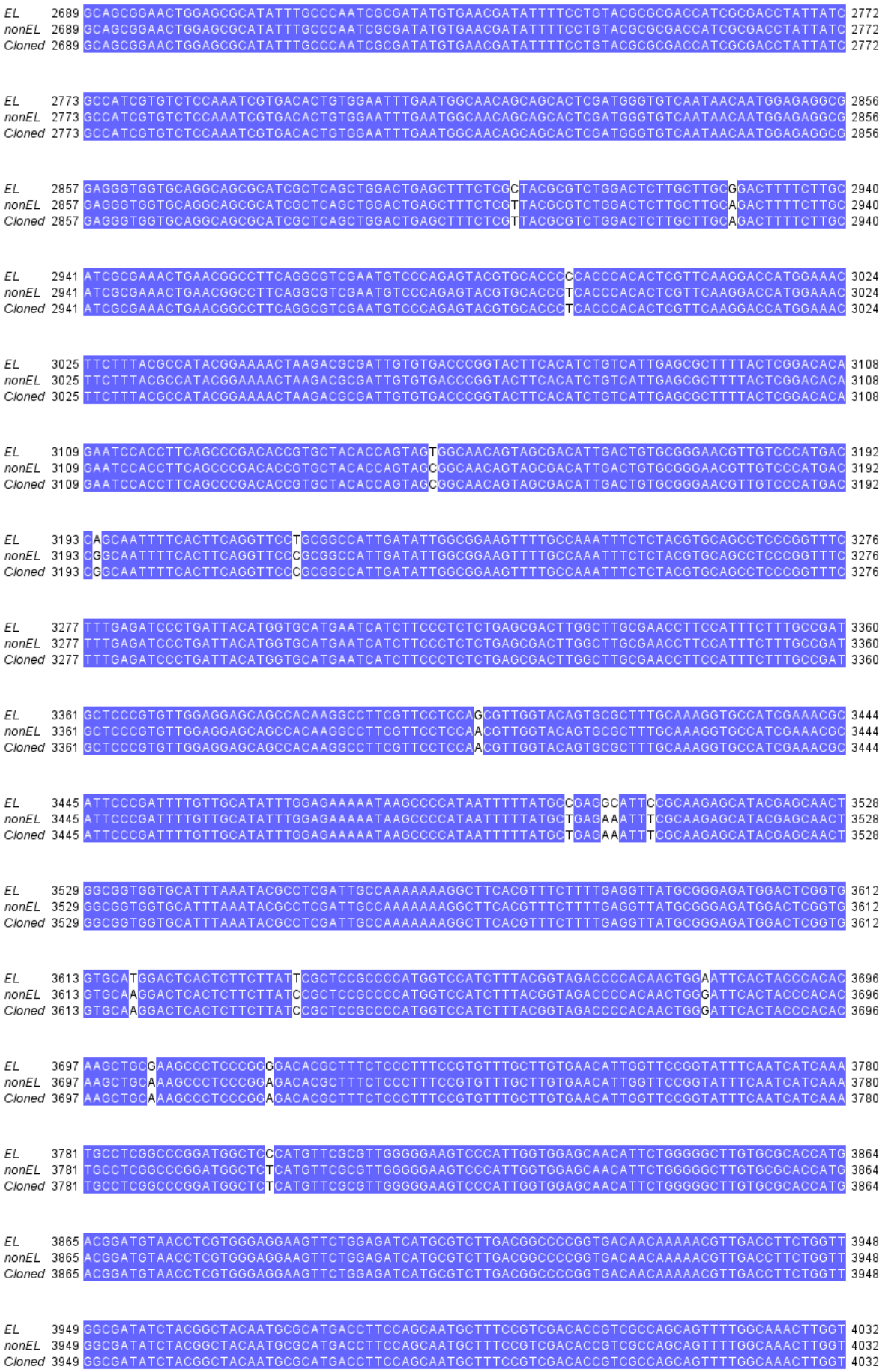


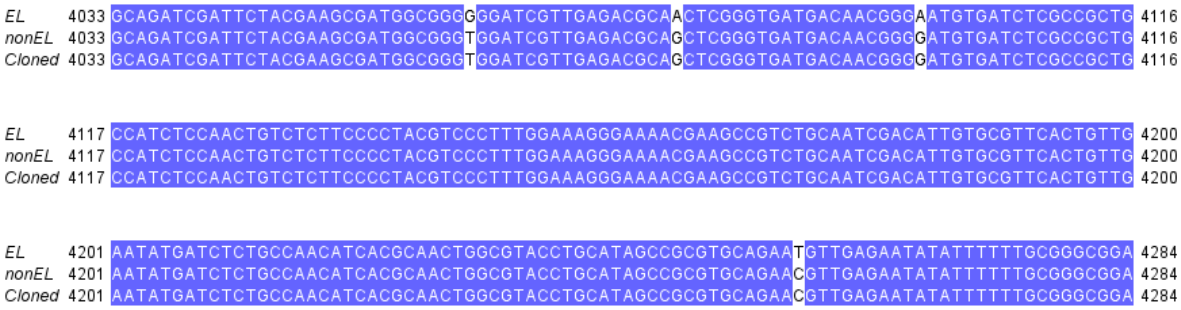


D.1. Alignment of PanK nucleotide sequences from *T. brucei* and *T. cruzi*: Aligned PanKs from *T. brucei* and *T. cruzi*, showing absence of conserved stretches >20 bp within the RNAi fragment region (outlined in red)


EL 1 ATGTCTTTCCATGCGTTTCAGGGGGGTGATAAAACCGATTCTATTGTCGGTGCTCATGTGAACCCTGTTTATTTAAAGGATACT 84
nonEL 1 ATGTCTTTCCATGCGTCTCAGGGGGGTGATAAAACCGATTCTATTCTCGGTGCTCAAGTGAACCCTGTTTATTTAAAGGATACT 84
Cloned 1 ATGTCTTTCCATGCGTTTCAGGGGGGTGATAAAACCGATTCTATTGTCGGTGCTCATGTGAACCCTGTTTATTTAAAGGATACT 84
EL 85 CCGAATATGGAAGCAGGCAGCATGCAGGTACGCGCTTTAAGCTACAACTTTAACATTCTTCCCAGGGGCTGTGGCGGGTTCCAA 168
nonEL 85 CCGAACATGGAGGCAGGCAGCACGCAGGTACGCGCTTTGAGCTACAACTTTAACATTCTTCCCAGGGGCTGTGGCGGGTTCCAA 168
Cloned 85 CCGAATATGGAAGCAGGCAGCATGCAGGTACGCGCTTTAAGCTACAACTTTAACATTCTTCCCAGGGGCTGTGGCGGGTTCCAA 168
EL 169 AATGAACGCATTGACAATTTTCTAAAGTGTGTCAATGAATACGATGTCTTGTTGTTTCAGGAGGTCTATGCTGCCAGTGTATTG 252
nonEL 169 AATGAACGCATTGACAATTTTCTAAAGTGTGTCAATGAATACGATGTCTTGTTGTTTCAGGAGGTCTATGCTGCCAGTGTATTG 252
Cloned 169 AATGAACGCATTGACAATTTTCTAAAGTGTGTCAATGAATACGATGTCTTGTTGTTTCAGGAGGTCTATGCTGCCAGTGTATTG 252
EL 253 CCGTACTTTGTGCAAAAACATATCTGTTTTCAGAAACGCCTTGTGGATGAGTTGAAGGAACGTGGATTTACACATTATGTTATA 336
nonEL 253 CCGTACTTTGTGCAAAAACATATCTGCTTTCAGAAACGCCTTGTGGATGAGTTGAAGGAACGTGGATTTACACATTATGTTATA 336
Cloned 253 CCGTACTTTGTGCAAAAACATATCTGTTTTCAGAAACGCCTTGTGGATGAGTTGAAGGAACGTGGATTTACACATTATGTTATA 336
EL 337 TCAAAGCAACCCTCGTACCCGACGATTTTTCGACACAATGTACACTCCGATAACGGCCTCATTATCGCCTCACGATTTCCCATT 420
nonEL 337 TCAAAGCAACCCTCATACCCGACGATTTTTCGACACAATGTACACTCCGATAACGGCCTCATTATCGCCTCACGATTTCCCATT 420
Cloned 337 TCAAAGCAACCCTCATACCCGACGATTTTTCGACACAATGTACACTCCGATAACGGCCTCATTATCGCCTCACGATTTCCCATT 420
EL 421 GAGCAATGTGGCTCATATACATTTGGCTGCAATAAGCGCGGCAGCCAGTCTGTTCGAAGGGGTTGTCTTTTTGCGGAGGTTATG 504
