UNIVERSITY OF NAIROBI

INTEGRATING BIOINFORMATICS MECHANISM-BASED SCREENING INTO KINETOPLASTIDS ANTIPROTOZOAN rRNA TARGETS COMPOUND DISCOVERY PARADIGM

By

Harrison Ndung'u Mwangi (MSc. University of Nairobi)
Reg No: H80/51971/2017

Thesis/Dissertation submitted to the Department of Biochemistry for the award of Ph.D. degree in Biochemistry, University of Nairobi.

©March 2020

DECLARATION AND APPROVAL

I declare that this Thesis/Dissertation is my own work, has not been submitted for research at any university and relevant literature has been duly referenced.

Name: Harrison Ndung'u Mwangi

Registration: H80/51971/2017

Department of Biochemistry,

School of Medicine,

University of Nairobi

Signature: _____ **Date**___10th March, 2020_____

This Thesis/Dissertation has been submitted with our approval as University Supervisors:
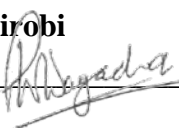
Prof, Francis Mulaa, PhD

Department of Biochemistry

University of Nairobi

Signature: _____ **Date**___20th March, 2020_____

Prof Peter Waiganjo Wagacha,

School of Computing & informatics

University of Nairobi

Signature: _____ **Date**___23rd March,2020_____

Dr Edward Kirwa Muge,

Department of Biochemistry
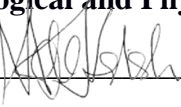
University of Nairobi

Signature: _____ **Date**___20th April,2020_____

Dr. Ndakala Albert,

Department of Chemistry, School of Physical Sciences,

College of Biological and Physical Science,

Signature: _____ **Date**___27th April,2020_____

ii

AKNOWLEDGEMENTS

DEDICATION

To my family especially my Parents Mr. James Mwangi & Mrs. Naomi Njeri and Children who put not only relentless effort but also undying support and encouragements for my achievement and my siblings: Jefferson Wanyoike, Tobias Job Wanyoike, & Stephen Chege, for their love and constant counseling during the course and all my mentors both past and present for their guidance

Abstract

Kinetoplastids are human pathogens with devastating economic and health effects, which include *Leishmania* and *Trypanosoma* species from flagellated protozoans. With the developed technology platform that allows the generation of high atomic level resolution of pathogen ribosome's crystal structures, we demonstrate that rRNA is a target of choice for the development of next-generation drugs. In addition, using several novel and transformative technologies we have developed, we demonstrate that the modular nature of rRNA facilitates the development of *in vitro* assays, structure determination, molecular-modeling, and compound screening studies for drug design. We employ computational homology and *de novo* modeling to reveal an atomic-level view of *Leishmania* and *Trypanosoma* ribosome and use the information of the rRNA structure and movement to design anti-infective-like compounds that target biologically functional ribosome RNA motifs in a predictable manner. This was performed by screening the pathogen box and the natural product databases where we got the best 40 compounds that bind well to the predicted motifs. Further analysis was conducted and mode of action of how the binding happens explained at the conclusion. Therefore, developing additional measures to control these "neglected tropical diseases" becomes increasingly clear, and we believe that the opportunities for developing drugs, diagnostics, vaccines, and other tools necessary to expand the knowledge base to combat these diseases have never been better.

Keywords: Neglected tropical diseases; homology and *de novo* modeling; rRNA motif; In vitro; Screening

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

HAT		Human African Trypanosomiasis	snRNA	Small nuclear RNAs
CNS	-	Central Nervous system	gRNA	Guide RNA
WHO	-	World Health Organization	ASL	Anticodon stem loop
CL	-	Cutaneous leishmaniasis	A-Site	Acceptor site), binds to the aminoacyl tRNA
VL	-	Visceral leishmaniasis or Kala-azar		
rRNA	-	Ribosomal Ribonucleic Acid	P Site	Peptidyl site
DNA	-	Deoxyribonucleic Acid	E Site	Exit-site is the third and final binding site for t-RNA
FRET	-	Fluorescence resonance energy transfer (FRET)		
			PTC	Peptidyl transferase center
NTDs	-	Neglected Tropical Diseases	CDHMS	
3DHMS	-	Three-Dimensional Homology Model Screening	ACE	Atomic contact energies
rDOCK	-	A software tool that performs docking		
ELISA	-	Enzyme-linked immune assay		
PDB	-	Protein Data bank		
SPP		Species		
RNA		Ribonucleic acid		
NMR		Nuclear magnetic resonance		
tRNA		Transfer Ribonucleic acid		
E.Coli		Escherichia coli		
ML		Mucosal Leishmaniasis		
DMFO		Difluoromethylornithine		
SSG		Sodium stibogluconate		
DNA		Deoxyribonucleic acid		
GA		Guanine-Adenine bond		
AU		Adenine-Uracil bond		
GU		Guanine-Uracil bond		
AC		Adenine-Cystine bond		
SiRNA		Small interfering RNA		
mRNA		Messanger RNA		
miRNA		Micro RNA		
piRNA		Piwi-interacting RNA		

CHAPTER ONE

1   Introduction

1.1   Background to the study

Kinetoplastids are a group of flagellated protozoans that are distinguished by the presence of a DNA-containing region, known as a "kinetoplast," in their single large mitochondrion (De Gaudenzi, Frasch, et al. 2005; Stuart, Brun, et al. 2008). Kinetoplastida order organism's protists have a microscopically visible complex of the mitochondrion DNA. Different pathogens in kinetoplastid undergo morphological changes during their cycles of life and have similar cellular structures and genomic organization (Akiyoshi and Gull 2014). This group includes a number of both animal and plant pathogens that are transmitted by different insect vectors and cause different human diseases (De Gaudenzi, Frasch, et al. 2005; Matthews 2005). Among this group of pathogens that causes diseases are; *Leishmania major,* which causes a range of diseases in Europe and the tropics; *Trypanosoma brucei* which causes trypanosomiasis (sleeping sickness) in Africa; *Trypanosoma cruzi* in Latin America which is an etiological agent of Chaga's disease (Simpson, Lukeš, et al. 2002; Ivens, Peacock, et al. 2005). It's important to note that the three parasites undergo a life cycle that involves transmission from one mammal to another by an arthropod vector known as digenetic (Lake, De La Cruz, et al. 1988). All other kinetoplastids are referred to as bodonids collectively, and among these, the best studied are the cryptobiids, some which parasitize fish that is commercially important (Woo 2001; Simpson, Lukeš, et al. 2002). Primarily in the tropical and subtropical areas of the world as documented half a billion people are at risks of contracting these diseases (Stuart, Brun, et al. 2008; Brun and Blum 2012). Further estimates of more than twenty million individuals are infected with the pathogens that cause them, resulting in more than one hundred thousand deaths per year and extensive suffering. With the different kinetoplastids having almost similar cellular biology, for example, all have a single flagellum originating near the large mitochondrion thus are all motile protozoans (Stuart, Brun, et al. 2008). The flagellum emits from a pocket in the cell membrane where the endocytosis occurs (Gull 2001). They have modified peroxisomes known as glycosomes which perform glycolysis and their cell membrane is underlain with a sheet of microtubules (Ivens, Peacock, et al. 2005). The cell membrane is critical for their survival hence it's further reinforced with species-specific molecules (Vickerman and Preston 1976; Stuart, Brun, et al. 2008).

1.2 Life cycles of kinetoplastids

Kinetoplastids typically grow asexually, although it is shown for *T brucei* and inferred for *T cruzi* sexual recombination occurs (Stuart, Brun, et al. 2008; Weedall and Hall 2015). This also might occur in some species of *Leishmania* although it is not obligated to either (Ivens, Peacock, et al. 2005). Kinetoplastids divide by binary fusion during which the nucleus does not undergo membrane dissolution or chromosome condensation (Vickerman and Coombs 1999; Stuart, Brun, et al. 2008). These similarities are expected given that of the more than 8,000 genes in the genome of each kinetoplastid there are more than 6,000 orthologs that are common (Stuart, Brun, et al. 2008; Gehrke 2012).
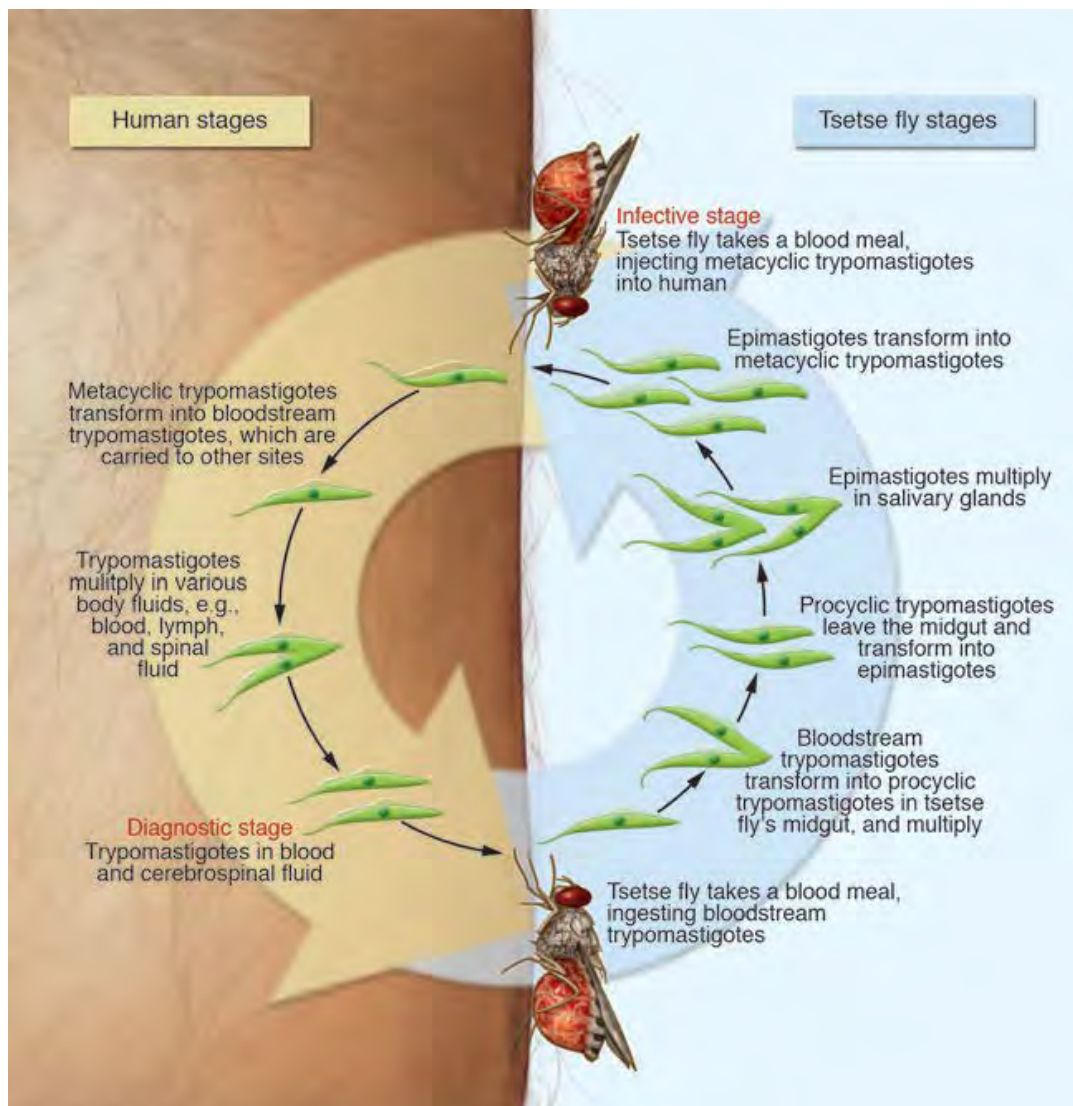
Figure 1 *T brucei* life cycle as reviewed by Stuart *(Stuart et al., 2008)*

T brucei causes Human African trypanosomiasis infection which occurs when an individual is infected with *T.b rhodesiense* or *T.b. gambiense*. These parasites are transmitted by over 20 species and subspecies of the tsetse flies (Brun, Blum, Chappuis, & Burri, 2010; Stuart et al., 2008). Out of these species, an approximation of 1 per 1000 flies can transmit the parasite because they have a mature salivary gland infection required to transmit human infection *(Stuart et al., 2008)*. During a blood meal from either an infected human or an animal, the trypanosomes enter the fly's system *(Aksoy, 2003; Franco, Simarro, Diarra, & Jannin, 2014)*.

The parasite undergoes transformations to become infective in both morphology and physiology in the salivary glands and alimentary tract *(Steverding, 2017; Stuart et al., 2008)* See figure 1. An infected fly will inject the parasite into the mammalian host during a blood meal on the skin tissue *(Stuart et al., 2008)*. The parasite spread throughout the body by the blood stream from the lymphatic system and continues to replicate by binary fusion *(Brun et al., 2010; Stuart et al., 2008)*. The parasite in the later stages resides in the cerebrospinal fluid and intercellular spaces by invading the central nervous system.

T. cruzi infection causes Chagas disease. During a blood meal by the *Triatominae* insect, the feces which contain trypomastigotes of the parasite is transmitted to humans *(Noireau, Diosque, & Jansen, 2009)*. The wound allows entry of trypomastigotes which invade nearby cells, within which they differentiate into intracellular amastigotes that multiply by binary fission *(Noireau et al., 2009; Stuart et al., 2008)*. Trypomastigote is differentiated from amastigote and enter the blood system from where they infect cells and multiple organs and tissues becoming amastigotes once again *(Tyler & Engman, 2001)*.

These include the gut, heart, CNS, adipose tissue and smooth muscles. Once again the triatomite bug gets infected during a blood meal from an infected host *(Stuart et al., 2008; Tyler & Engman, 2001)*. Transformations both morphological and physiological of the trypomastigotes occur in the midgut of the insect vector and differentiate into infective trypomastigotes in the hindgut *(Bern, 2015)*. *T. cruzi* life cycle is shown in Figure 2 below.
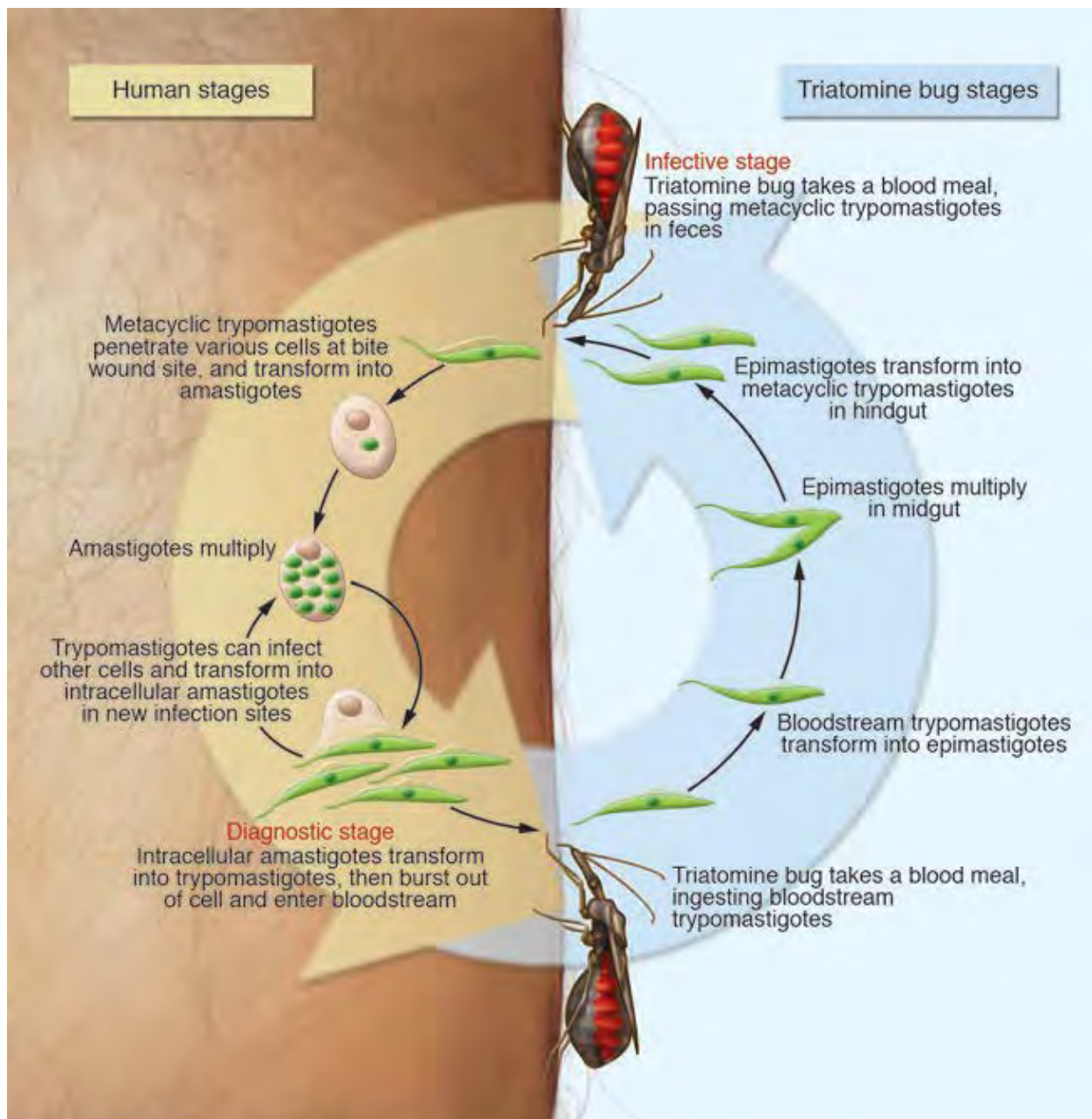
Figure 2 Life cycle of *T cruzi* as illustrated in *(Stuart et al., 2008)*

Leishmania spp infections from different species cause a distinct type of leishmaniasis (Bates, 2018). for example, cutaneous leishmaniasis is caused by any one of the several different *leishmania spp* while visceral leishmaniasis is caused by infection with either *L. donovani* or *L. infantum* (Bates, 2018; Stuart et al., 2008). The vectors liable for the transmittance of these diseases are the sand flies during a blood meal where they inject infective promastigotes (Bates, 2018; Stuart et al., 2008). Using the receptor-mediated endocytosis, the parasite infects the mammalian

macrophages and later transform into amastigotes which multiply by binary fission (Bates, 2018; Matamoros, 2016; Stuart et al., 2008). The vector acquires the infection by ingesting infected cells during a blood meal from the mammalian host (Matamoros, 2016). Transformation of the amastigotes into promastigotes happens in the gut where they eventually become metacyclic promastigotes that are now ready to infect humans as shown in Figure 3 below (Matamoros, 2016; Stuart et al., 2008).
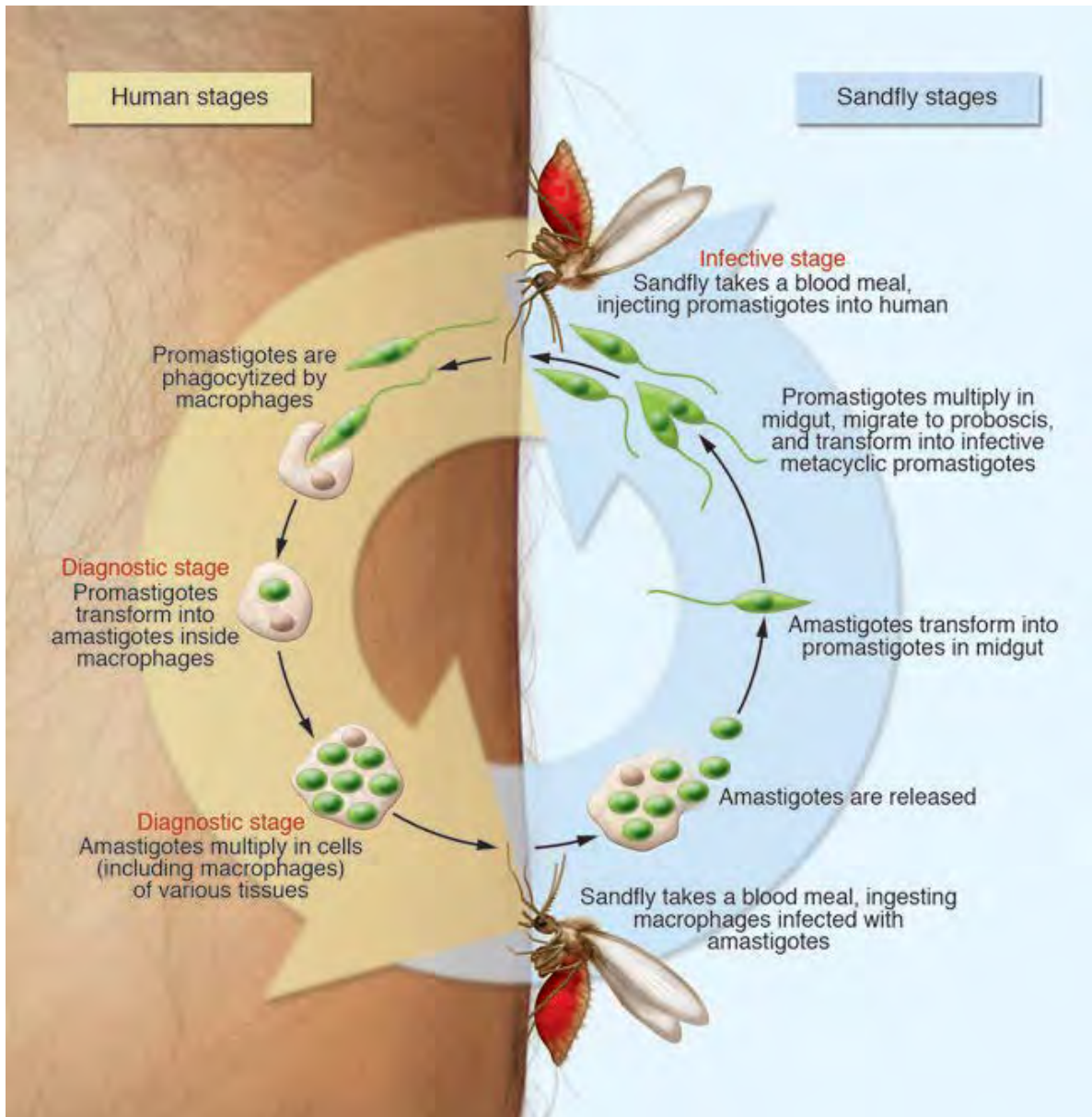
Figure 3 showing the life cycle of *Leishmania spp*

It's important to note that diseases that are caused by different kinetoplastids are very distinctive despite the similarities documented (Brun and Blum 2012; Stich 2013). Thus understanding differences between these pathogens that cause human diseases at the cellular, molecular and genetic levels will definitely provide approaches that are novel in the development of newer interventions be it in diagnostic, vaccines, drug among other tools required to deal with this existing problem. New knowledge of the parasite genome, through biochemistry and bioinformatics, provides new insights in the pathogen and host immune response will enable in achieving the overall goal. Together with the major advances, the elimination of this menace can be achieved in compliment with epidemiological studies, surveillance, vector control, and other public health programs.

1.3 Introduction to the Ribosome

The ribosome is a universal organelle that performs the function of protein synthesis. All kinetoplastids have ribosomes that do this process of translation and they are highly conserved due to their functional aspect. Ribosomes are composed of primarily a ribonucleic acid (RNA) and several proteins complexed together to achieve a functional unit. These components are either indirectly or directly involved in the process of peptide synthesis. From common knowledge, nucleotide monomers combine to form a polymer known as RNA while amino acid monomers combine to form polymers known as proteins. Due to the diverse functions that the RNA molecules play, these molecules are central to the biology of the cell. These functions, for example, may include a host of catalytic roles and various aspects of involvement in gene expression. RNA can adopt different three-dimensional structures to concur with their functional diversity while they interact with other molecules such as protein, cofactors or with other RNAs. It has been observed that there is high sequence conservation in many relevant functional motifs, and this evolutionary conservation points to the important roles that these RNAs have performed throughout time. The desire to unravel underlying mechanisms the various roles that RNA(s) play and the RNA functions has sparked immense interest in structural biology. This is especially due to the surge in the availability of complete genomes sequences. The basic primary methods of elucidating the three-dimensional structure of biological molecules and macromolecules are X-Ray crystallography and Nuclear Magnetic Resonance Spectroscopy though there are some concerted efforts from the cryo-electron microscopy. NMR for isolated RNA domains still remains the

method of choice for solution structure determination. However due to the size limitation of NMR towards solving larger RNA newer ways have to be invented (Davis et al., 2005). Newer theoretical methods have been developed that have worked to help minimize the time it takes to obtain the three-dimensional structure of more complex RNA. These include both Homology and *de novo* modeling. The work presented in this dissertation is based on the theoretical methods taking to account highly conserved and functionally relevant ribosomal motifs from the small subunit of the selected kinetoplastids.

1.4   The rationale of the Study

This dissertation presents research work performed as a collaborative project with an eventual goal of developing newer drugs that offer treatment of kinetoplastids infections with lesser side effects, less susceptibility to the mechanism of resistance. The novelty of this work is that it brings to bear the state of art RNA and compound modeling methodologies some of which was developed in the laboratory, to yield high-resolution models of biological molecules and macromolecules. In our case, the smaller subunit ribosomal RNA of the selected kinetoplastids was modeled to enable perform virtual screening for new compounds or existing drugs that would allow therapeutic switching. Regions of 18S RNA that have demonstrated the functional significance and highest conservation within the ribosome were identified in this study as targets. The essential nucleotide motifs were selected based on a genetic *in vivo* system from high throughput identification (Denny & Greenleaf, 2018; Lee, Varma, SantaLucia Jr, & Cunningham, 1997). From various studies conducted earlier, there are some regions that have been identified to be essential for the facilitation of protein synthesis form the small subunit. This includes a region that binds the initiation factor, the tRNA and ribosomal proteins located on helix 23 and helix 24 of on *E .coli* which are essential to protein synthesis (Schluenzen et al., 2000; Stark et al., 1995). There are several structural differences between human and kinetoplastids 18S rRNA that could be exploited to tailor specific anti-infectives that only target the rRNA of this pathogenic kinetoplastids that are classified as a major class of tropical neglected diseases. To realize this goal we split this into two parts one is the determination of the three-dimensional structure of the selected kinetoplastids and identify the important motif which lies adjacent to helix 23 and 24 of the *E.coli* and other 2 motifs. Importantly as shown from various researches human helix 24 has a base modification (Pseudouridine) which is lacking to the prokaryotic homolog (Brown et al., 2017; Charette & Gray, 2000; Ofengand &

Bakin, 1997) it's therefore important to see how it appears for the selected kinetoplastids. Determining structural implication of this particular modified residue toward protein synthesis of these eukaryotes by solving the structural of all the three chosen kinetoplastids hence is important. The next part is performing virtual screening for selected lead compounds by performing lead compound-target recognition. These studies will yield results which will be used to refine further the design strategy of the compound libraries.

Identifying interesting structural domains from the 18S RNA of the chosen kinetoplastids, such as helix 23 which forms a domain known as the kink-turn motif (Huang, Ashraf, & Lilley, 2018; Klein, Schmeing, Moore, & Steitz, 2001)is important. Studies have shown this motif is highly dynamic (Matsumura, Ikawa, & Inoue, 2003; Xuesong Shi, Huang, Lilley, Harbury, & Herschlag, 2016), hence an excellent target for drugs reason being stabilization of one conformer by a target affects the process of protein synthesis (Teague, 2003). In addition to this no structure that has been solved completely for an isolated kink-turn (Klein et al., 2001; Xuesong Shi et al., 2016), therefore characterizing the RNA dynamics and ligand recognition *in silico* and *in vitro* offers some difficulty but this could be helped by this study. *In silico* structure prediction will be what we will use to determine the RNA dynamic structures for the kinetoplastid. These will be followed by identification of interesting motifs sites for drug targets, which will help perform *in silico* virtual screening of identified compound databases, and finally the *in vitro* validation of this motif–compound binding. All these studies were done at the University of Nairobi Department of biochemistry and center for biotechnology and bioinformatics.

CHAPTER TWO

2 LITERATURE REVIEW

2.1 Disease, Pathogenesis, Diagnosis, and Treatment

2.1.1 Trypanosomiasis

Kinetoplastid *Trypanosoma brucei*, transmitted by tsetse flies (*Glossina*) causes Human African trypanosomiasis (HAT; sleeping sickness) which occurs in sub-Saharan Africa (Brun, Blum, et al. 2010; Brun and Blum 2012; Stich 2013). *Trypanosoma brucei gambiense* almost causes all cases and is indigenous to the west and central Africa. In areas with periodic political instability resurgence and prevalence of this disease is strongly dependent on control measures. This could be cited in the 1990s where there were flair ups of civil unrest and lack of awareness in Angola, Democratic Republic of Congo, Sudan and Uganda (Stuart, Brun, et al. 2008; Organization 2013). Due to greater political stability, surveillance funding, greater control, and treatment programs reduced reported cases of new infection to 17,600 in 2004 (Stuart, Brun, et al. 2008; Steinmann, Stone, et al. 2015). Trypanosomiasis is among the most neglected tropical diseases with documentation indicating fewer than 12,000 cases of this disabling and fatal disease reported per year (Murray, Murray, et al. 1977; Jordan 1986; Organization 2013). If Typanosomiasis is left untreated it is fatal, the clinical presentation is complex, diagnosis and treatment difficult. The haemo-lymphatic stage which is the first stage is accompanied by fever, headache adenopathy, joint pain, and pruritus (Dumas and Boa 1988; Burri and Brun 2003). In this stage, there is the rapid growth of parasites which is countered by the host immune response, but parasite antigenic variation is enabling it to evade the immune system resulting to parasitemic waves (Dumas, Bouteille, et al. 2013; Lutumba, Matovu, et al. 2016). This is followed by the CNS stage which is accompanied by symptoms that are neurological severe including mental, sensory and sleep glitches (Stuart, Brun et al. 2008; McDonald and rH Stone 2016).

The current HAT diagnostic tools aren't satisfactorily completely, the reason being the second stage methods are insensitive and evasive (Kennedy 2008; Brun, Blum, et al. 2010, Kennedy 2013). The first stages ones do not discriminate the *T brucei* subspecies and even detect the parasite in blood or lymph aspirates from patient's microscopy and are readily not translatable into the field (Stuart, Brun, et al. 2008; Nagle, Khare, et al. 2014). The available drugs are old, requiring parasitological confirmation before beginning treatment, complicated to administer since drug toxicity is high, and can cause severe adverse reactions (Pepin 1993; Babokhov, Sanyaolu, et al.

2013; Kennedy 2013). New diagnostic methods, safe and effective drugs are urgently needed currently, the cerebrospinal fluid which is difficult to introduce in the fields is very evasive. Organized vector control on a Pan-African basis is needed to reduce the number of flies in existing foci (Buguet, Cespuglio, et al. 2015). WHO states that if national control strategies, international organizations, institutes of research, and other partners engage in collaborative action, elimination of this disease might even be possible (Organization 2013; Organization 2015).

1.1.1 Chagas disease

Chagas disease is a chronic, systemic, infection discovered in 1909 caused by *Trypanosoma cruzi* protozoan, prevalent in the Americas (Organization 2014). The disease affects about 8 million people in Latin America whose housing provides a habitat for the Triatomite insect known as "Kissing bug" commonly, which acts as the vector (Lent and Wygodzinsky 1979). *T cruzi* has many reservoirs that include both domestic and wild mammals (Stuart, Brun, et al. 2008; Rassi and de Rezende 2012; Organization 2014). Infection can be transmitted through a number of ways that include; Vertical transmission, ingestion of contaminated foods or drink, organ transplantation and blood transfusion (Control and Prevention 2006; Stuart, Brun, et al. 2008). From the infected patients 30–40% either has or will develop digestive mega syndromes, cardiomyopathy, or both (Control and Prevention 2006; Young, Losikoff, et al. 2007; Shikanai-Yasuda and Carvalho 2012; Murillo, Bofill, et al. 2016). Management and control of Chagas disease in the past three decades has undergone several improvements. Prevalence and incidence have been reduced through screening donor blood and programs involving large-scale vector control (Coura 2013). Treatment with benznidazole (nifurtimox) is reasonably safe and effective, although more effective trypanocidal drugs are needed (Bern 2015). Development of novel technologies and information, more effective tailor-made interventions could be made. Certain guided treatments could reverse or halt disease progression with improved risk stratification models availability. By contrast, detection of parasites after the acute stage of infection is very difficult due to their low numbers (Pirard, Iihoshi, et al. 2005). Using different platforms of multiple serological tests are used to try to build a consensus result to amount to a diagnosis (Pirard, Iihoshi, et al. 2005). However, the status of positive result based on only one test is inconclusive, whereas multiple tests on an individual could amount to a negative result (Avila, Pereira, et al. 1993; Pirard, Iihoshi, et al. 2005).

Thus challenges remain due to inadequate monitoring techniques during treatment efficiency and other interventions.

Growing population movements in non-endemic areas are becoming an emerging health problem on Chagas disease (Rassi and Marin-Neto 2010). Due to these reasons, there is underused early detection and treatment of asymptomatic individuals, and there is an assessment need in prospective randomized trials and potential benefits of novel therapies.

1.1.2 Leishmaniasis

Leishmaniasis disease is transmitted to humans by the bite of infected female phlebotomine sandflies by a protozoa parasite from over 21 species of *Leishmania* (Alvar, Yactayo, et al. 2006). There are over 90 known sandflies species known to transmit *Leishmania* parasites (Murray, Berman, et al. 2005; Dawit, Girma, et al. 2013). This disease range from self-healing cutaneous leishmaniasis (CL) to debilitating and lethal visceral leishmaniasis (VL or Kala-azar)(Chang and Bray 1986; Desjeux 2004).

Visceral leishmaniasis, also known as kala-azar if left untreated is fatal in over 95% of cases (Desjeux, Ghosh, et al. 2013). It is characterized by weight loss, spells of irregular fever, spleen and liver enlargement, and anemia (Chappuis, Sundar, et al. 2007; Stuart, Brun, et al. 2008). In East Africa and the Indian subcontinent, it is highly endemic and estimated 200,000 to 400,000 new cases of VL occur worldwide each year. Bangladesh, Brazil, Ethiopia, India, South Sudan, and Sudan account for over 90% new cases (Stuart, Brun, et al. 2008; Ready 2014). The various programs in South-East Asia for elimination are making sustained progress towards elimination, and cases are declining in Bangladesh, India, and Nepal (Desjeux, Ghosh, et al. 2013).

The most common form of leishmaniasis is **Cutaneous leishmaniasis** causing skin lesions, mainly ulcers, on exposed parts of the body, leaving life-long scars and serious disability (Haddad, Herbert, et al. 2013). The Americas, the Mediterranean basin, the Middle East and Central Asia account for about 95% of its cases (Pavli and Maltezou 2010). Algeria, Brazil, Afghanistan, Colombia, Iran, and the Syrian Arab Republic account for over two-thirds of new cases and estimated 0.7 million to 1.3 million new cases occur worldwide annually (Desjeux 2004; Stuart, Brun, et al. 2008).

Mucocutaneous leishmaniasis leads to partial or total mucous membranes of nose, mouth and throat destruction. Bolivia, Peru, and Brazil account for almost 90% of mucocutaneous leishmaniasis cases (Alvar, Yactayo, et al. 2006; Hailu, Dagne, et al. 2016).

Diagnosis entails microscopic detection of the organism in the lesions and skin tests and also seen in splenic and bone marrow aspirates and immunologic studies that include ELISA (Goto and Lindoso 2012). An accurate test and cure is useful for evaluation of novel therapeutics of leishmaniasis

2.2 Role of chemotherapy in kinetoplastid control

From section 2.1 the diseases caused by these selected kinetoplastids are among the most important parasitic diseases. This is because they affect millions of people and are considered to be within the most relevant group of neglected tropical diseases (Castillo, A Dea-Ayuela, Bolás-Fernández, Rangel, & E Gonzalez-Rosende, 2010; Hammarton, Mottram, & Doerig, 2003). To control such parasitosis chemotherapy has been used over the years. However, current chemotherapeutic treatments are far from being satisfactory offering many undesired side effects (Hammarton et al., 2003). Various studies have outlined their understanding of different drugs against leishmaniasis, African sleeping sickness, and Chagas disease, their mechanism of action and resistance (Castillo et al., 2010; Hammarton et al., 2003). Recent approaches in the area of anti-leishmanial and trypanocidal therapies are also enumerated, new modulators from the mode of action, development of new formulations of old drugs, therapeutic switching and "*in silico*" drug design.

2.3 The threat of Drug Resistance and Drug Toxicity

Various drugs have been used over a period and some of the challenges are documented based on their efficacy and toxicity during administration. An example is in the treatment of Chagas disease using nifurtimox and benznidazole which are derivatives of Nitroheterocyclic (Castillo et al., 2010). The activity of this drug has been shown to significantly vary in efficacy according to geographical locations. It has significant activity in the acute and early chronic stages; however has a very low antiparasitic activity in the chronic form of the disease (Alcântara, Ferreira, Gadelha, & Miguel, 2018; Castillo et al., 2010; S. L. Croft, 2008; Simon L. Croft, Barrett, & Urbina, 2005). This now warrants the further need for research and to improve its efficacy.

Current drugs that we rely on to treat HAT present problems in administration, efficacy, compliance and have adverse effects (Castillo et al., 2010; Hammarton et al., 2003). The treatment

of the first stage of the infection Pentamide and suramin are used. During the second stage of infection approved treatment by melarsoprol and eflornithine (DMFO) are used.