nonEL 421 GAGCAATGTGGCTCATATACATTTGGCTGCAATAAGCGCGGCAGCCAGTCTGTTCGAAGGGGTTGTCTTTTTGCGGAGGTTATG 504
Cloned 421 GAGCAATGTGGCTCATATACATTTGGCTGCAATAAGCGCGGCAGCCAGTCTGTTCGAAGGGGTTGTCTTTTTGCGGAGGTTATG 504
EL 505 GTGCCTTTAAGGGAAGGTGGAAGTGTACCTATCCTTTTCTTTAATGTTCACCTTCGTCAGGATGAGAGCACCGTTGCGACGTTT 588
nonEL 505 GTGCCTTTAAGGGAAGGTGGAAGTGTACCTATCCTCTTCTTTAATGTTCACCTTCGTCAGGATGAGAGGCCGTTGCGACGTTT 588
Cloned 505 GTGCCTTTAAGGGAAGGTGGAAGTGTACCTATCCTCTTCTTTAATGTTCACCTTCGTCAGGATGAGAGGCCGTTGCGACGTTT 588
EL 589 TCGCAGGTGATGGAAACCCGTCGCTTTATTGATTCCGTCATTAGCAACCTGTACGGAGAGAACAACGAGGCGAGTAAAATTCCA 672
nonEL 589 TCGCAGGTGATGGAAACCCGTCGCTTTATTGATTCCGTCATTGGCAACCTGTACGCAGAGAACAACGAGGCGAGTAAAATTCCA 672
Cloned 589 TCGCAGGTGATGGAAACCCGTCGCTTTATTGATTCCGTCATTGGCAACCTGTACGCAGAGAACAACGAGGCGAGTAAAATTCCA 672
EL 673 CTTGTGGTGGCTGGGGATTTTAACATCAATGGCATTGATCCCCACAACGGAGGCAAGCCGACAAAGATATTCATGGAACTTATG 756
nonEL 673 CTTGTGGTGGCTGGAGACTTTAACATCAATGGCATTGATCCTCACAACGGAGGCAAGCCGACAAAGATATTCATGGAACTTATG 756
Cloned 673 CTTGTGGTGGCTGGAGACTTTAACATCAATGGCATTGATCCTCACAACGGAGGCAAGCCGACAAAGATATTCATGGAACTTATG 756
EL 757 CAAGAGCTGCAGCCACTGGGAGGCGGGCTGAGTGAAGTCATTCTCGACACGCACGGGTATCACCCCTCGACCCGCCCATCCAAG 840
nonEL 757 CAAGAGCTGCAGCCACTGGGAGGCGGGCTGAGTGAAGTCATTCTCGACACGCACGGATATCACCCCTCGACCCGCCCATCCAAG 840
Cloned 757 CAAGAGCTGCAGCCACTGGGAGGCGGGCTGAGTGAAGTCATTCTCGACACGCACGGATATCACCCCTCGACCCGCCCATCCAAG 840
EL 841 TTGTTTTTCCCGTCGCAGTCAAAACTGAACCGAGACTCACTGACGCCGCAAAGACAAGACTTTTTTTTTGTTACCCCTGCGGTG 924
nonEL 841 TTGTTTTTCCCGTCGCAGTCAAAACTGAACCGAGAATCACTGACGCCGCAAAGACAAGACTTTTTTTTTGTTACCCCTGCGGTG 924
Cloned 841 TTGTTTTTCCCGTCGCAGTCAAAACTGAACCGAGACTCACTGACGCCGCAAAGACAAGACTTTTTTTTTGTTACCCCTGCGGTG 924
EL 925 GATGTTCGTATGGCATGCATCAAGAAATTTATAGCAAGTAGTCGAAGCCCGTATGTATACCTTTCTGACCACTTTGGTGTGTCA 1008
nonEL 925 GATGTTCGTATGGCATGCATCAAGAAATTTATAGCAAGTAGTCGAAGCCCGTATGTATACCTTTCTGACCACTTTGGTGTGTCA 1008