2.3.1 Chemotherapies in Use and drugs in clinical trials

In the treatment of both leishmaniasis and trypanosomiasis synthetic organic arsenicals and antimonials are used (Castillo et al., 2010; S. L. Croft, 2008). Examples of these drugs are pentavalent antimonials; meglumine antimonate, and sodium stibogluconate (SSG) which are still the preferred first-line treatments against leishmaniasis (Bates, 2018; S. L. Croft, 2008). Importantly to note is that these drugs mode of action is not clearly elucidated yet, and it's thought that they are pro-drugs and the site and mechanisms of reduction remain unclear(Castillo et al., 2010; S. L. Croft, 2008). Preceding research work suggested the inhibition of macromolecular biosynthesis in amastigotes by antimonials (Herman, Gallalee, & Best, 1987), but recent data suggests induction of efflux of intracellular trypanothione and inhibition of trypanothione reductase by antimony (Wyllie & Fairlamb, 2006). In addition to these two mechanisms, pentavalent antimony has been shown to bind to the ribosome moiety forming a stable complex which induces apoptosis by DNA fragmentation and externalization (Sudhandiran & Shaha, 2003). Resistance to these drugs has been documented in various areas and their uses are very limited due to efficacy issues and have been shown to have reduced activity in absence of t-cell immune response (Cruz et al., 2006). Resistant strains also offer challenges causing a high rate of therapeutic failure as has been shown by selective uptake of the therapeutics in various pathways (Legros et al., 2002; Nok, 2003; Sundar & Goyal, 2007; Wyllie, Vickers, & Fairlamb, 2008).

Antibiotics such as *amphotericin B* and *Paromomycin* have been used as common drugs for leishmaniasis. Amphotericin B is a natural metabolite derived from *Streptomyces nodosus*. *Amphotericin B* molecule works best on the cell membrane due to its structure. It works on the parasite membrane sterols and inserts ergosterol causing an increase in permeability leading to death (Castillo et al., 2010; Romero & Morilla, 2008). *Amphotericin B* mechanism of action is responsible for toxicity in mammals and hence its use is restricted due to toxic side effects (Castillo et al., 2010; Wasan, Morton, Rosenblum, & Lopez-Berestein, 1994). There is no documentation of resistance to AmB for *leishmania* infections but resistant promastigotes have been selected *in vitro* (Castillo et al., 2010; Kaur & Rajput, 2014; Mbongo, Loiseau, Billion, & Robert-Gero, 1998). *Paromomycin* is an aminoglycoside broad-spectrum antibiotic that belongs to the neomycin family (Castillo et al., 2010; Sundar & Chakravarty, 2008). From studies conducted in both Africa and

India by the World Health Organization and Drugs for Neglected Disease initiative, treatment using *paromomycin* has poor outcomes with parenteral administration (Castillo et al., 2010). Ribosomes have been shown to be primary targets of paromomycin especially the smaller subunit interfering with the process of translation of mRNA to protein (Mwangi, Wagacha, Mathenge, Sijenyi, & Mulaa, 2017a). Paromomycin has undesired side effects such as permeability, membrane fluidity alterations and induction of respiratory dysfunction (Maarouf, Lawrence, Brown, & Robert-Gero, 1997)

2.3.2   Recent Development and the need for newer Drugs

Recent developments in both taxonomic and functional profiling from sequenced data have given good highlights on the biochemical pathways of kinetoplastids. The availability of completed whole genomes sequencing techniques and the emergence of *In silico* bioinformatics methodologies enables the rational design or screening of potential new targets against kinetoplastids that cause disease. Studies on established drugs for other disease treatment (Pathologies) also known as therapeutic switching, together with the high throughput screening of chemical libraries, along novel constructs or combination therapy are all representative of alternative strategies that have been embraced. Table 1 summarize new formulations from High Throughput Screening, an effective way to save cost and time for already approved drugs for other uses to assess their potential as drugs for newer pathologies

Origin	Chemical group	Compound	Disease
Anticancer	Alkylphospholipids	Miltefosine1,2	Leishmaniasis
	Ornithine derivatives	Eflornithine3	HAT
	Imidazoloquinolinone	Tipifarnib	Chagas
Antifungal	Azoles	Ketoconazole Posaconazole	Leishmaniasis/ Chagas
	Polienes	Amphotericin2	Leishmaniasis
	Allylamines	Terbinafine	Chagas
Amebicide	Aminoglucoside	Paromomycin1	s Leishmaniasis
Osteoporosis	Bisphosphonates	Pamidronate, Risedronate Ibandronate	Leishmaniasis/ HAT
Hypercholesterolemia		Mevinolin	Chagas
Hyperuricemia	Purine derivatives	Allopurinol4	Leishmaniasis/ HAT Chagas
Pneumonia	Diamidines	Pentamidine2	Leishmaniasis/ HAT Chagas

| Onchocercosis | Quinolines | Sitamaquine5 | Leishmaniasis |
| mmunomodulator | Quinolines | Imiquimod | Leishmaniasis |

Table 1 Therapeutic Switching examples from high throughput sequencing *(Castillo et al., 2010)*

Another example shown in table 2 is the use of virtual screening to identify antiparasitic molecules towards kinetoplastids *(Alcântara et al., 2018; Castillo et al., 2010; Simon L. Croft et al., 2005).*

Compound	Pharmacological action	Anti-Parasite activity
Alexidine	Antiseptic Disinfectant	*T. brucei [167]*
Aklavine	Antineoplastic	*T. brucei [167]*
Anthothecol	Natural product (antitumor)	*T. brucei [167]*
Dequalinium chloride	Antiseptic Disinfectant	*T. brucei [167]*
7,8-Dihydroxyflavone	Flavonoid	*T. brucei [167]*
2,3,29-Triacetoxy-24-nor-1,3,5,7-friedelatetraene	Natural product (antitumor)	*T. brucei [167]*
Gambogic acid	Natural product (antitumor)	*T. brucei [167]*
Lycorine	Protein synthesis inhibitor	*T. brucei [167]*
Orlistat	Lipid inhibitor	*T. brucei [167]*
Patulin	Natural product (antitumor)	*T. brucei [167]*
Pristimerin	Natural product (antitumor)	*T. brucei [167]*
Pyrvinium	Antihelmintic	*T. brucei [167]*
Rutilantin	Antibiotic (antineoplastic)	*T. brucei [167]*
Rutilantone	Antibiotic (antineoplastic)	*T. brucei [167]*
Thimerosal	Antiseptic Disinfectant	*T. brucei [167]*
NSC# 13512	Anticancer assays	*L. major [168]*
NSC# 83633	Anticancer assays	*L. major [168]*

| NSC# 351520 | Anticancer assays | *L. major [168]* |

Table 2 Virtual Screening of chemical libraries to identify anti kinetoplastids Compounds *(Castillo et al., 2010; S. L. Croft, 2008)*

2.4   Introduction to the RNA Structure

A polymer of nucleotides that contain ribose sugar can be attributed to as an RNA molecule. The polymer chains are connected by a phosphodiester bonded ribose sugar to the four common bases purines: guanine and adenine pyrimidines: uracil and cytosine. It is important to note that the RNA structure differs from the DNA in several ways which are: RNA is single-stranded whereas DNA is double-stranded predominantly (Schuwirth et al., 2005), RNA contains ribose sugar whereas DNA contains 2' deoxyribose sugar, and finally RNA contains a uracil base instead of the thymine found in the DNA. Both DNA and RNA form hydrogen bonds between pairs of bases which are governed by Watson crick rules (Watson & Crick, 1953) where a complementary system is achieved for guanine and cytosine in both, adenine and uracil in RNA whereas for DNA you have adenine and thymine. Sometimes we have bonding that does not follow the Watson –crick rules giving rise to mismatches known as a non-canonical base pairing (Leontis, Stombaugh, & Westhof, 2002; Leontis et al., 2002). Common ones are GA, imino, GA, AU reverse hoogsteen, GU and AC wobble pairs (Leontis et al., 2002; Nagaswamy et al., 2002) as shown in Figure 4.

The RNA forms duplex structures and can adopt secondary structural elements that are different such as bulges, hairpins, internal loops, mismatches, pseudoknots and multi-branched junctions (Chastain & Tinoco Jr, 1991; Hendrix, Brenner, & Holbrook, 2005; Steger, Riesner, Maurel, & Perreault, 2017; Xu & Chen, 2015) as shown in Figure 5. These structures now can form very diverse RNA motifs that can participate in functional roles such as interaction recognition, enzymatic activity, and metal binding. RNA can be of different kinds playing part in different functional roles such as gene expression (coding RNA) (Crick, 1970), and non-coding RNAs (Dinger et al., 2008; Eddy, 2001; Mattick & Makunin, 2006; Rassoulzadegan et al., 2006). The non-coding include Small interfering RNA (siRNA) which are short and form a double-stranded binding to mRNA playing a role in gene silencing (Paul, Good, Winer, & Engelke, 2002; Roeder, 2005; Valencia-Sanchez, Liu, Hannon, & Parker, 2006; Wittrup & Lieberman, 2015). Micro RNA (miRNA) regulates expression of genes by mRNA cleavage or translational repression (Hausser & Zavolan, 2014; He & Hannon, 2004). Guide RNA (gRNA) in trypanosomes perform the

function of guiding uridine insertion into mRNAs (McAdams et al., 2015; Weng et al., 2008). Piwi-interacting RNA (piRNA) in germline cells perform the process of gene silencing of retrotransposons and other genetic elements (Iwasaki, Siomi, & Siomi, 2015; Kim, 2006). Small nuclear RNAs (snRNA) in eukaryotes are transcribed by RNA polymerases and are involved in the regulation of transcription factors and RNA splicing (Hernandez, 2001). Small nucleolar RNAs guide the process of methylation and pseudouridylation of ribosomal RNAs which are both chemical modification (Dupuis-Sandoval, Poirier, & Scott, 2015; Jády & Kiss, 2001). Transfer RNA tertiary structure adopts an L shape with the 3' ends being sites for amino acids attachment and anticodon stem loop (ASL) sequence that recognizes the correct codon in the mRNA for fidelity during protein synthesis (Jobe, Liu, Gutierrez-Vargas, & Frank, 2018; Rozov, Demeshkina, Westhof, Yusupov, & Yusupova, 2015).
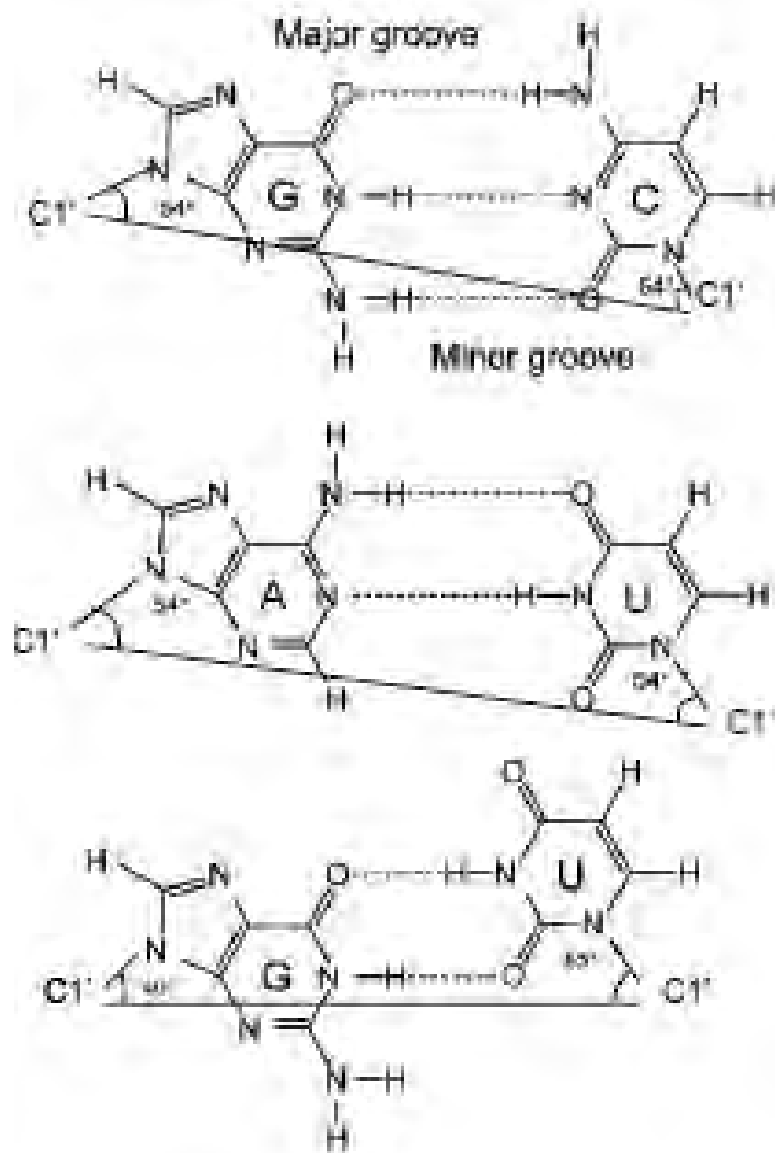
Figure 4 Watson Crick base pairing shown with a similar angle of 54 and a wobble base pair gives a different shape
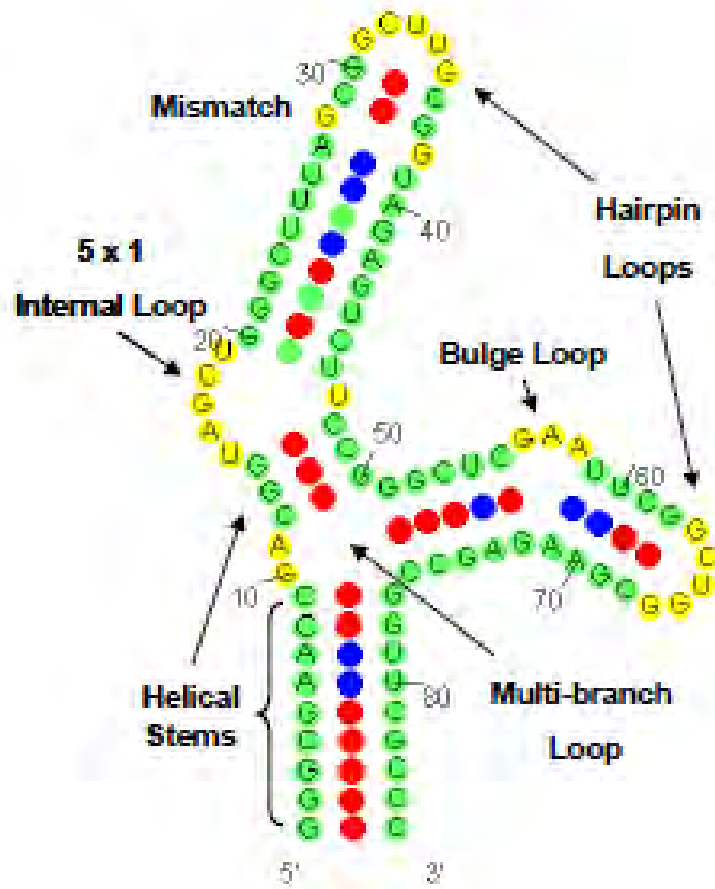
Figure 5 Secondary structure of RNA showing various motifs

2.5 The Eukaryotic Ribosome

The ribosome as a macromolecular organelle catalyzes peptide bond formation during protein synthesis. The eukaryotic ribosome has two subunits and designated as the 80S. The two subunits contain a larger one known as the 60S subunit and the smaller one known as the 40S subunit. The smaller 40S ribosomal subunit contains 18S ribosomal RNA and approximately 33 proteins, whereas the larger 60S ribosomal subunit contains three RNA 5S, 5.8S, and 28S and approximately 49 proteins (Ben-Shem et al., 2011; Melnikov et al., 2012; Mwangi et al., 2017a). Shown in Figure 6 below In recent times various experimental techniques such as X-Ray crystallography and cryo-electron microscopy has been used to solve the three dimensional structure of ribosome without complexed cofactors successfully (Ben-Shem, Jenner, Yusupova, & Yusupov, 2010; de Loubresse et al., 2014; Schluenzen et al., 2000; Wimberly et al., 2000). With the increase of this structures, theoretical methods of structure prediction that have given a very good insight have been

developed increasing the efficiency and reducing the time and cost of performing the experimental methods (Mwangi et al., 2017a; Sijenyi et al., 2012). These first ribosome structures revealed additional moieties of proteins, and expansion segments shapes and locations (de Loubresse et al., 2014; Spahn et al., 2001).

Recent structures have been able to offer more clarity that has helped understand the process of protein synthesis (Ben-Shem et al., 2010; Mwangi et al., 2017a). These structures have helped show the ribosomal RNA with all the expansion segments and how they complex with the ribosomal proteins for both the small and the larger subunit, the A.site, E.site which have helped decipher the process of protein synthesis (Hinnebusch, 2014; Rozov et al., 2015; Rozov, Demeshkina, Westhof, Yusupov, & Yusupova, 2016). The different segments of the ribosome will be addressed in the following sections in more detail.
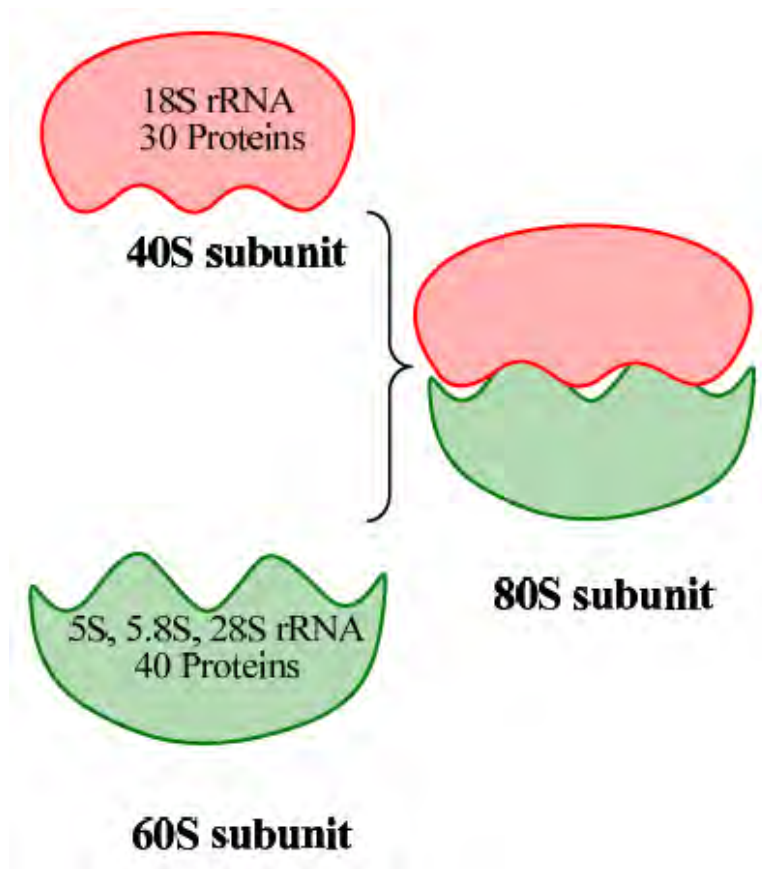
Figure 6 An oversimplified structure illustration of the eukaryotic ribosome showing the different subunits and their content *(Mwangi et al., 2017a)*.

2.5.1 The Eukaryotic 40S Small Subunit

From the various experimental methods that have been used to obtain the crystal structure of the ribosome, much atomic detail to the functional states has been revealed from the whole ribosome. Examples are the consorted efforts that obtained the structure of the *thermos thermophillus* 30S one from MRC, Cambridge (Wimberly et al., 2000) and that from Max Plank/Weizmann Institutes group (Schluenzen et al., 2000). Both groups gave evidence that the three dimensional structures could be described clearly in personalized forms which have been further illustrated by other groups that have pursued to entangle the eukaryotic ribosome. The eukaryotic small ribosomal subunit could be subdivided into three regions the head, neck and body which incorporates a spur, shoulder and platform, (Jobe et al., 2018; Mwangi et al., 2017a; Scaiola et al., 2018). The subunit contains the 18S rRNA which has approximately 1700 nucleotides for *Plasmodium falciparum* and can be divided into four domains the 3' minor domain, the 3'major, the central domain and 5'domain (Mwangi et al., 2017; Rabl, Leibundgut, Ataide, Haag, & Ban, 2011). The 40S ribosomal subunit also contains 34 proteins which are associated with the 18S rRNA (Mwangi et al., 2017a; Rabl et al., 2011) The main purpose of the eukaryotic 40S ribosomal subunit is to decode the genetic information from mRNA transcript (Jackson, Hellen, & Pestova, 2010; Rabl et al., 2011). The 40S subunit interacts with factors of initiation, initiator tRNA and mRNA playing a key role in the assembly of the initiation complex before the commencement of the synthesis of the proteins (Hinnebusch, 2014; Hinnebusch & Lorsch, 2012). The subunit also plays a key role of ensuring translational fidelity in protein synthesis by ensuring that the right anticodon-codon interaction are maintained (Dale & Uhlenbeck, 2005; Fahlman, Dale, & Uhlenbeck, 2004; Olejniczak, Dale, Fahlman, & Uhlenbeck, 2005) In the 18S rRNA, the helix 44 has been shown to be involved in the protein synthesis (Decatur & Fournier, 2002; Rabl et al., 2011; Yoshizawa, Fourmy, & Puglisi, 1999) which forms part of the A. site discussed in detail in the next section.

2.5.2 Inter subunit association

The small ribosomal subunit interacts with the large subunit to achieve the functionality in the course of protein synthesis using a magnesium dependent fashion (Ben-Shem et al., 2010; Selmer et al., 2006). Approximately 12 intersubunit bridges have been identified for the prokaryotic 70S ribosome (Schuwirth et al., 2005; Spahn et al., 2001). In the process of elongation these bridges have been proposed to undergo breakage and rearrangement during protein synthesis (Schuwirth et al., 2005; Valle et al., 2003; Yusupov et al., 2001) resulting to a ratchet motion (B. Chen et al.,

2015; Frank & Agrawal, 2000). The bridges include interaction of RNA-RNA, RNA-protein both RNA-RNA to RNA-protein and finally protein to protein interactions (Khatter, Myasnikov, Natchiar, & Klaholz, 2015; Yusupov et al., 2001; Yusupova & Yusupov, 2017). The small subunit rRNA has been shown to have residue at a specific position that is kingdom specific as shown for position 702 residue (Cannone et al., 2002).This position residue forms a helical region (helix 23) which adopts an interesting V shape motif known as a kink turn which is considered to be dynamic (Huang & Lilley, 2016; Klein et al., 2001). In the course of translocation of both the mRNA and tRNA both the two ribosomal subunits have been shown to undergo a ratchet like motion (Spahn et al., 2004; Valle et al., 2003). This process of contact breaks has been shown to continue during the process of translocation where bridges at both heads of the subunits are in contact (Cannone et al., 2002; Rozov et al., 2015). One of the bridge contacts of the small subunit includes the A-site and the P-site (Carter et al., 2000; Khatter et al., 2015) which is discussed in the following section.

2.5.3 Interactions of the tRNA and the ribosome

From the section above tRNA molecules interact with the ribosome during protein synthesis at three functionally distinct sites namely, the P site, A site, and the E site. This sites that bind to the tRNA (Carter et al., 2000; Khatter et al., 2015) are explained in details in the following sections

2.5.3.1 A-Site

The A site because of its role in ensuring that bases correctly pair between the mRNA codon and the tRNA having the cognate anticodon is sometimes known as the decoding site. Studies show a portion of the A site is located in the small subunit 16S rRNA at helix 44 for the *E.coli* stretching from position 1404-1412 and 1488-1497 (Clos II, 2007; Sijenyi, 2008). In this location, it is ensured that the addition of the new amino acid takes the correct tRNA to grow the polypeptide chain (Ramakrishnan, 2002). The new amino acid at 3'end of the tRNA at the A site is made available for the formation of the peptide bond at the peptidyl transferase center (PTC) of the larger subunit (de Loubresse et al., 2014; Selmer et al., 2006). At the small ribosomal subunit decoding center, there is a clearly visible anticodon stem loop (ASL) comprising nucleotides 26-44 of the tRNA (de Loubresse et al., 2014; Selmer et al., 2006). The translation initiation has been suggested to happen when a stable cognate pair between mRNA/tRNA complex five hydrogen bonds binds to the three universally 16S rRNA conserved bases G530, A1492, and A1493 (Ramakrishnan, 2002; Xinying Shi, Chiu, Ghosh, & Joseph, 2009; Sijenyi, 2008). Translation occurs only when a

stable pairing happens meaning a non-similar anticodon and codon pairing is not stable enough to propagate this process.

2.5.3.2 P Site

The site at which the initiator tRNA binds before translocation begins on the small ribosome is known as the P site. The AUG codon recognition site on the mRNA to be translated by the small subunit signals the first amino acylated binding to the initiator tRNA (Dever, 2002; Ramakrishnan, 2002). The P site that ensures codon fidelity is achieved during initiation and translation (Nissen, Hansen, Ban, Moore, & Steitz, 2000; Sijenyi, 2008). This site spans in the larger subunit becoming the second site binding the tRNA in the ribosome during translation (Nissen et al., 2000).

The P site maintains the reading frame as it has evolved to hold the tRNA tightly in a position which is important for peptidyl transfer (Selmer et al., 2006). The P site bound to the tRNA in the small 30S subunit is stabilized by A-minor type I and type II interactions between the 16S rRNA G1338 and A1339 and GC base pair 30-40 and 29-49 of the P site tRNA (ABDI & FREDRICK, 2005; Khatter et al., 2015). Residues A1339 and G1338 in the 16S rRNA Interact with the ASL (Anti-codon Stem-Loop) of the P site tRNA on one side and nucleotide 790 found at helix 24 of the 30S platform on the other side (Nierhaus, 2004; Sijenyi, 2008; Weisser, 2017). This ensures the tRNA does not move into the E site. For the translocation of the tRNA from the P site to the E site the two elements act as a translocation switch and as an initiator by moving apart (Fredrick & Noller, 2002; Noller, Yusupov, Yusupova, Baucom, & Cate, 2002; Steitz, 2008). Ribosomal proteins S13 and S9 interact with the P site bound tRNA via phosphate oxygen's (For S13 position 29 and position 33 and 34 for the S9 protein) (Fredrick & Noller, 2002; Selmer et al., 2006). The P site acceptor end interacts with the larger 50S subunit peptidyl transferase center (PTC) (Schmeing, Huang, Strobel, & Steitz, 2005; Schmeing et al., 2002; Schmeing & Ramakrishnan, 2009). The 23S rRNA D site at H69 makes direct minor groove interaction with the helix at nucleotide 11 and 12 and also the adjacent nucleotide 24 and 25 in the anti-codon stem, apart from just ASL and PTC interactions (Sijenyi, 2008; Yusupov et al., 2001; Yusupova & Yusupov, 2017).

2.5.3.3 Peptidyl Transferase Center

The PTC center is the point at which the peptide bond formation takes place growing the peptide to the incoming tRNA from the A site. This site is located at the larger ribosomal subunit. The PTC comprises 23S rRNA domain V primarily which facilitates the formation of the peptide bond by positioning first the 3' CCA ends of the aminoacyl-tRNA and peptidyl via the P- and A- loops

binding (Schmeing et al., 2002; Sijenyi, 2008). The formation of the peptide bond in the PTC occurs via the primary amine of the A site nucleophilic attack bound tRNA's amino acid on the carbonyl carbon of the P site bound tRNA's amino acid (Selmer et al., 2006; Weixlbaumer et al., 2008). Studies done previously show that the larger ribosomal subunits (the 50S) are active in peptide bond formation, but the rate of catalysis is lower by a factor of about 1000 than that of intact (the 70S) ribosome (Katunin, Muth, Strobel, Wintermeyer, & Rodnina, 2002; Kazemi, Sočan, Himo, & \AAqvist, 2018; Rodnina, Beringer, & Wintermeyer, 2007). It is a bit unclear how ribosome accelerate the process of peptide formation, but current hypothesis suggests mostly if not entirely due to the substrate positioning within the active site (\AAqvist & Kamerlin, 2015; Doerfel et al., 2015; Sievers, Beringer, Rodnina, & Wolfenden, 2004; Youngman, Brunelle, Kochaniak, & Green, 2004) coupled to substrate-assisted catalysis (Kazemi, Himo, & \AAqvist, 2016; Kazemi et al., 2018; Polikanov, Steitz, & Innis, 2014; Weinger, Parnell, Dorner, Green, & Strobel, 2004)

2.5.3.4 E site

The area occupied by the translocated deacylated tRNA from the P site is known as the E site. Both the small (30S) and larger (50S) ribosomal subunits have been proposed to bind to the E site bound tRNA making interactions with the mRNA codon (Geigenmüller & Nierhaus, 1990; Koch et al., 2015). Recent studies from *Thermus thermophilus* 2.8 Å resolution crystalline structure (Selmer et al., 2006) however, show no E site codon-anticodon interaction. The middle anticodon base A35 was found to be closer to 16S RNA G693 than to the E-site codon (B. Chen, 2015; Selmer et al., 2006). This information is thus consistent with previous studies (Carter et al., 2000) where the small subunit (the 30S) E site was claimed to consist of S11 and S7 ribosomal proteins primarily (Carter et al., 2000; Robert & Brakier-Gingras, 2003). The E site tRNA acceptor end in the large (50S) subunit interact with residues at the base of H82, D and T loop basses interact with the L1 stalk, these interactions are thought to stabilize the stalk in a closed conformation (Selmer et al., 2006).

2.6   Homology and *De novo* Modeling

De novo modeling also referred to as *ab initio* modeling, predict the structure from the sequence alone and seek to build models from scratch. It relies on physical principles without any previously solved structures. However, *de novo* modeling procedures require vast computational resources

and only been carried out for tiny molecules. Homology modeling also known as comparative modeling, predict the tertiary structures by relying on any significant similarity; usually more than 70% similarity, at the fold level between known structures and the modeled sequence. Homology modeled structures in most cases have an accuracy that can be comparable to low resolution, experimentally determined structures.

2.6.1 Compound Screening against rRNA

A critical problem in combating protozoan infections is that many of the existing leishmaniasis and Typanosomiasis drugs target eukaryotic processes common to both the protozoa and mammals. Several different anti-protozoan drug classes have no known target and the potential reality of drug toxicity are readily apparent because these drug targets are likely to be homologous between this flagellated protozoa and mammals (Cavalli and Bolognesi 2009). Picking new targets is complicated by the many other similarities among all eukaryotes at the levels of cell biology and biochemistry. These facts point to a need to identify and utilize new protozoan targets that are absent in humans for the development of more effective and safer anti-protozoan agents.

Meanwhile, new drug targets in apicomplexan parasites are difficult to validate given the significant gaps in our knowledge of pathogen biology. To overcome these challenges, we developed a novel synthetic ribosome chemical screen that opens new avenues for identification of specific ribosome inhibitors and validation of ribosome drug targets. We propose to move the *Trypanosomiasis and Leishmaniasis* ribosome into the computer and back and develop both a digital and synthetic ribosome chemical screen to identify new inhibitors with a novel mode of action targeting ribosome function.

The future of treatment for many Apicomplexan diseases lies in uncovering mechanisms that are newly emerging or are still to be discovered in the parasite (Elsheikha 2014). With combination-focused technology pipeline that exploits the power of four scientific platforms to help address unmet drug discovery needs in a host (Mwangi 2013). We believe that the best way to help patients is to focus on breakthrough science to discover these mechanisms and develop novel, targeted therapies that interact with them (Guiguemde, Shelat, et al. 2010).

One of the major challenges in Apicomplexan research is to leverage advances in genome sequencing into lead therapeutic modalities to treat Apicomplexan infections and antibiotic resistance. This functional genomics approach holds great promise to not only advance

apicomplexan therapeutics, but to provide highly selective chemical probes of function to study disease biology. These technical challenges are compounded by a lack of well-validated drug targets as well as a poor understanding of the optimal pharmacology for the treatment of apicomplexan infections (Mwangi 2013).

Our laboratories have developed a technology for the generation of high-resolution 3D atomic-level view structures of pathogen Ribosomes, by **moving the ribosome into the digital world and back** using a combination of homology and *De novo* modeling. This developed RNA platform allows for systematically studying RNA-ligand interactions, for the rapid screening of compounds against ribosome and RNA motifs (Cruz, Blanchet, et al. 2012; Sijenyi, Saro, et al. 2012). The invention of homology and *de novo* modeling of biological macromolecules high-resolution crystal structures such as the *P falciparum* 40S and *M. tuberculosis* 30S offers another novel process that uses the three-dimensional structure for identifying ribosome-related ligands and for designing ligands (Mwangi 2013; Mathenge, Mwangi, et al. 2015). These ligands have specific ribosome binding properties that may act as protein synthesis inhibitors. Thus the combination of this unique process may be used to produce ligands that could be designed to specifically kill or inhibit the growth of any target organism especially the causative agents of tropical neglected diseases (Warrier, Martinez-Hoyos, et al. 2015). In addition, this process could be used to realign prior used drugs to a specific target that have developed resistance to others that could be possible targets.

We used this technology platform to determine how small molecules interact with and alter the protein synthesis ability of the *Trypanosoma* and *Leishmania* ribosome RNA structure and its movement over time. Based on the results of these ligand-RNA interactions, we designed compounds that, even in small amounts, significantly inhibit the *Trypanosoma and Leishmania* target cells. We screened compounds from pathogen box screening program and compound libraries, against the generated 3D structures of *Trypanosomiasis and Leishmaniasis* Ribosome subunits using *in silico* techniques. Targets were validated using the immobilized compound array and identified RNA motifs.

Our group was involved in the rational design of small molecules that target ribosome RNA and in solving the structures of these small molecules in complex with their RNA targets. These studies not only allow us to understand the molecular and atomic level interactions that drive association of complexes but also allow us to rationally design improved small molecules that target RNA (Scott, Hünenberger, et al. 1999).

2.6.2 Screening and structural characterization methodology and *in vivo* studies

In order to effectively characterize potential lead compounds in structural analysis and *in vitro* validation studies, the focus was on the development of one rRNA target per pathogen. *In vitro* techniques allow specific biological and mechanistic pathways to be isolated and tested under controlled conditions (Oberdörster, Maynard, et al. 2005). We examined the potential targets in *Trypanosoma and Leishmania* 18S rRNA from modeling studies performed by our research group, for suitability as targets for small molecule screening based on a number of criteria including size and complexity, presence or absence of modified nucleotides, known interactions with proteins, RNA or other ligands, and differences from corresponding human 18S rRNA sequence.