Cloned 925 GATGTTCGTATGGCATGCATCAAGAAATTTATAGCAAGTAGTCGAAGCCCGTATGTATACCTTTCTGACCACTTTGGTGTGTCA 1008
EL 1009 GCAACACTTTGCGTCTCGGCTAGCGATCAAAGGATACCCCACAAAAACTCAACGCGTCCTCTTCTGAAGACGTCTATGTCACGT 1092
nonEL 1009 GCAACACTTTGCGTCTCGGCTAGCGATCAAAGGATACCCCACAAAAACTCAACGCGTCCTCTTCTGAAGACGTCTATGTCACGT 1092
Cloned 1009 GCAACACTTTGCGTCTCGGCTAGCGATCAAAGGATACCCCACAAAAACTCAACGCGTCCTCTTCTGAAGACGTCTATGTCACGT 1092
EL 1093 GGGATTGTACATGAGCACTCCAATCCTCTTTTAAGCATGAAAATGCAATTTATTCCTCTGCTTCTAACGGCTTGGGCTGCGATG 1176
nonEL 1093 GGGATCGTACATGAGCACTCCAATCCTCTTTTAAGCATGAAAATGCAATTTATTCCTCTGCTTCTAACGGCTTGGGCTGCGATG 1176
Cloned 1093 GGGATCGTACATGAGCACTCCAATCCTCTTTTAAGCATGAAAATGCAATTTATTCCTCTGCTTCTAACGGCTTGGGCTGCGATG 1176
EL 1177 TACTTCAACTCCCTCACACTCTGTGTGCTGGGAGCCGTATGGTTTCTCTTGTGGTGCTTCGTTTCACGTATTCATGCTTTAAGT 1260
nonEL 1177 TACTTCAACTCCCTCACACTCTGTGTGCTGGGAGCCGTATGGTTTCTCTTGTGGTGCTTCGTTTCACGTATTCATGCCTTAAAT 1260
Cloned 1177 TACTTCAACTCCCTCACACTCTGTGTGCTGGGAGCCGTATGGTTTCTCTTGTGGTGCTTCGTTTCACGTATTCATGCCTTAAAT 1260
EL 1261 AACGAGCGACAATTTGCTACCGTGAGAAGCGCTGTTGTAGAGGGCAAAGAGGCCGTCACATTAAATCAACGAAGGAGTACAAA 1344
nonEL 1261 AACGAGCGACAATTTGCTACCGTGAGAAGCGCTGTTGTAGAGGGTAAAGAGGCCGTCACATTAAATCACCGAAGGAGTACAAA 1344
Cloned 1261 AACGAGCGACAATTTGCTACCGTGAGAAGCGCTGTTGTAGAGGGTAAAGAGGCCGTCACATTAAATCACCGAAGGAGTACAAA 1344


EL 1345 ACCCTTCAAGGGGTTAAGAGTCTTGCGGAATTGTGGCAAAAAACGGTGCAGCATCACAGTATGTTACGCTGTTTGGGGCAGAAC 1428
nonEL 1345 ACCCTTCAAGGGGTTAAGAGTCTTGCAGAATTGTGGCAAAAAACGGTACAGCATCACAGTATGTTACGCTGTTTGGGGCAGAAC 1428
Cloned 1345 ACCCTTCAAGGGGTTAAGAGTCTTGCAGAATTGTGGCAAAAAACGGTACAGCATCACAGTATGTTACGCTGTTTGGGGCAGAAC 1428
EL 1429 AAAGAGTCAGGAGAATCTGAGTGGCTTACGTACGCGGATGTAGACGCAAAGGTGCGGGAACTTGGCTCCGGACTATGTGCTCTT 1512
nonEL 1429 AAAGAGTCAGGAGAATCTGAGTGGCTTACGTACGCGGATGTAGACGCAAAGGTGCGGGAACTTGGCTCCGGACTATGTGCTCTT 1512
Cloned 1429 AAAGAGTCAGGAGAATCTGAGTGGCTTACGTACGCGGATGTAGACGCAAAGGTGCGGGAACTTGGCTCCGGACTATGTGCTCTT 1512