2.6.3 Homology Modeling of RNA Domains

RNA homology modeling software developed by Dr. John Santa Lucia at Wayne State University was used to predict the structures of the wild-type target, viable mutants, and the corresponding human sequences (Sijenyi, Saro, et al. 2012; Mwangi 2013). The software requires the input of a known 3D structure, followed by sequence alignment to determine where in the 3D structure substitutions, deletions or insertions were to be made (Cruz, Blanchet, et al. 2012; Sijenyi, Saro, et al. 2012). The sequence alignment for each domain was done using a structure-based sequence alignment program, which automatically performs the sequence alignment subject to the constraint that the paired residues from both secondary structures are in the proper register. The accuracy of the predictions has been validated for *P. falciparum* and *Mycobacterium Tuberculosis* rRNAs with sequence identity as low as 50% and predicted structures for RNA domains with all-atom RMSDs near 3 Å were typically observed (Mathenge 2013; Mwangi 2013). We used this modeling software in the studies to predict the structures of the target. The study primarily focused on two RNA targets. One prioritized target was chosen for further studies and while the additional targets were considered as time and resources permitted.
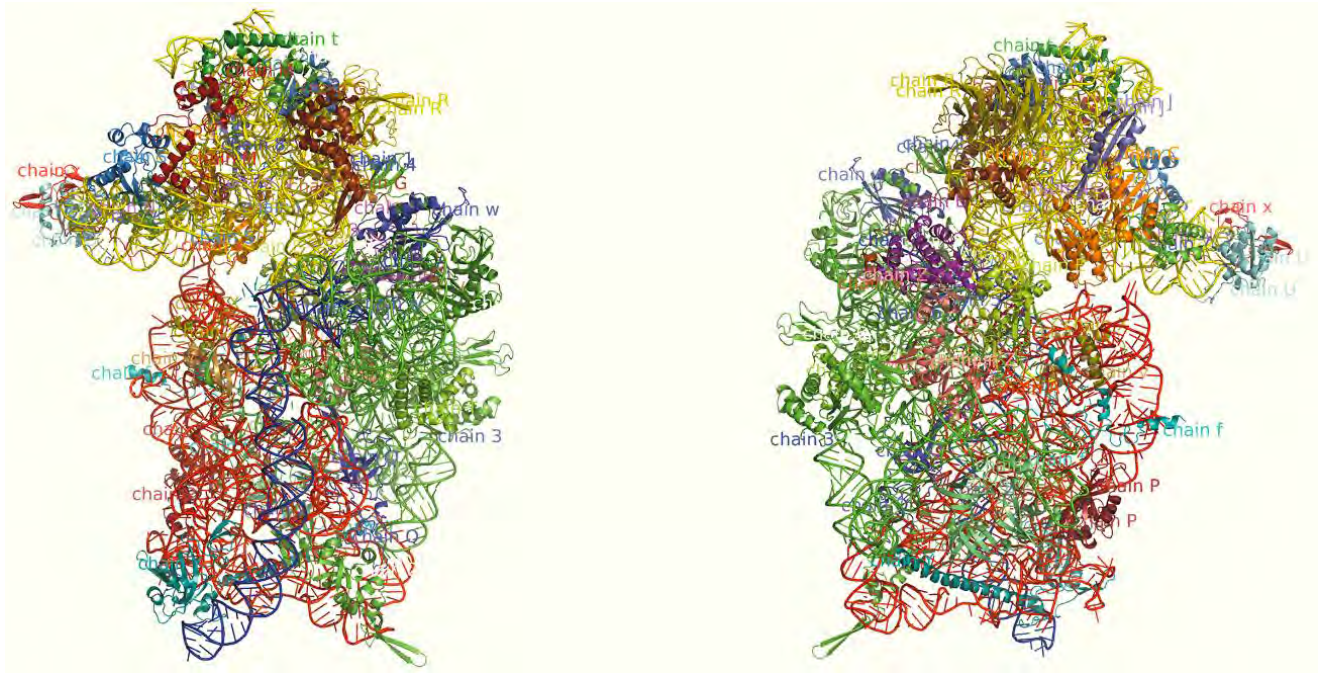
Figure 7: Tertiary structure of *Plasmodium falciparum* 40S front and back view. 18S rRNA is colored differently by the domain (5' major –red, Central-green, 3'major-yellow and 3' minor-blue). Also shown are the 40S ribosomal 34 proteins interacting with the 18S rRNA (Mwangi 2013).
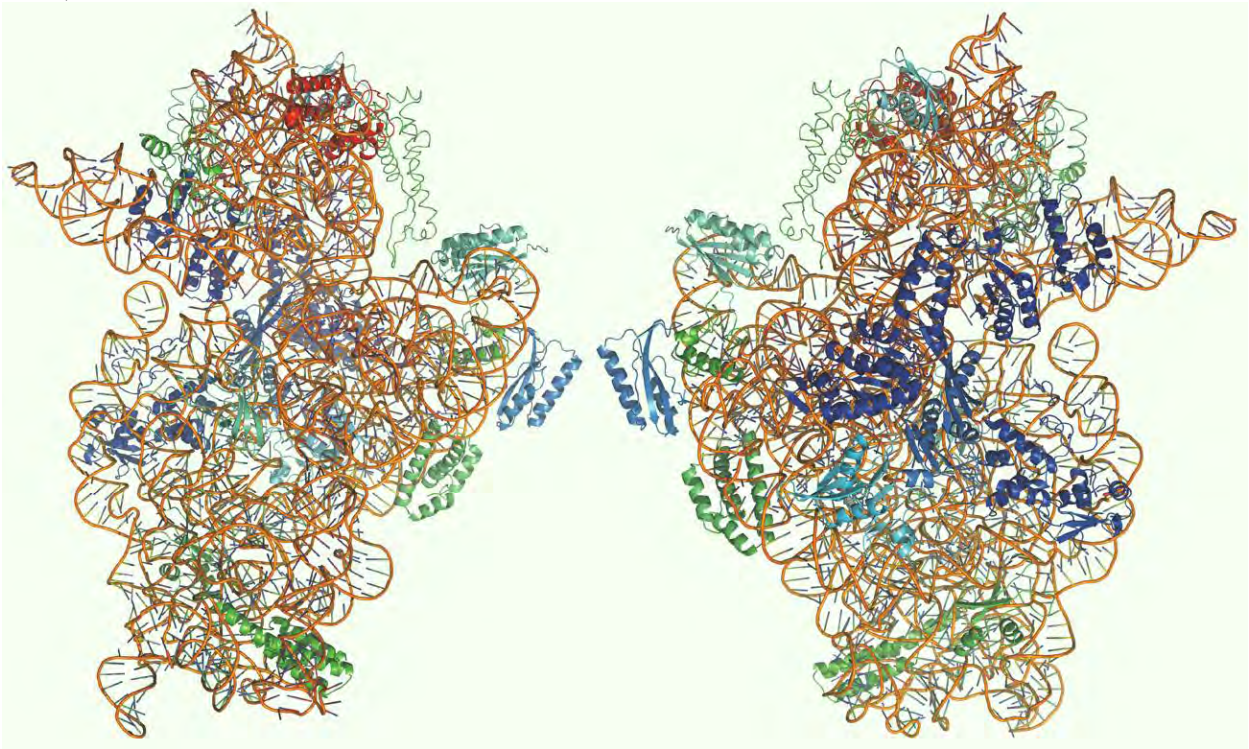
Figure 8: *Mycobacterium tuberculosis* 16S rRNA. Domains are colored differently for identification. 5' domain (Green), central domain (Yellow), 3' major domain (cyan) 3' minor domain (red) (Mathenge 2013).

2.7 Docking and Binding

Docking is the computational simulation of a potential ligand binding to a receptor site, which in our case could be the receiving molecule the ribosome or any of its content (Ruiz-Carmona, Alvarez-Garcia, et al. 2014; Irwin and Shoichet 2016). Once the binding has occurred, the orientation of the small molecule (ligand) relative to the receptors well as the conformation of the complex, when bound to each other is determined by the pose which is the candidate binding mode (Irwin and Shoichet 2016). This is then scored by evaluating particular poses by calculating the number of favorable intermolecular interactions such as the hydrophobic contacts and hydrogen bonds (Perryman, Yu et al. 2015; Irwin and Shoichet 2016). The ranking is then done by classifying the ligands that are highly likely to interact favorably to a particular receptor based on predicted free-energy required in binding itself. Finally, a docking assessment is performed to quantify the predictive capability of the whole process(Irwin and Shoichet 2016).

2.7.1 Target screening:

Discovery of compounds that act on rRNA motifs can be achieved through mechanism-based screens (Erlanson, McDowell, et al. 2004). We screened the pathogen box library against our targets using docking software tools such as Schrödinger's suite, Accelrys Discovery studio and rDOCK and then identified lead compounds (Schrödinger 2013; Ruiz-Carmona, Alvarez-Garcia, et al. 2014). This reduced our compounds to a manageable number of lead compounds. At this stage, we carried out *in vitro* screening. We then screened the selected compounds, against a labeled synthetic target in binding assays by capillary electrophoresis. The Compounds will be screened using selection and fluorescence assays to identify lead compounds with high affinity for the rRNA target, but low affinity for human ribosome (Mwangi 2013). The nature of the interaction of the hits with their targets was determined by capillary electrophoresis gel retardation binding analysis of lead-target complexes (Krull, Gendreau, et al. 2000). After a collection of initial hits is generated and the sequence information obtained, compound libraries containing RNA-binding molecules were synthesized and accessed from commercial and non-commercial sources. By using a limited set of input building blocks with structural and functional characteristics that resemble the monomers of the initial leads, the binding interactions with the RNA targets were further enhanced and optimized *in silico* using discovery studio and Schrödinger's suite.

Paromomycin bound on the A site of *M tuberculosis* h44

Streptomycin bound to two regions on the H27 OF *M tuberculosis* I6S rRNA

Kanamycin bound to two regions on the A site of *M tuberculosis* h44
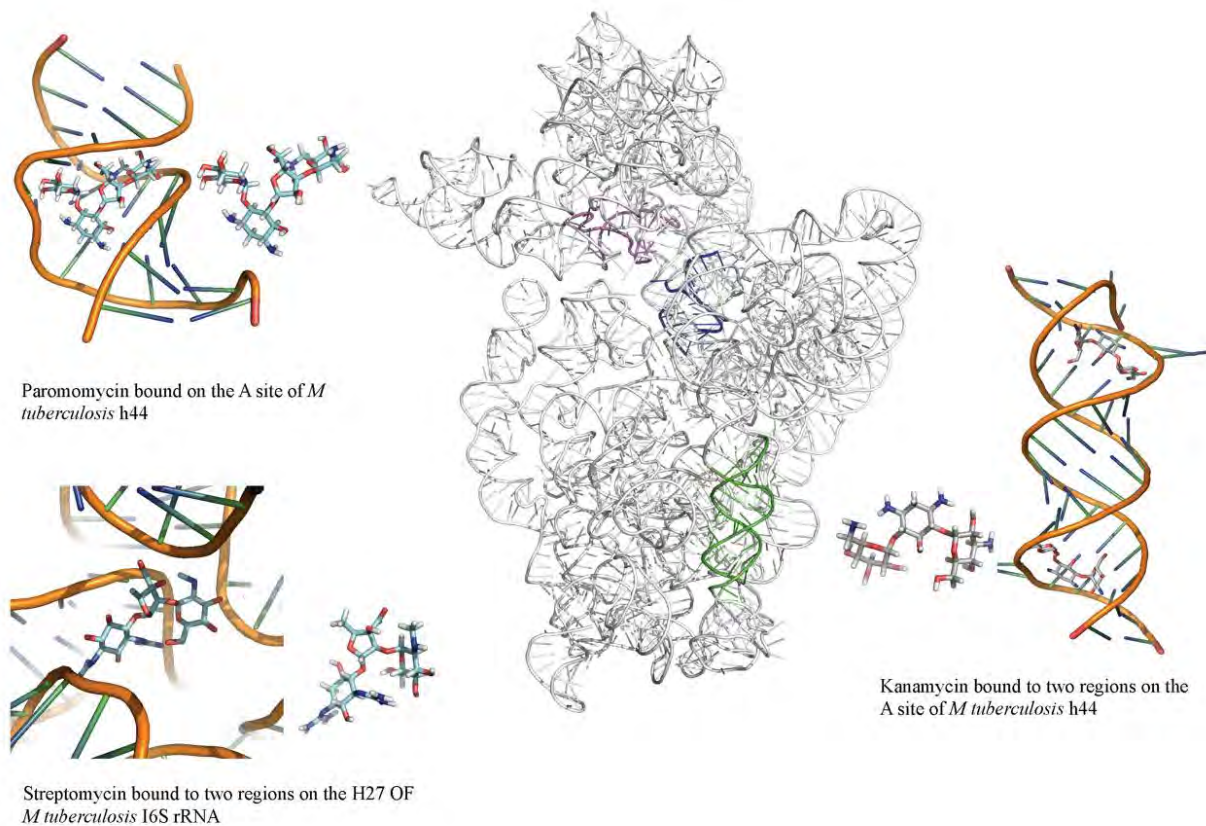
Figure 9: Graphical summary showing homology and De novo modeled *Mycobacterium tuberculosis* 16S rRNA motifs bound to antibiotics (19).

2.7.2   RNA motif synthesis.

RNA motifs identified will be synthesized chemically or cloned into a plasmid that facilitates run-off transcription (de Silanes, Zhan, et al. 2004). The method of synthesis will depend on the specific target chosen for screening, and whether it is being used for protein screening or high throughput library screening (de Silanes, Zhan, et al. 2004; Siegfried, Busan, et al. 2014).

2.7.3   Preparation and labeling of RNA targets

The target motif sequences will be designed so that the transcript will form a stable clamp sequence in the stem with an RNA tail at the 3' end. RNA targets generated by transcription will be annealed

30

with a 5' fluorescent-linked DNA or targets will be synthesized with a 5' fluorescent tag (Kramer, Weber, et al. 2007; Mulaa and Krämer 2011; Simmons, Arriza, et al. 2013).

2.8 Compound Library Screening

2.8.1 Fluorescence-based binding assay

Fluoresce in-tagged RNAs was synthesized by chemical methods from commercial sources (Blakeley, DePorter, et al. 2012). Fluorescent labels were added to the 5' end of the RNAs motif during the last step of the chemical synthesis by means of a carbon linker. Remote placement of the dye molecule at the end of the substrate RNA allows detection of drug or ligand-induced, conformational changes in the RNA (Brown, Diemer, et al. 2013). Binding of a ligand could be monitored by capillary electrophoresis binding assay or a decrease in the fluorescence emission intensity of the attached fluorophore dye (Kariuki, Ogwang, et al. 1998; Kariuki, Kiaira, et al. 1998; Mulaa and Krämer 2011). An alternative to end labeling the RNA molecule was to insert a fluorescent nucleoside analog within the RNA target (Kramer, Weber, et al. 2007; Mulaa and Krämer 2011). This would most likely increase the sensitivity in detecting a binding event. This approach was used to detect the binding of a compound to one of the designated 18S target RNAs. The addition of an internal label within the target would also allow the design of a FRET-based assay using a labeled ribosomal protein that binds to the target (Bao, Rhee, et al. 2009)., The ability of a ribosomal protein to bind the target RNA would also be an indicator of a correctly folded RNA molecule (Dorywalska, Blanchard et al. 2005).

3 RESEARCH PROBLEM

3.1 Statement of the Problem

Kinetoplastids parasites *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major* are a growing and increasingly serious public health problem. Recent World Health Organization identified *Trypanosomiasis* and *Leishmaniasis* as a neglected major cause of life-threatening disease. Moreover, Kinetoplastids has been shown to affect both animals and plants with devastating economic and health effects. Currently, the treatment of Trypanosomiasis and leishmaniasis disease has no ideal effective drugs due to the toxic to a lethal side effect. **Nifurtimox and benznidazole** the current standard drugs of care for Chagas have serious undesirable effects, which have been reported during their clinical use. Despite the many decades of use of most of the current trypanocides, we know little of their mode of action and Antibiotic resistance caused by HAT and other Apicomplexan parasites responsible for many deaths each year, and much of this mortality is due to the rise of antibiotic-resistant organisms. Because antibiotic-resistant infections double the duration of hospital stay, mortality, and morbidity as compared with drug-susceptible infections, economic costs of Kinetoplastid infections and antibiotic resistance are estimated to be in the billions of dollars. One of the major challenges in Kinetoplastid research is to leverage advances in genome sequencing into lead therapeutic modalities to treat *Trypanosomiasis* and *Leishmaniasis* infections and antibiotic resistance. This functional genomics approach holds great promise to not only advance kinetoplastid therapeutics, but to provide highly selective chemical probes of function to study disease biology. These technical challenges are compounded by a lack of well-validated drug targets as well as a poor understanding of the optimal pharmacology for the treatment of kinetoplastid infections. The overall goal of this project is to develop new anti-infectives that is highly effective and refractory to antibiotic resistance using a combination of homology modeling and microarray motifs that allows the identification of new rRNA

3.2 Overall objective

The overall goal of this project is to develop new anti-infectives that are highly effective and refractory to antibiotic resistance using a combination of homology and de novo modeling, *in silico*

screening and microarray motifs technologies that allow the identification of new rRNA target sites.

This will involve: Generate and predict target *Trypanosome and Leishmania* ribosome crystal RNA structure from sequence to be a powerful platform to design small molecules that target them. The approach we are taking for identification of *Trypanosome and Leishmania* ribosome drug targets is to utilize the RNA homology modeling software to develop a high throughput method to identify ribosome RNA motif-small molecule partners, or "paired" motifs. Namely three-Dimensional Homology Model Screening (3DHMS). We shall utilize this innovative approach to implement more broadly, a screening strategy with the goal of identifying *Trypanosome and Leishmania* ribosome drug targets. This will culminate in custom synthesis of highly selective and potent lead compounds by a disease-causing agent by using a link between digital ribosome and array screening of a disease-causing ribosome as a catalyst. That is, the disease-causing ribosome serves as a reaction vessel and disease-causing RNA motifs as a catalyst to allow for the synthesis of its own treatment. This technology will be applied to develop compounds to treat and *Kinetoplastids* that affect millions of people worldwide. We describe the application of these approaches towards kinetoplastids a major cause of life-threatening disease in populations.

3.3   Specific objectives

The specific objectives of this study are to:

1. Develop a high-resolution functional Trypanosome and Leishmania 18S rRNA homology and de novo modeled crystal structure.
2. Identify essential rRNA motifs from the Trypanosome and Leishmania ribosome crystal structure and potential targets compounds for these motifs
3. Optimize and validate hit compounds against the targeted ribosome 18S rRNA motifs.

3.4 Rationale

Determining the first structure of the Trypanosome and Leishmania 18S rRNA will lead to better understanding of the structural basis for its protein-synthesizing roles in the cell. This will enable research in the field of drug development run *In silico* ligand screening experiments using the solved Trypanosome and Leishmania 18S rRNA structure as a target against a library of potential anti-kinetoplastid compounds. The use of structure prediction and modeling technologies in this study will dramatically reduce the time it takes from target identification to drug lead determination. Additionally, millions of compounds that can be tested *In silico* against the generated structure would have been impossible to test experimentally

3.5 Innovation and Expected Output

The approaches used in this study to validate and screen rRNA targets are highly innovative. They are superior to traditional screening approaches because of the likelihood of success in finding translation inhibitors that are highly efficacious, but at the same time unlikely to select for target mutations, is very high. Its innovation is based on conserved ribosomal sequences. The exploitation of these sequences could lead to new anti-protozoan drugs refractory to resistance. Drug resistance significant or more likely the highest drug-related priority to date. The production of new anti-protozoan compounds refractory to drug resistance would be a major advance in the repertoire of NTDs. Directing this approach towards other organisms including eukaryotic infective agents could make a significant positive worldwide health impact.

This is an application exploiting the conserved regions of ribosomal rRNA that form functional motifs or loops to develop next-generation anti *Trypanosoma and Leishmania* candidate compounds that can be further developed to potent anti *Trypanosoma and Leishmania* to combat Kinetoplastids family against many protozoans that cause moderate to severe diseases and is applicable to a broad spectrum of apicomplexan protozoan's diseases. The application is unique as it proposes to use high-resolution macromolecular models that will lead to the identification of superb modeling systems to the identification of conserved functional motifs of Kinetoplastids ribosomal RNA that for functional reasons cannot be altered and therefore good targets for the screening of lead anti-Cryptosporidium compounds.

4   METHODOLOGY

4.1   Construction of 18S rRNA Crystalline Structures of the Three Kinetoplastids through Homology and De novo modeling

4.1.1   Obtaining and Verifying Query Sequence

The three kinetoplastids; *Leishmania major, Trypanosoma brucei* and *Trypanosoma cruzi* 18S rRNA sequences were obtained through blasting in the gene bank (NCBI) (Coordinators 2016). An important thing to note is that most of the sequences in the gene bank are not complete so a further process of verification was required. We checked the completeness of the sequences and this further led us to specialized groups that do verification of 18S rRNA sequences. Such a group is The Comparative RNA Web (CRW) Site that has a database that shows the completeness of sequences among other analyzed and verified annotation (Cannone, Subramanian, et al. 2002). The sequences picked from this site were further analyzed to ascertain the sequences and minimize the errors. Back to the gene bank the FASTA format of these refined sequences was picked and saved as text documents using a notepad++ text editor (https://notepad-plus-plus.org/). There is a possibility of having more than one complete sequence hence to get the final query sequence that was modeled we further carried out multiple sequence alignment to find which one deviates minimally from the consensus sequence. MultiAlign (Corpet 1988) software was used in alignment to show how similar or dissimilar various sequences were.

4.1.2   Selecting a Template

Selection of template structures for the three kinetoplastid rRNA was a rigorous exercise that is described below in section 4.1.2.1. These required a search of various structure libraries using the query sequence.