EL 1513 GGCGTTGTTCCGGGTGATGTGATTGGGGTAGACTGTCAGGCGAACAGGGAAGCTGTGATTCTTGAGATTGCTTGTGTAATGTAT 1596
nonEL 1513 GGCGTTGTTCCGGGTGATGTGATTGGGGTAGACTGTCAGGCGAACAGGGAAGCTGTGATTCTTGAGATTGCTTGTGTAATGTAT 1596
Cloned 1513 GGCGTTGTTCCGGGTGATGTGATTGGGGTAGACTGTCAGGCGAACAGGGAAGCTGTGATTCTTGAGATTGCTTGTGTAATGTAT 1596
EL 1597 GGTTTTACAACACTAACGCTGGTAGGGAGAGGGAATGCTTTTCGAACCCTCATTGATGAGCATAATGTGAAGGTTGTATTTGCA 1680
nonEL 1597 GGGTTTACAACGCTAACGCTGGTAGGGAGAGGGAATGCTTTTCGAACCCTCATTGATGAGCATAATGTGAAGGTTGTATTTGCA 1680
Cloned 1597 GGGTTTACAACGCTAACGCTGGTAGGGAGAGGGAATGCTTTTCGAACCCTCATTGATGAGCATAATGTGAAGGTTGTATTTGCA 1680
EL 1681 AGCCGAAACAATGTGGCGTCTATTCTAACCTGTCGTTCTCGGCATTTGGAGACCGTGGTGAATGTGCACTCTTTTTACGATGCG 1764
nonEL 1681 GGCCGAAACAATGTGGCCTCTATTCTAACCTGTCGTTCTCGGCATTTGGAGACCGTGGTGAATGTGCATTCTTTTTACGATGCA 1764
Cloned 1681 GGCCGAAACAATGTGGCCTCTATTCTAACCTGTCGTTCTCGGCATTTGGAGACCGTGGTGAATGTGCATTCTTTTTACGATGCA 1764
EL 1765 GAAGATGATGCCGTGGCACGAGATTTCGGTATTACACTTCTCTTGTACGATGGGGTCTGCTACAATGGTCGATTACGCCTTGTA 1848
nonEL 1765 GAAGATGATGCCGTGGCACGAGATTTCGGTATTACACTTCTCTTGTACGATGGGATCTGCTACAATGGTCGATTACGCCTTGTA 1848
Cloned 1765 GAAGATGATGCCGTGGCACGAGATTTCGGTATTACACTTCTCTTGTACGATGGGATCTGCTACAATGGTCGATTACGCCTTGTA 1848
EL 1849 CCACCACCGCTGCATGTTGCCAGCGAAACGGTCTTTTCGTTGGTTGTAGACACCACTACCACCAACATGGGTAGTTCTCTTAGT 1932
nonEL 1849 CCACCACCGCTGCATGTTGCCAGCGAAACGGTCTTTTCGTTGGTTGTAGACACCACTACCACCAACTTGGGTAGTTCTCTTAGT 1932
Cloned 1849 CCACCACCGCTGCATGTTGCCAGCGAAACGGTCTTTTCGTTGGTTGTAGACACCACTACCACCAACTTGGGTAGTTCTCTTAGT 1932
EL 1933 GTTGTTCGTATGACTCATGCGGATGTGCTTCGTGATATTTGCACTCTAGTCGCCACGTCTGTCTTGCCTACATCGCGGAAGAAG 2016
nonEL 1933 GTTGTTCGTATGACTCATGCGGATGTGCTTCGTGATATTTGCACTCTAGTCGCCACGTCTGTCTTGCCTACATCGCGGAAGAAG 2016
Cloned 1933 GTTGTTCGTATGACTCATGCGGATGTGCTTCGTGATATTTGCACTCTAGTCGCCACGTCTGTCTTGCCTACATCGCGGAAGAAG 2016
EL 2017 CATCTCGTTGTACATTTTACACCGTTCTCGATGGTGTTTAACCGCATCTTTGTTCTTGGCCTCTTTGCACACGGAAGTTCTGTG 2100