4.1.2.1 Obtaining and Verifying Template Sequences and three Dimensional coordinate files

The templates for all the three kinetoplastids stated in section 1.1.1 were again selected through an elaborate process that involved several steps. Firstly by blasting individual query sequences in the gene bank (NCBI) and finding the sequences that are highly similar to the query and not in any way the query by (Do not select the query if it is shown in the BLAST alignments) (Altschul, Gish, et al. 1990). These sequences infer and evolutionary relationship with the query but not specifically

the query if they cover a distinct region of the template to get a higher similarity score. Calculation of a local pair-wise alignment between the templates and their targets was performed. Secondly, checks and evaluations were done as the section above to check if these sequences were complete followed by a heuristic step which intended to improve the alignment. Insertions and deletions placement in the template was considered for optimization. Of particular interests are the isolated residues in the alignment (Islands) which were moved to the flanks for the facilitation of loop building. The next step was to find out if there was any crystal structure of the 18S rRNA template sequences that were available. This was done by checking the Research Collaborators for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) archive information that gives the 3D shapes of nucleic acids, proteins, and complex assemblies that helps researchers and students understand all aspects of biology (Berman, Henrick et al. 2006). Coordinate files of the template structures of the 18S rRNAs obtained from the PDB website available online (http://www.rscb.org/pdb/explore.do?) were saved as PDB files on a text editor. Depending on the complexity of the rRNA homology and *de novo* modeling was done in parts by dividing it into the different parts after cleaning: 5'major, central, 3'minor and 3'major domains. An important point noted was while splitting cut the rRNA of both the template and the query at similar points to achieve the best structure at the end. The structure was viewed in different software available that can read PDB file such as pymol and Accelrys among others. The crystal structures of the templates obtained had some challenges such as unresolved portions, gaps and require to be further optimized.

4.1.3   Homology and De novo Modeling

18S rRNA homology and *de novo* modeling were done using RNA123 version 2.0.1.3 and Genesilico software. RNA123 was able to predict the secondary and tertiary structure of the three kinetoplastids ribosomal RNA. RNA123 took three steps; Preprocessing, Alignment, and Modeling.

4.1.3.1 *Preprocessing* step was aimed at analyzing the template structure by checking and removing bad contacts, and residue clashes. This step takes time depending on the length and condition of the initial structure finally giving an optimized template structure. High energy structures optimize much slower that low energy structures overall. Once preprocessing was complete two output files in the working folder appeared one optimized template as a .opt.pdb and the other as an optimized secondary structure as opt.2D. PDB coordinate file can be analyzed for base pair type and classification once loaded in RNA123. This shows base pairing patterns of all the detected base pairs both tertiary and secondary interaction. Secondary interaction based on non-Watson-Crick, Watson-Crick paired bases and also list distorted pairs (Watson and Crick 1953). Tertiary interactions will show tertiary stacking interactions, pseudoknots, and other tertiary hydrogen bonds. The nomenclature of Leontis and Westhof is used to classify and annotate base pairs by types (Leontis and Westhof 2001). More information on the structure can be viewed by clicking on the 2D INFO window that may describe the inter-atomic distances and angles.

4.1.3.2 Alignment is performed between the now optimized template structure and the query. Within RNA123 a characteristic alignment known as structure-based sequence alignment (SBSA) is performed. These enable the generation of an optimal alignment that must be inspected for consistency before beginning the modeling. The software can generate more than one alignment if need be depending on user preference. A series of algorithm performs scoring where substitution matrices account for divergent evolutionary nucleotides. This takes care of gap opening, gap extensions, and gap terminals by penalizing them wherever they occur in the alignment. Opening a gap is very unfavorable compared to extending the initial gap, and the gap terminal controls the penalty at both ends of either template or query. The Structure-based sequence alignment score is arrived at using the equation shown below indicating the template sequence score compared to the query score.

$$\%SBSA\ Score = \frac{Template\ Vs.\ Query\ score}{Template\ Vs.\ template\ score} \times 100$$

To perform homology modeling the SBSA score should be above 60%, which means that there is conservation in base pairing, rather than nucleotide identity. Alignments can be edited before the homology process begins where visual inspection for alignment accuracy by paying attention to the secondary structure elements in both the query and template to ensure correct alignment.

4.1.3.3 Modeling is the next step that begins after a good alignment has been achieved. During modeling, all the required motifs to build the query tertiary structure have to be present. This motifs information is found in a database and checked against the template structure for presence. Lack of important motifs means that area must be *de novo* modeled (Shown in Figure 1). During the process of modeling, all the distant and hydrogen restrains are applied. This impacts the quality of the modeled structure. Concurrently as the modeling is happening a log file is created to save the data. To generate parts that could be inserted or deleted within the rRNA an ensemble of fragments compatible with the neighboring stems was constructed using the *de novo* algorithm. This was performed using a scoring scheme based on the favorable force field energy, steric hindrance and favorable interactions like hydrogen bond formation to select the best loop. Modeling the 18S rRNA for the kinetoplastids selected was done in sections known as domains because of its complexity. These are the 5'domain, central domain, 3'major domain and finally the 3'minor domain. This was a better approach than total 18S rRNA modeling as it allowed easier manipulation of alignments crucial for viable structure generation.

4.1.3.4 Energy minimization is performed to adjust deviations in the rRNA structure geometry. This is introduced in the modeling algorithm at the last modeling stage to regularize rigid fragment when joining. RNA123 uses the empirical force field to detect model parts with conformational errors. Neither does molecular dynamics or energy minimization improve the model accuracy all they do is to regularize the structure (Mathews and Turner 2006, Das and Baker 2007). Recently however, restricted molecular dynamics has been applied to improve the quality of homology models (Flohil, Vriend, et al. 2002). RNA123 has an integrated algorithm structure that runs this process with the help of a verified database for known motifs force fields (Sijenyi, Saro, et al. 2012) as shown by Figure 10 below.

Structure validation was performed once the model was generated and energy minimized. Several validation tools were used such as SimRNA which refined the model structure further (Magnus, Boniecki, et al. 2016). Discovery studio Accelrys (Studio 2013) software help view the structure, add hydrogen's and measure distances between molecules among other functionalities. RNA123 performed geometric refinement within itself (Sijenyi, Saro, et al. 2012) together with [ModeRNA](#) (Rother, Rother, et al. 2011). Here the residues are optimized according to the ideal atom distances. Once modeling and all the other processes are complete a new PDB file was generated as <Template_name>_aln1.pdb in the output folder for all the three kinetoplastid 18S

rRNA. Three-dimensional de novo models of RNAs of sizes up to 300 nucleotides, given secondary structure and experimental constraints, can be obtained. It can be carried out reasonably automatically, but human curation of submodels along the build-up path may improve accuracy.
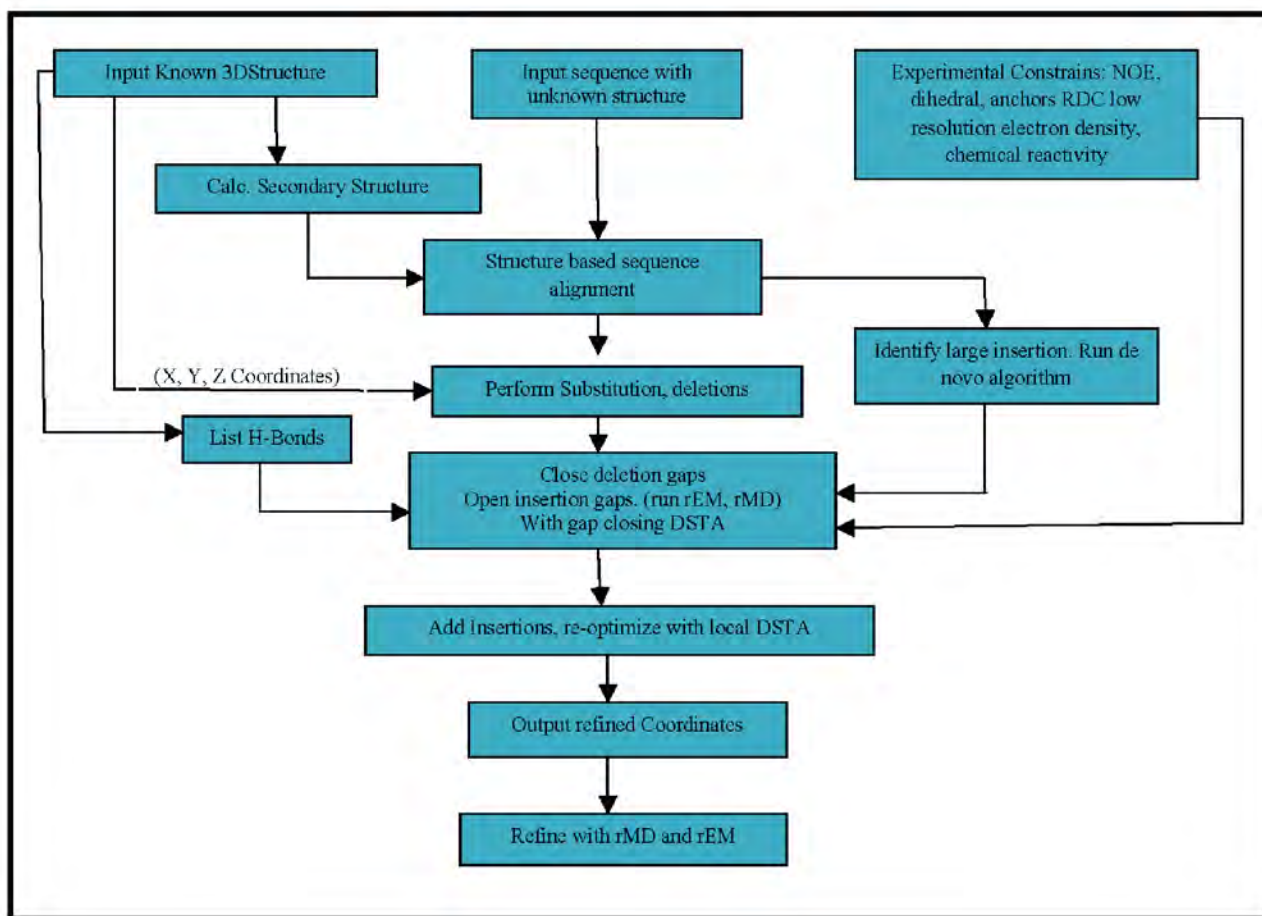
Figure 10 Diagrammatic Scheme showing the steps involved in the process of RNA123 modeling (Mwangi, Wagacha, et al. 2017).

4.1.3.5 Refinement of structures using Molecular Dynamics simulations The structures obtained was further run through Molecular Dynamics simulations and refinement for refinement and equilibration. A program that uses graphical processing units (GPUs) known as AceMD[50] was used to perform the equilibrium MD simulations (Harvey, Giupponi, et al. 2009). The expected speed of simulation was 30 nanoseconds per day, for system taking ~50,000 atoms. This warranted that to achieve a ready steady-state simulation for all the 18S rRNA kinetoplastid crystalline structures several hundreds of nanoseconds were required. Further simulations were carried out using NAMD (Nelson, Humphrey, et al. 1996), which performed temperature replica exchange molecular dynamics simulations. A simulation that used NAMD was conducted at several temperature values but limited in terms of simulation options. Only the 300K physiological simulation was used for structure evaluation. Other solvent representatives were used to mimicking the physiological response of the modeled structures.
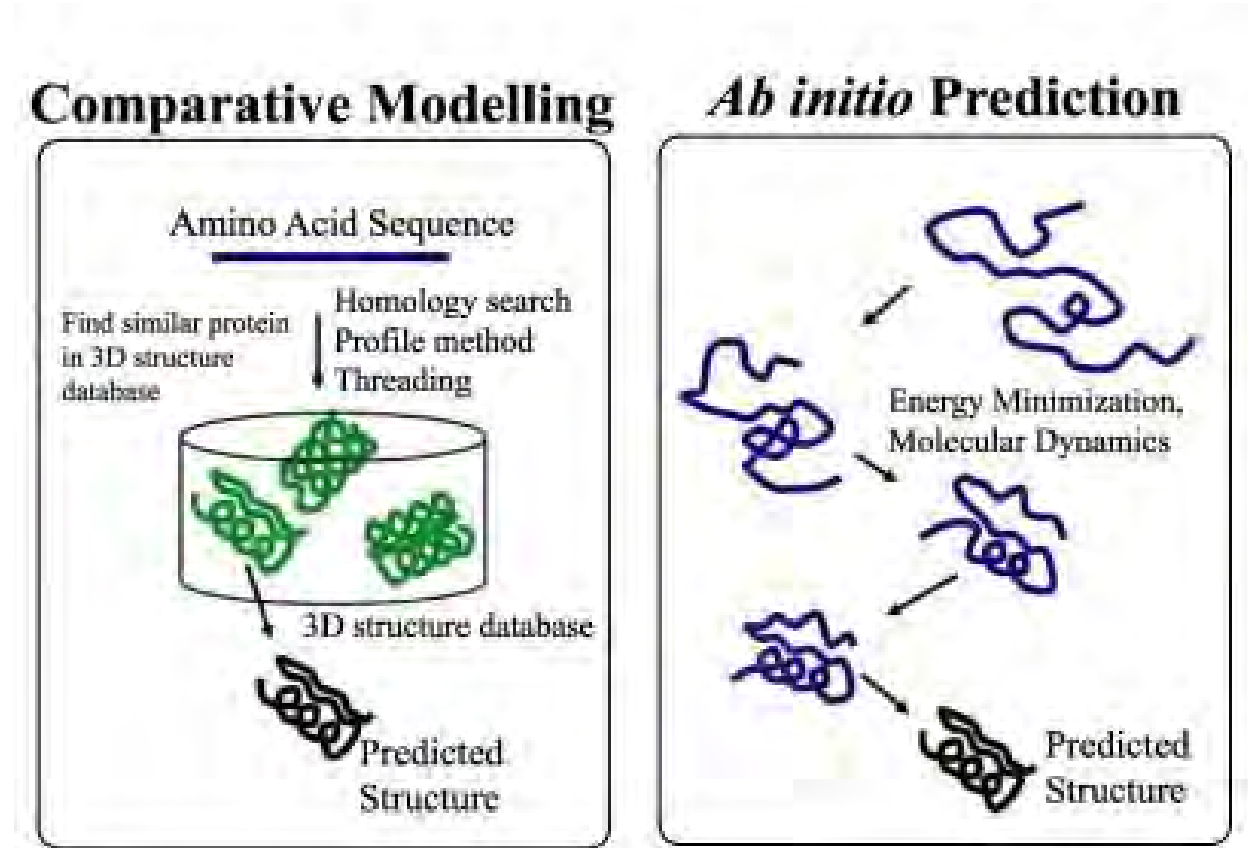
Figure 11: Illustrations showing the various processes of homology and *de novo* modeling

4.2   Target selection and refinement

To identify areas that could be used as drug binding sites for all the three kinetoplastids 18SrRNA, we had to align the three using their sequences and both their secondary and tertiary structures. The reason for this is to be able to have a common area that could be used as a target position for all the kinetoplastids. One RNA subdomain for all the three kinetoplastids was selected as a prioritized target from the four potential targets chosen from the RNA "regions of interest" identified.

We used CGT to identify every mutation of the target that could lead to drug resistance, and used homology and *de novo* modeling to determine the essential structural components of the target.

4.3   Target screening: Docking and Binding Methodology

To screen the compounds, compound and peptide libraries were made against the modeled wild type target and its viable mutants. Docking is the computational simulation of a potential ligand binding to a receptor site, which in our case could be the receiving molecule the ribosome (18S rRNA) or any of its content.

Once the binding has occurred the orientation of the small molecule (ligand) relative to the receptors well as the conformation of the complex when bound to each other was determined by the pose which was the candidate binding mode.

This was then scored by evaluating particular poses by calculating the number of favorable intermolecular interactions such as the hydrophobic contacts and hydrogen bonds.

The ranking was then done by classifying the ligands that were highly likely to interact favorably to a particular receptor based on predicted free-energy required in binding itself.

Finally, a docking assessment was performed to quantify the predictive capability of the whole process. See Figure 12 below
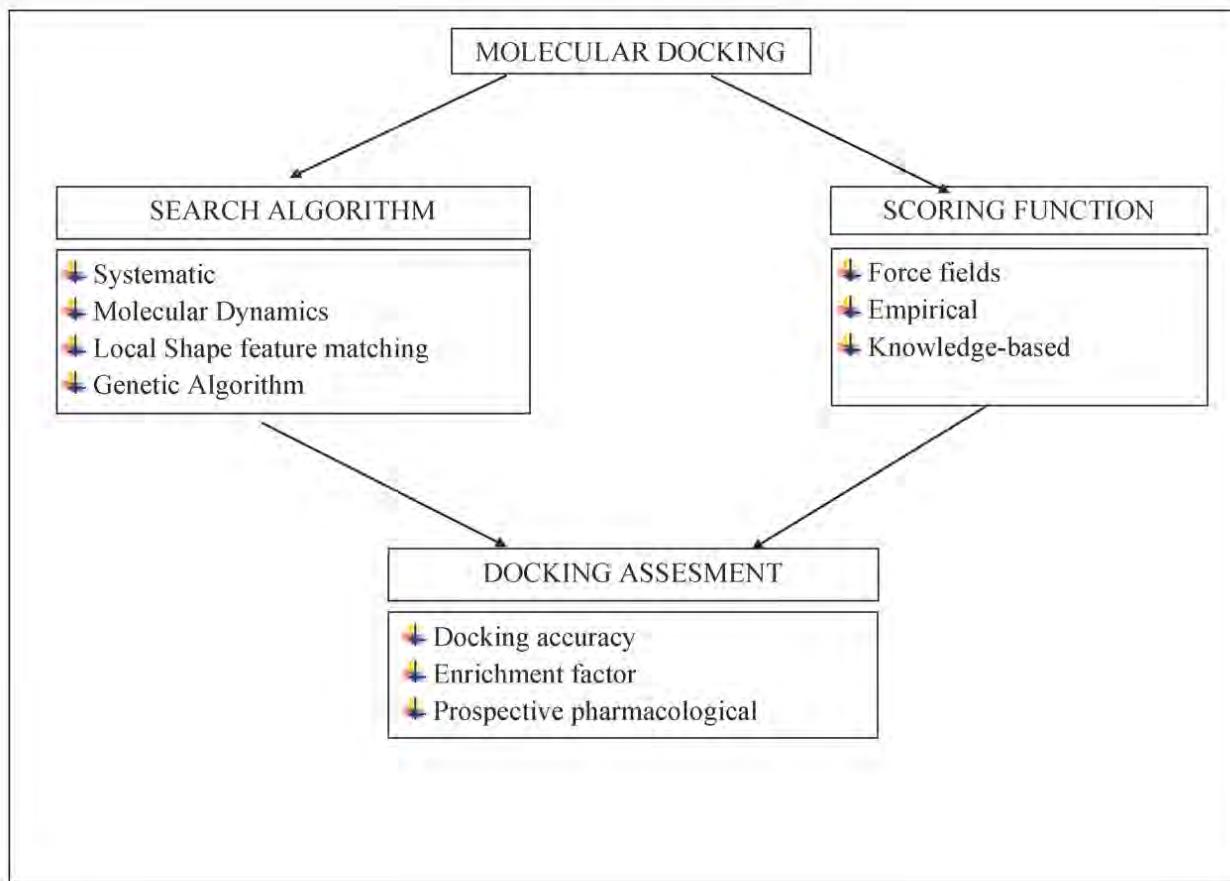
Figure 12: Docking and binding flow chart showing the different processes of *in silico* screening

4.3.1 Toxicity Prediction

The prediction of compound toxicities is an important part of the drug design development process. Computational toxicity estimations are not only faster than the determination of toxic doses in animals but can also help to reduce the amount of animal experiments.

Toxicity prediction can be done with ProTox server: http://tox.charite.de/protox_II/

(1) Click on the url address above

(2) Identify canonical SMILES of compound (using Drugbank/PubChem) or draw 2D structure of the drug

(3) Paste in SMILES (See picture below)

(4) Select the toxicity target of interest that you would like to analyze.

4.3.2 Predictive Pharmacokinetic Analysis

1. Go to http://www.swissadme.ch/

2. Identify canonical SMILES of compound (using Drugbank/PubChem) or identify 2D structure if newly designed drug

3. Paste in SMILES or draw structure in the space provided (See picture below) and press enter

4. If information is correctly entered, RUN! button will show red.

5. Click on the RUN! button.

6. Once the molecule has been put through the algorithm, the physiochemical and pharmacokinetic information may be extracted (data seen below).
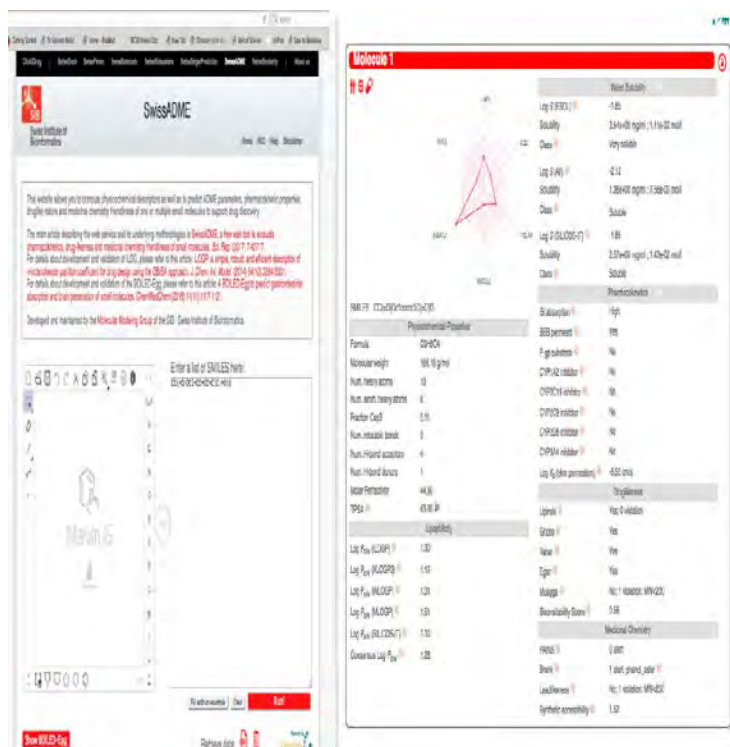
Figure 13 Physiochemical and pharmacokinetic information of a specific compound as seen in swissADME

43

4.3.3 Identifying Potential Targets of a Compound

This computer-aided drug design technique may take a few minutes at no cost, compared to protein analysis of each enzyme within a metabolic pathway. With this being said, it is always important to validate these predictive techniques with experimental analysis.

The PASSONLINE software predicts both the physiological target and biological activities of a compound. The key principle of this technique is that similar proteins would bind structurally similar molecules.

The PASSONLINE software identifies and rates potential biological targets based on the probability of action (Pa) and probability of action (Pi) of the compound to the target. The target may then be accessed from virtual libraries, or modelled, and further computational analysis on the molecule and target may be carried out.

Procedure:

1. Go to http://www.pharmaexpert.ru/passonline

2. Register to use the server and logging

3. Identify canonical SMILES of compound (using Drugbank/PubChem) or identify 2D structure if newly designed drug

4. Paste in SMILES (See picture below) and press enter

5. Select the target species that you would like to analyze.

6. If information is correctly entered, "target predict" button will show red.

7. Click on the "target predict`' button.

8. Once the molecule has been put through the algorithm, the biological targets, categorized by probability, will be demonstrated.
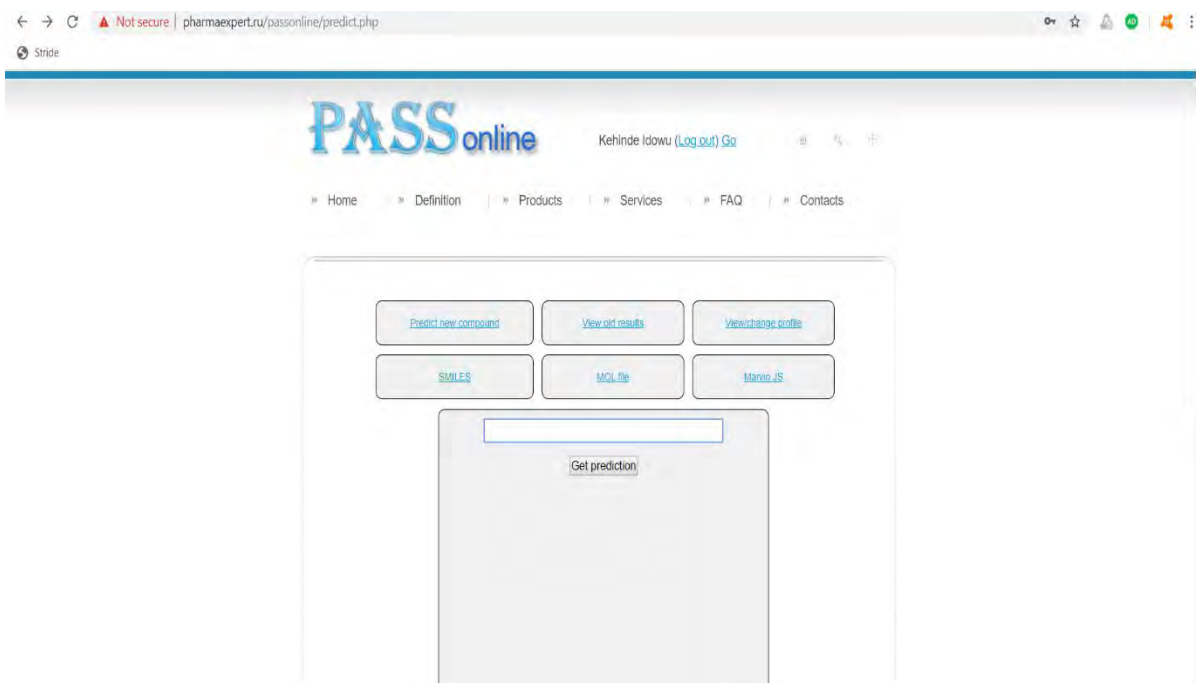
Figure 14 The PASS online tool As it appears once the link is open

Paste the SMILE of Paracetamol, the result below will be gotten

Figure 15 The PASS online tool As it appears once the link is open and the data about the compound is inserted and results presented

4.3.4 Molecular Docking

There are two main steps involved in docking:

1. Sampling conformations of a ligand in the active site of protein- different algorithms may be used when sampling the numerous conformations of the docked complex: the "lock and key" model which describes the ligand and receptor as rigid structures, or the ligand may be flexible either through random or simulation-based methods. The latter algorithm is the most commonly used method as it allows for a more realistic fit of the ligand to the protein.

2. Ranking the different conformations by scoring function- the scoring function may be based on statistically preferred contacts, MM force fields or pre-existing protein-ligand binding affinities.

Over the past decade, there has been flood of molecular docking related publications and although these papers may add to the structural information about a biological target or new lead compound, there are still many inconsistencies that arise. Frequent criticism associated with docking includes incorrect binding sites, choice of docked complex (conformational pose) and choice of small molecule (inhibitor or agonist).

Due to these concerns, all docked complexes should be verified with MD simulations or experimental analysis.

There are many different tools that can be used for molecular docking with popular examples being Autodock, Shrodinger Maestro with Glide, GOLD, discovery studio among others. We use also free online tool and provide knowledge on software that is freely available, the Patchdock server and Chimera with the Autodock plugin kindly see a detailed example in the appendices section.

4.4 Validation of targets and compounds

We performed structural studies of target/hit complexes, optimized hit compounds, and validated the target using *in vitro* assays. One or more compounds that were potential leads and target validated.

5   RESULTS

5.1   Sequence selection and analysis of the 18S rRNA of the selected kinetoplastids *Leishmania Major*, *Trypanosoma brucei,* and *Trypanosoma cruzi*

5.2   Introduction

The study described in the current paragraph has compared the differential status of selected kinetoplastids (*Leishmania Major*, *Trypanosoma brucei,* and *Trypanosoma cruzi*) 18S rRNA sequences from available selected databases. The sequences were selected based on the completeness of each of the 18S sequenced genes deposited in verified databases. The three best-ranked sequence based not only with the completeness but also verified by other peer-reviewed researchers databases were selected for alignments to see the deviation in terms of genetic variation. This was performed for all the three kinetoplastids selected in this study.

5.2.1   *Leishmania Major Sequence* Selection and alignment

Three sequences deposited in the Gene bank with the defined criteria were selected and analyzed for their differences by aligning them to understand if there were very drastic differences within the sequences.

Row #	Organism (2)	L(3)	RT(4)	RC	Size	Cmp	Acc	Common Name	Phylogeny[M] (1)
1	Leishmania major	N	R	16S	882	40	M81427	kinetoplasts	cellular organisms ...»
2	Leishmania major	N	R	16S	2137	97	X53915	kinetoplasts	cellular organisms ...»
3	Leishmania major	N	R	16S	2203	100	AC005806	kinetoplasts	cellular organisms ...»

Table 3: Showing the three selected *leishmania major* spp sequences selected L(3) means the cell location which is nucleus, RT(4)means RNA type R = ribosomal RNA (rRNA), RC=RNA Class 16S, Nucleotide size, Cmp means % Complete, Acc means gene bank accession number, common name and the Phylogenetic Classification, m.
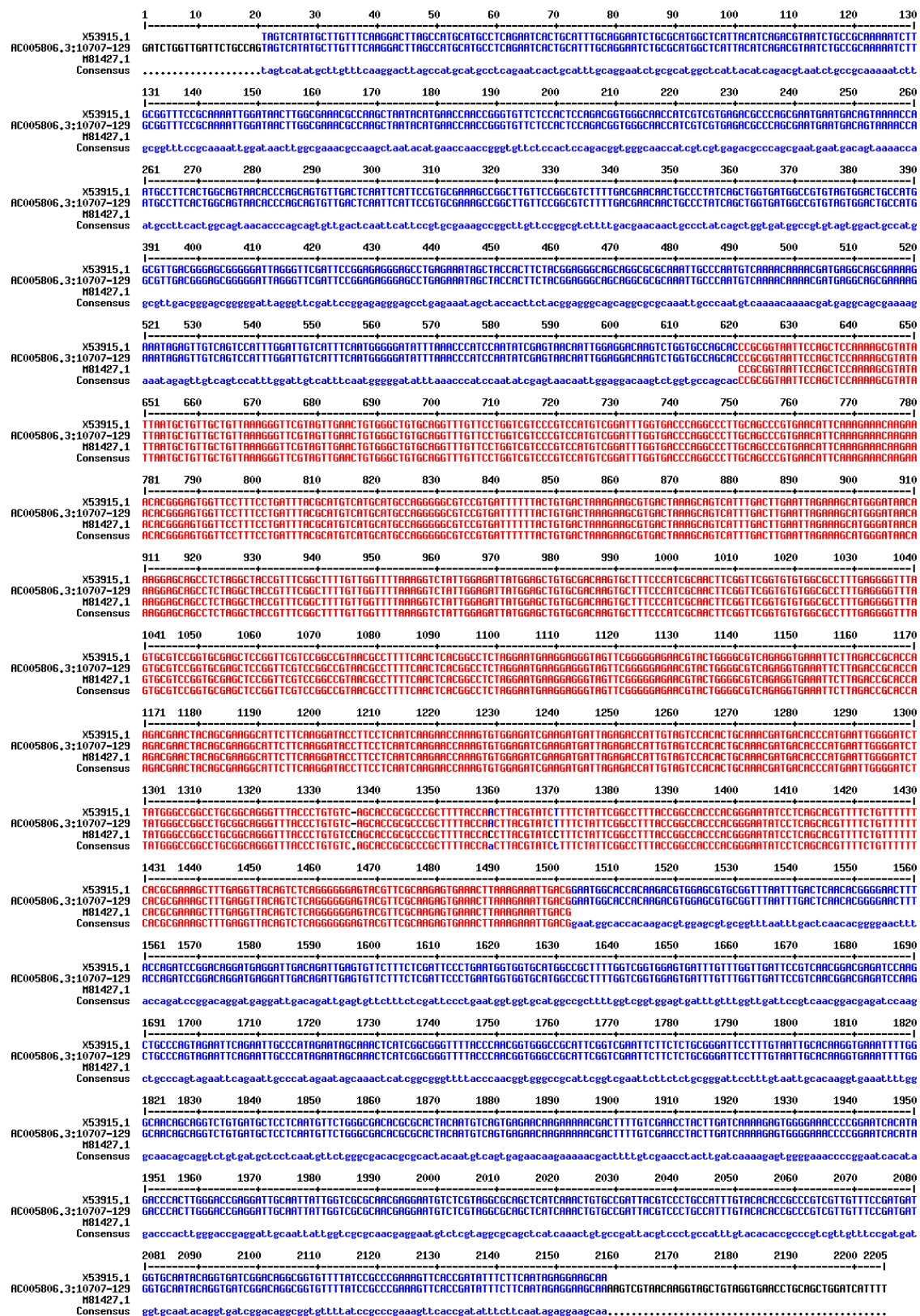
Figure 16 Alignments for the three top most complete gene sequences of leishmania major 18S rRNA together with the consensus sequences *accession number* **X53915**, **AC005806**, and **M81427**

48

5.2.2 *Trypanosoma brucei and Trypanosoma cruzi Sequence* **Selection and alignment**

Three sequences deposited in the Gene bank with the defined criteria were selected and analyzed for their differences by aligning them to be able to understand if there were very drastic differences within the sequences.

Row #	Organism (2)	L(3)	RT(4)	RC	Size	Cmp	Acc	Common Name	Phylogeny[M] (1)
1	Trypanosoma brucei	N	R	16S	2251	100	M12676	kinetoplasts	cellular organisms …»
2	Trypanosoma brucei	N	R	16S	2188	97	AJ009142	kinetoplasts	cellular organisms …»
3	Trypanosoma brucei	N	R	16S	2188	97	AC005806	kinetoplasts	cellular organisms …»
4	Trypanosoma brucei	N	R	16S	2252	100	AL929605	kinetoplasts	cellular organisms …»

Table 4: Showing the three selected *Trypanosoma brucei* spp sequences selected L(3) means the cell location which is nucleus, RT(4)means RNA type R = ribosomal RNA (rRNA), RC=RNA Class 16S, Nucleotide size, Cmp means % Complete, Acc means gene bank accession number, common name and the Phylogenetic Classification, m.

Row #	Organism (2)	L(3)	RT(4)	RC	Size	Cmp	Acc	Common Name	Phylogeny[M] (1)
1	Trypanosoma cruzi	N	R	16S	2234	97	AF228685	kinetoplasts	cellular organisms …»
2	Trypanosoma cruzi	N	R	16S	2240	97	X53917	kinetoplasts	cellular organisms …»
3	Trypanosoma cruzi	N	R	16S	2240	97	AF245383	kinetoplasts	cellular organisms …»
4	Trypanosoma cruzi	N	R	16S	2246	97	AF239980	kinetoplasts	cellular organisms …»
5	Trypanosoma cruzi	N	R	16S	2253	97	AF239981	kinetoplasts	cellular organisms …»
6	Trypanosoma cruzi	N	R	16S	2246	97	AF245381	kinetoplasts	cellular organisms …»
7	Trypanosoma cruzi	N	R	16S	2248	97	AF245380	kinetoplasts	cellular organisms …»
8	Trypanosoma cruzi	N	R	16S	2315	100	AF245382	kinetoplasts	cellular organisms …»
9	Trypanosoma cruzi	N	R	16S	2319	100	M31432	kinetoplasts	cellular organisms …»