nonEL 2017 CATCTCGTTGTGCATTTTACACCGTTCTCGATGGTGTTTAACCGCATCTTTGTTCTTGGCCTCTTTGCACACGGAAGTTCTGTG 2100
Cloned 2017 CATCTCGTTGTGCATTTTACACCGTTCTCGATGGTGTTTAACCGCATCTTTGTTCTTGGCCTCTTTGCACACGGAAGTTCTGTG 2100
EL 2101 GCGACGTCGGAGGCTGCCAGCAGACGGGCATTTACAGTTTTTCGTCCGACCATCATGTTGGCTGTGCCCTCTTTATTTGCCACA 2184
nonEL 2101 GCGACGTCGGAGGCCGCCAGCAGACGGGCATTTACAGTTTTTCGTCCGACCATCATGTTGGCTGTGCCCTCTTTATTTGCCACA 2184
Cloned 2101 GCGACGTCGGAGGCCGCCAGCAGACGGGCATTTACAGTTTTTCGTCCGACCATCATGTTGGCTGTGCCCTCTTTATTTGCCACA 2184
EL 2185 AGTGAGACGCAACTAAAACGGAAAAATGAGCGATACGGGTGCGTTTACTCCTGGCTGTTCGAGAAGGCGTTCCAGCTGCGCTCT 2268
nonEL 2185 AGTGAGACGCAACTAAAACGGAAAAATGAGCGATACGGGTGCGTTTACTCCTGGCTTTTCGAGAAGGCGTTCCAGTTGCGCTCT 2268
Cloned 2185 AGTGAGACGCAACTAAAACGGAAAAATGAGCGATACGGGTGCGTTTACTCCTGGCTTTTCGAGAAGGCGTTCCAGTTGCGCTCT 2268
EL 2269 GTCCTTATCAACGTGCACCGTCGGGATTCGCCGGTGCTGCGTCTTATCTTTTTTCGCTCCACTCAACGTCTTCTGGGTGGCTGC 2352
nonEL 2269 GTCCTTATCAACGTGCACCGTCGGGATTCGCCGGTGCTGCGTCTTATCTTTTTTCGCTCCACTCAACGTCTTCTGGGTGGCTGC 2352
Cloned 2269 GTCCTTATCAACGTGCACCGTCGGGATTCGCCGGTGCTGCGTCTTATCTTTTTTCGCTCCACTCAACGTCTTCTGGGTGGCTGC 2352
EL 2353 GTGGAAAAAGTCGTGCTGTGCGTATCGGAGGAAAGCACGTCGTATCGGCTGGAGGAACACATTGGGGTCTGCTACGCGCCGTGC 2436
nonEL 2353 GTGGAAAAAGTTGTGTTGTGCGTATCGGAGGAAAGCACGTCGTATCGGCTGGAGGAACACATTGCAGTCTGCTACGCGCCGTGC 2436
Cloned 2353 GTGGAAAAAGTTGTGTTGTGCGTATCGGAGGAAAGCACGTCGTATCGGCTGGAGGAACACATTGCAGTCTGCTACGCGCCGTGC 2436
EL 2437 CTGCGCGAGGTGTTTTTTATACCCTCCGAAGGCGTTTTCTGCGTTGATGGCATCCCAGCACCCGGCGTGCGGGTGGAATTGGAG 2520
nonEL 2437 CTGCGCGAGGTGTTTTTTATACCCTCCGAAGGCGTTTTCTGCGTTGATGGCATCCCAGCACCCGGCGTGCGGGTGGAATTGGAA 2520
Cloned 2437 CTGCGCGAGGTGTTTTTTATACCCTCCGAAGGCGTTTTCTGCGTTGATGGCATCCCAGCACCCGGCGTGCGGGTGGAATTGGAA 2520
EL 2521 CCGTTTGATGAGCCTTCCACGAAGAACACGTTCGGGCAGTTGACTCTTTTGCGTCGCAGTGCAAAGCATACTTTGCCGATCGCG 2604
nonEL 2521 CCGTTTGATGAGCCCTCCACGAAGAACACGTTTGGGCAGTTGACCCTTGTGCGTCACAGTGCAAAGCATACTTTGCCGATCGCG 2604
Cloned 2521 CCGTTTGATGAGCCCTCCACGAAGAACACGTTTGGGCAGTTGACCCTTGTGCGTCACAGTGCAAAGCATACTTTGCCGATCGCG 2604
EL 2605 GCCAAATGGAACGATCAGCGGACACTGCGTCTAATGGGACCTCCGGGTGGTATACTTTTACCTCTTAAGGGTGAATATGTTGTG 2688
nonEL 2605 GCCAAATGGAACGATCAGCGGACACTGCGTCTCATGGGACCTCCGGGTGGTATACTTTTACCTCTTAAGGGTGAATATGTTGTG 2688
Cloned 2605 GCCAAATGGAACGATCAGCGGACACTGCGTCTCATGGGACCTCCGGGTGGTATACTTTTACCTCTTAAGGGTGAATATGTTGTG 2688


EL 2689 GCAGCGGAACTGGAGCGCATATTTGCCCAATCGCGATATGTGAACGATATTTTCCTGTACGCGCGACCATCGCGACCTATTATC 2772
nonEL 2689 GCAGCGGAACTGGAGCGCATATTTGCCCAATCGCGATATGTGAACGATATTTTCCTGTACGCGCGACCATCGCGACCTATTATC 2772
Cloned 2689 GCAGCGGAACTGGAGCGCATATTTGCCCAATCGCGATATGTGAACGATATTTTCCTGTACGCGCGACCATCGCGACCTATTATC 2772
EL 2773 GCCATCGTGTCTCCAAATCGTGACACTGTGGAATTTGAATGGCAACAGCAGCACTCGATGGGTGTCAATAACAATGGAGAGGCG 2856
nonEL 2773 GCCATCGTGTCTCCAAATCGTGACACTGTGGAATTTGAATGGCAACAGCAGCACTCGATGGGTGTCAATAACAATGGAGAGGCG 2856
Cloned 2773 GCCATCGTGTCTCCAAATCGTGACACTGTGGAATTTGAATGGCAACAGCAGCACTCGATGGGTGTCAATAACAATGGAGAGGCG 2856
EL 2857 GAGGGTGGTGCAGGCAGCGCATCGCTCAGCTGGACTGAGCTTTCTCGCTACGCGTCTGGACTCTTGCTTGCGGACTTTTCTTGC 2940
nonEL 2857 GAGGGTGGTGCAGGCAGCGCATCGCTCAGCTGGACTGAGCTTTCTCGTTACGCGTCTGGACTCTTGCTTGCAGACTTTTCTTGC 2940
Cloned 2857 GAGGGTGGTGCAGGCAGCGCATCGCTCAGCTGGACTGAGCTTTCTCGTTACGCGTCTGGACTCTTGCTTGCAGACTTTTCTTGC 2940
EL 2941 ATCGCGAAACTGAACGGCCTTCAGGCGTCGAATGTCCCAGAGTACGTGCACCCCACCCACACTCGTTCAAGGACCATGGAAAC 3024
nonEL 2941 ATCGCGAAACTGAACGGCCTTCAGGCGTCGAATGTCCCAGAGTACGTGCACCCTCACCCACACTCGTTCAAGGACCATGGAAAC 3024
Cloned 2941 ATCGCGAAACTGAACGGCCTTCAGGCGTCGAATGTCCCAGAGTACGTGCACCCTCACCCACACTCGTTCAAGGACCATGGAAAC 3024
EL 3025 TTCTTTACGCCATACGGAAAACTAAGACGCGATTGTGTGACCCGGTACTTCACATCTGTCATTGAGCGCTTTTACTCGGACACA 3108
nonEL 3025 TTCTTTACGCCATACGGAAAACTAAGACGCGATTGTGTGACCCGGTACTTCACATCTGTCATTGAGCGCTTTTACTCGGACACA 3108
Cloned 3025 TTCTTTACGCCATACGGAAAACTAAGACGCGATTGTGTGACCCGGTACTTCACATCTGTCATTGAGCGCTTTTACTCGGACACA 3108
EL 3109 GAATCCACCTTCAGCCCGACACCGTGCTACACCAGTAGTGGCAACAGTAGCGACATTGACTGTGCGGGAACGTTGTCCCATGAC 3192
nonEL 3109 GAATCCACCTTCAGCCCGACACCGTGCTACACCAGTAGCGGCAACAGTAGCGACATTGACTGTGCGGGAACGTTGTCCCATGAC 3192
Cloned 3109 GAATCCACCTTCAGCCCGACACCGTGCTACACCAGTAGCGGCAACAGTAGCGACATTGACTGTGCGGGAACGTTGTCCCATGAC 3192
EL 3193 CAGCAATTTTCACTTCAGGTTCCTGCGGCCATTGATATTGGCGGAAGTTTTGCCAAATTTCTCTACGTGCAGCCTCCCGGTTTC 3276
nonEL 3193 CGGCAATTTTCACTTCAGGTTCCCGCGGCCATTGATATTGGCGGAAGTTTTGCCAAATTTCTCTACGTGCAGCCTCCCGGTTTC 3276
Cloned 3193 CGGCAATTTTCACTTCAGGTTCCCGCGGCCATTGATATTGGCGGAAGTTTTGCCAAATTTCTCTACGTGCAGCCTCCCGGTTTC 3276
EL 3277 TTTGAGATCCCTGATTACATGGTGCATGAATCATCTTCCCTCTCTGAGCGACTTGGCTTGCGAACCTTCCATTTCTTTGCCGAT 3360
nonEL 3277 TTTGAGATCCCTGATTACATGGTGCATGAATCATCTTCCCTCTCTGAGCGACTTGGCTTGCGAACCTTCCATTTCTTTGCCGAT 3360
Cloned 3277 TTTGAGATCCCTGATTACATGGTGCATGAATCATCTTCCCTCTCTGAGCGACTTGGCTTGCGAACCTTCCATTTCTTTGCCGAT 3360
EL 3361 GCTCCCGTGTTGGAGGAGCAGCCACAAGGCCTTCGTTCCTCCAGCGTTGGTACAGTGCGCTTTGCAAAGGTGCCATCGAAACGC 3444
nonEL 3361 GCTCCCGTGTTGGAGGAGCAGCCACAAGGCCTTCGTTCCTCCAACGTTGGTACAGTGCGCTTTGCAAAGGTGCCATCGAAACGC 3444
Cloned 3361 GCTCCCGTGTTGGAGGAGCAGCCACAAGGCCTTCGTTCCTCCAACGTTGGTACAGTGCGCTTTGCAAAGGTGCCATCGAAACGC 3444
EL 3445 ATTCCCGATTTTGTTGCATATTTGGAGAAAAATAAGCCCCATAATTTTTATGCCGAGGCATTCCGCAAGAGCATACGAGCAACT 3528
nonEL 3445 ATTCCCGATTTTGTTGCATATTTGGAGAAAAATAAGCCCCATAATTTTTATGCTGAGAATTTCGCAAGAGCATACGAGCAACT 3528
Cloned 3445 ATTCCCGATTTTGTTGCATATTTGGAGAAAAATAAGCCCCATAATTTTTATGCTGAGAATTTCGCAAGAGCATACGAGCAACT 3528
EL 3529 GGCGGTGGTGCATTTAAATACGCCTCGATTGCCAAAAAAAGGCTTCACGTTTCTTTTGAGGTTATGCGGGAGATGGACTCGGTG 3612
nonEL 3529 GGCGGTGGTGCATTTAAATACGCCTCGATTGCCAAAAAAAGGCTTCACGTTTCTTTTGAGGTTATGCGGGAGATGGACTCGGTG 3612
Cloned 3529 GGCGGTGGTGCATTTAAATACGCCTCGATTGCCAAAAAAAGGCTTCACGTTTCTTTTGAGGTTATGCGGGAGATGGACTCGGTG 3612
EL 3613 GTGCATGGACTCACTCTTCTTATTCGCTCCGCCCCATGGTCCATCTTTACGGTAGACCCCACAACTGGAATTCACTACCCACAC 3696
nonEL 3613 GTGCAAGGACTCACTCTTCTTATCCGCTCCGCCCCATGGTCCATCTTTACGGTAGACCCCACAACTGGGATTCACTACCCACAC 3696
Cloned 3613 GTGCAAGGACTCACTCTTCTTATCCGCTCCGCCCCATGGTCCATCTTTACGGTAGACCCCACAACTGGGATTCACTACCCACAC 3696
EL 3697 AAGCTGCGAAGCCCTCCCGGGGACACGCTTTCTCCCTTTCCGTGTTTGCTTGTGAACATTGGTTCCGGTATTTCAATCATCAAA 3780
nonEL 3697 AAGCTGCAAAGCCCTCCCGGAGACACGCTTTCTCCCTTTCCGTGTTTGCTTGTGAACATTGGTTCCGGTATTTCAATCATCAAA 3780
Cloned 3697 AAGCTGCAAAGCCCTCCCGGAGACACGCTTTCTCCCTTTCCGTGTTTGCTTGTGAACATTGGTTCCGGTATTTCAATCATCAAA 3780
EL 3781 TGCCTCGGCCCGGATGGCTCCCATGTTCGCGTTGGGGGAAGTCCCATTGGTGGAGCAACATTCTGGGGGCTTGTGCGCACCATG 3864
nonEL 3781 TGCCTCGGCCCGGATGGCTCTCATGTTCGCGTTGGGGGAAGTCCCATTGGTGGAGCAACATTCTGGGGGCTTGTGCGCACCATG 3864
Cloned 3781 TGCCTCGGCCCGGATGGCTCTCATGTTCGCGTTGGGGGAAGTCCCATTGGTGGAGCAACATTCTGGGGGCTTGTGCGCACCATG 3864
EL 3865 ACGGATGTAACCTCGTGGGAGGAAGTTCTGGAGATCATGCGTCTTGACGGCCCCGGTGACAACAAAAACGTTGACCTTCTGGTT 3948
nonEL 3865 ACGGATGTAACCTCGTGGGAGGAAGTTCTGGAGATCATGCGTCTTGACGGCCCCGGTGACAACAAAAACGTTGACCTTCTGGTT 3948
Cloned 3865 ACGGATGTAACCTCGTGGGAGGAAGTTCTGGAGATCATGCGTCTTGACGGCCCCGGTGACAACAAAAACGTTGACCTTCTGGTT 3948
EL 3949 GGCGATATCTACGGCTACAATGCGCATGACCTTCCAGCAATGCTTTCCGTCGACACCGTCGCCAGCAGTTTTGGCAAACTTGGT 4032
nonEL 3949 GGCGATATCTACGGCTACAATGCGCATGACCTTCCAGCAATGCTTTCCGTCGACACCGTCGCCAGCAGTTTTGGCAAACTTGGT 4032
Cloned 3949 GGCGATATCTACGGCTACAATGCGCATGACCTTCCAGCAATGCTTTCCGTCGACACCGTCGCCAGCAGTTTTGGCAAACTTGGT 4032



D.2. Alignment of PanK alleles from *T. cruzi CL Brener* and cloned *Tc*PanK used in *T. brucei* add-back experiments: *Tc*PanK alleles: EL: Esmeraldo like, nonEL: non Esmeraldo-like. Cloned: Cloned *Tc*PanK gene inserted into pTubEX for add-back of RNAi cell lines (Materials and Methods section 2.3.5, Results Chapter 4, section 4.4).