Table 5: Showing the three selected *Trypanosoma cruzi* spp sequences selected L(3) means the cell location which is nucleus, RT(4)means RNA type R = ribosomal RNA (rRNA), RC=RNA Class 16S, Nucleotide size, Cmp means % Complete, Acc means gene bank accession number, common name and the Phylogenetic Classification, m.
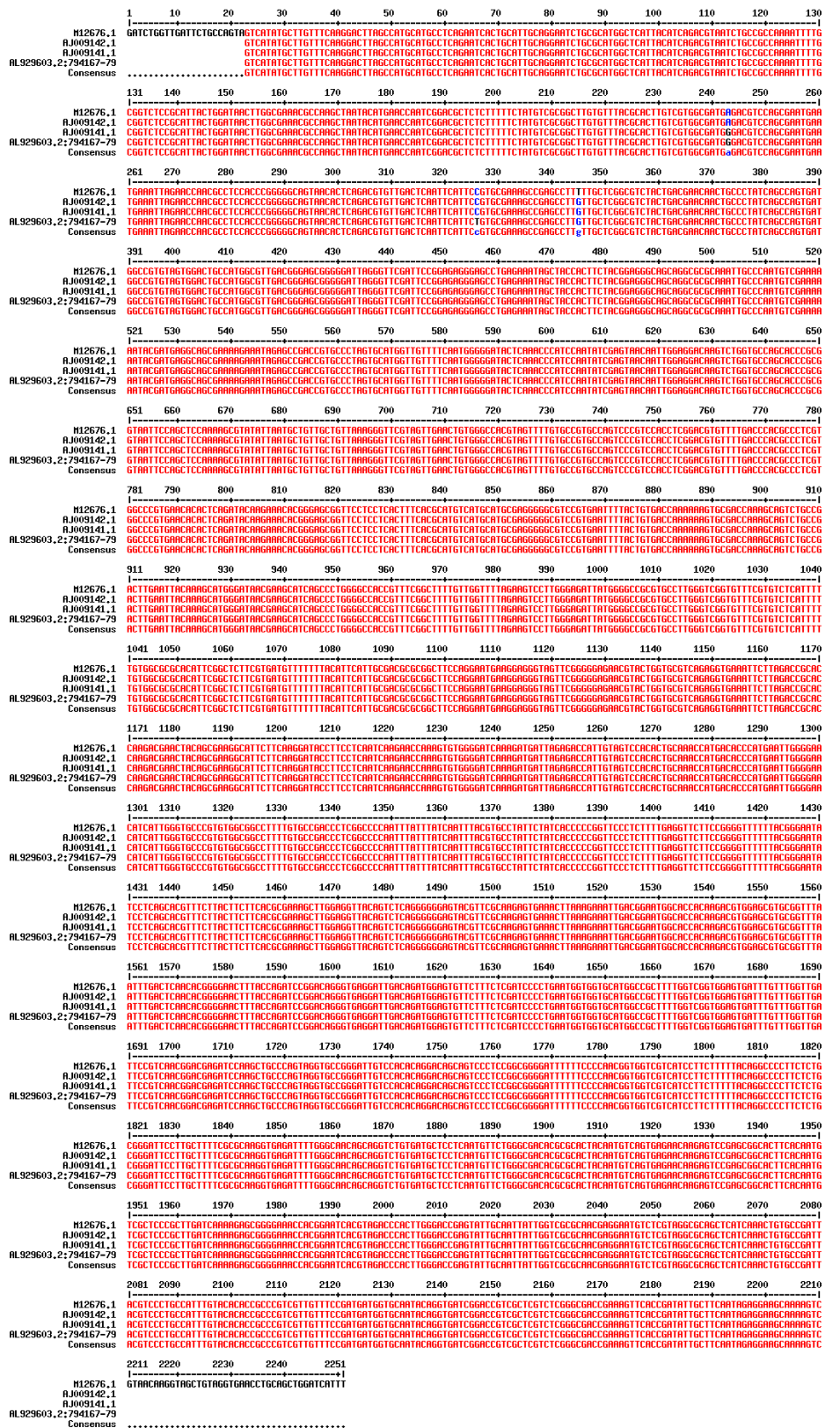
Figure 17 Alignments for the three top most complete gene sequences of *Trypanosoma brucei* 18S rRNA together with the consensus sequences *accession number* **M12676**, **AJ009142**, **AC005806** and **AL929605**
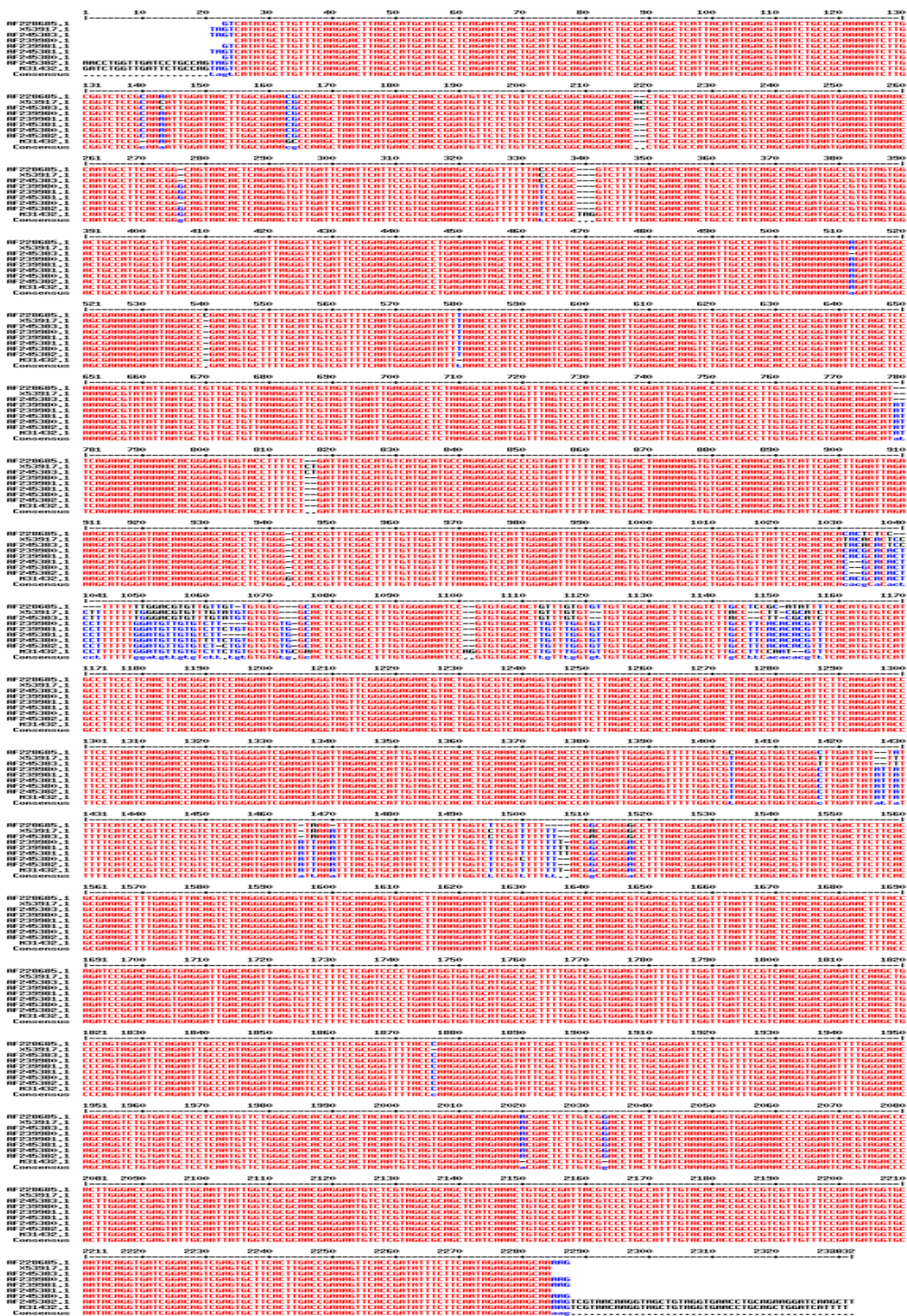
Figure 18 Figure 19 Alignments for the three top most complete gene sequences of *Trypanosoma cruzi* 18S rRNA together with the consensus sequences *accession number* **AF228685**, **X53917**, **AF245383**, **AF239980**, **AF239981**, **AF245381**, **AF245380**, **AF245382**, and **M31432**

51

5.3   18S rRNA Secondary structure
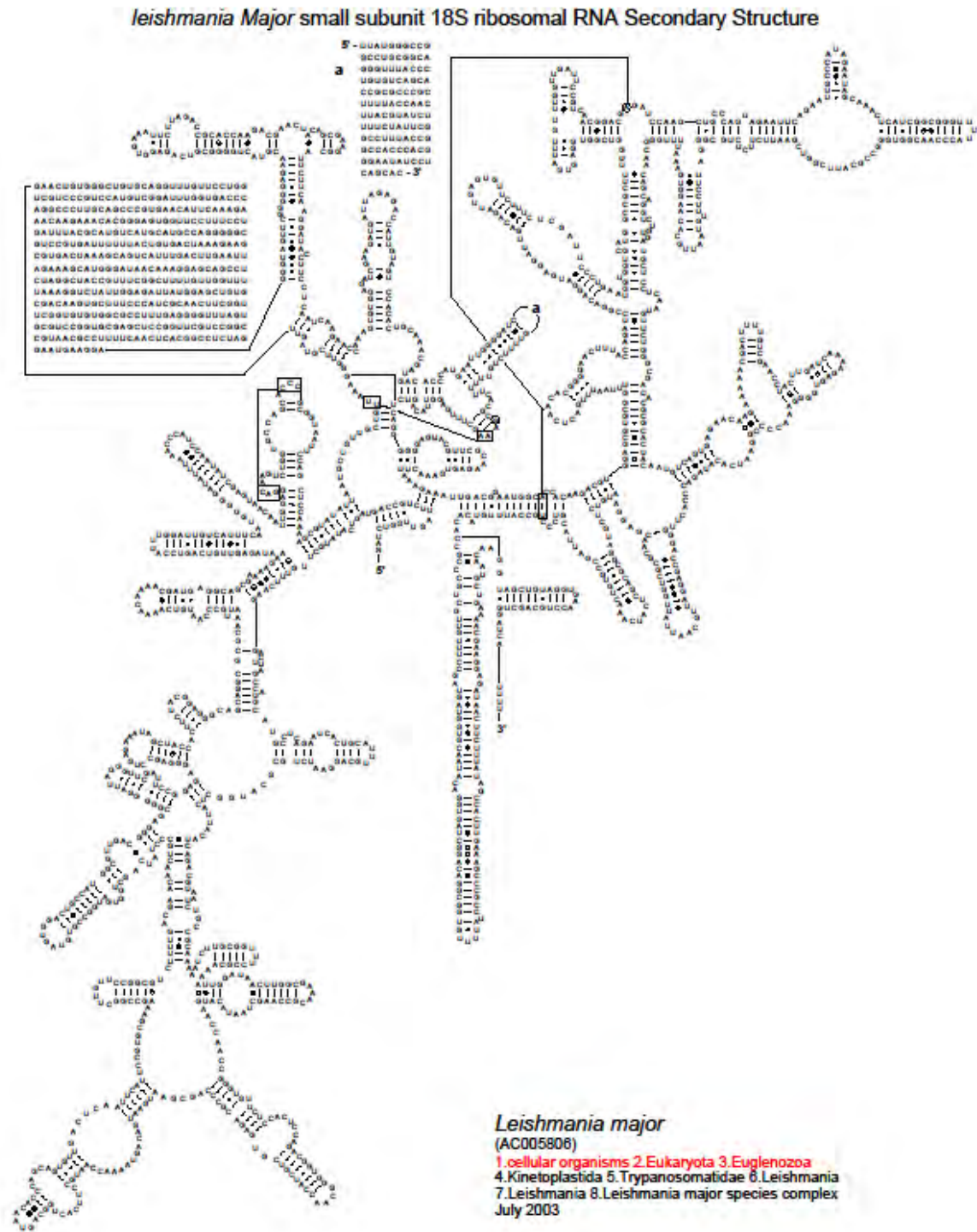
Figure 20 Secondary structure of 18S rRNA of *leishmania major* selected for the three-dimensional modeling after alignment and analysis for completeness
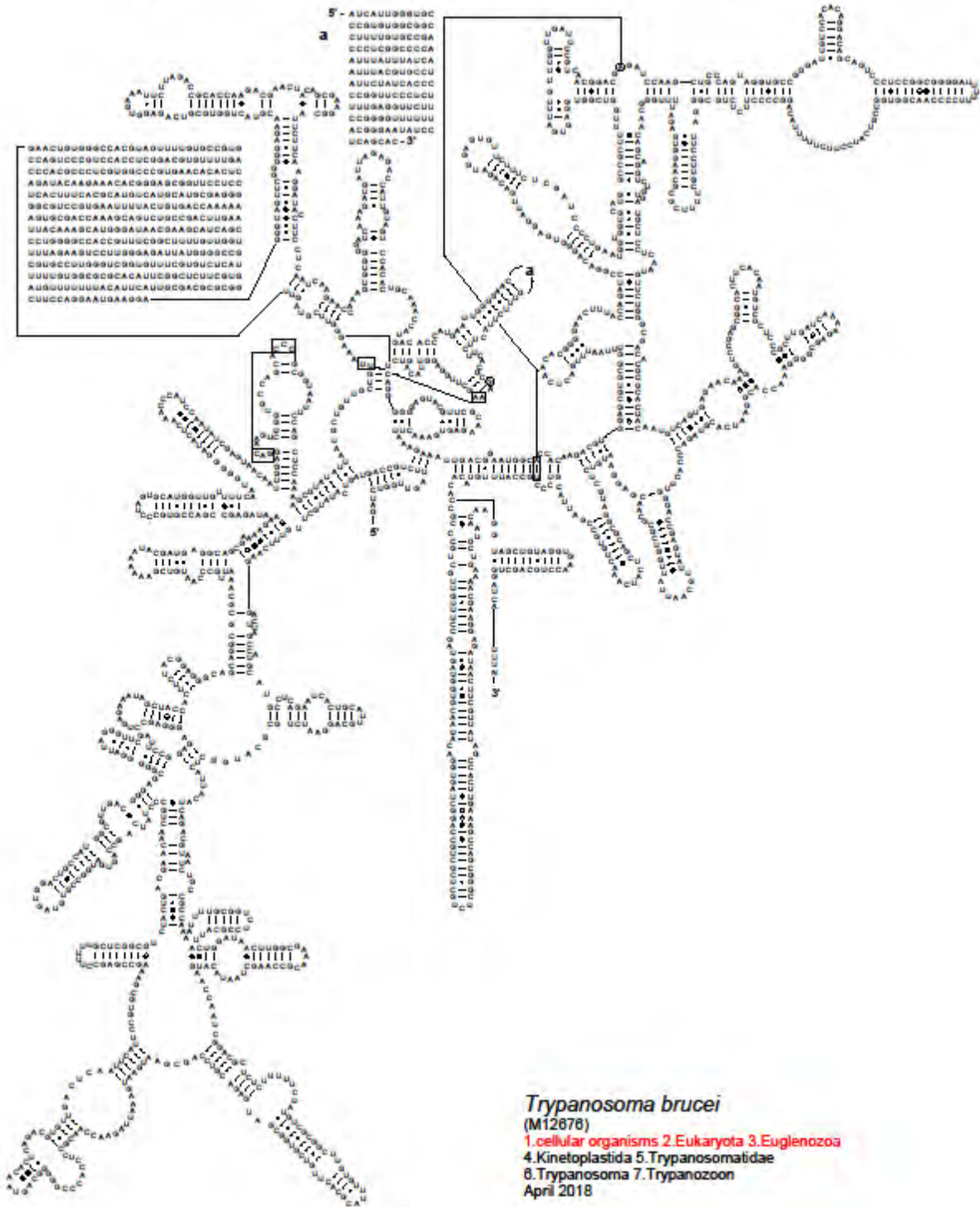
Figure 21 Secondary structure of 18S rRNA of *Trypanosoma brucei (Ac M12676)* selected for the three-dimensional modeling after alignment and analysis for completeness

Figure 22 Secondary structure of 18S rRNA of *Trypanosoma cruzi* (Ac AF245382) selected for the three-dimensional modeling after alignment and analysis for completeness

5.4 Three Dimensional structures of the modeled kinetoplastids

5.4.1 *Leishmania Major* 18S rRNA Three dimensional image
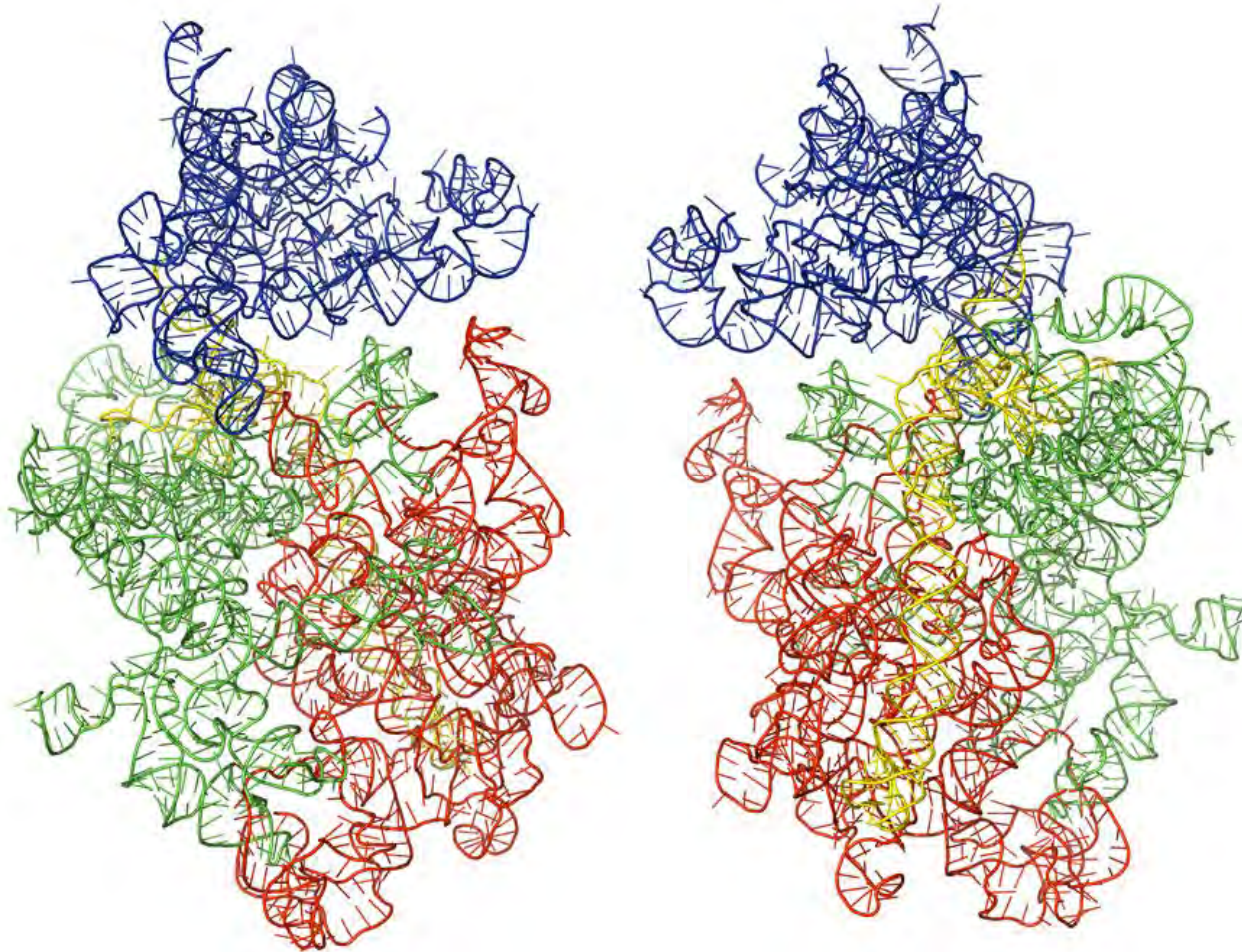
Figure 23 Architectural tertiary structure of *Leishmania major* 18S rRNA front and back view. Shown is the 18S rRNA colored differently depending with domains (5'major –red, Central-green, 3'major-blue, and 3'minor-yellow)

5.4.1.1 *Leishmania major***18S** rRNA Energy Minimization

Name	18SrRNA.std.egy	18SrRNA.opt.egy
Total Inter energy	908487.31823	-83582.63882
Total intra energy(-Gamma en	-17864.6686	-17828.1188
Total Gamma Terms Energy	1746.8654	1743.13568
Total Gap Geometry Penalty	3108.37922	2746.02559
Total Restraint Energy	0	3550.87301
TOTAL STRUCTURE ENERGY	895477.89425	-96921.59635

Table 6 *Leishmania major* 18S rRNA Energy optimization Table obtained from results of RNA 123that helps minimize the energy from a large positive figure to a more acceptable negative figure that is biologically functional
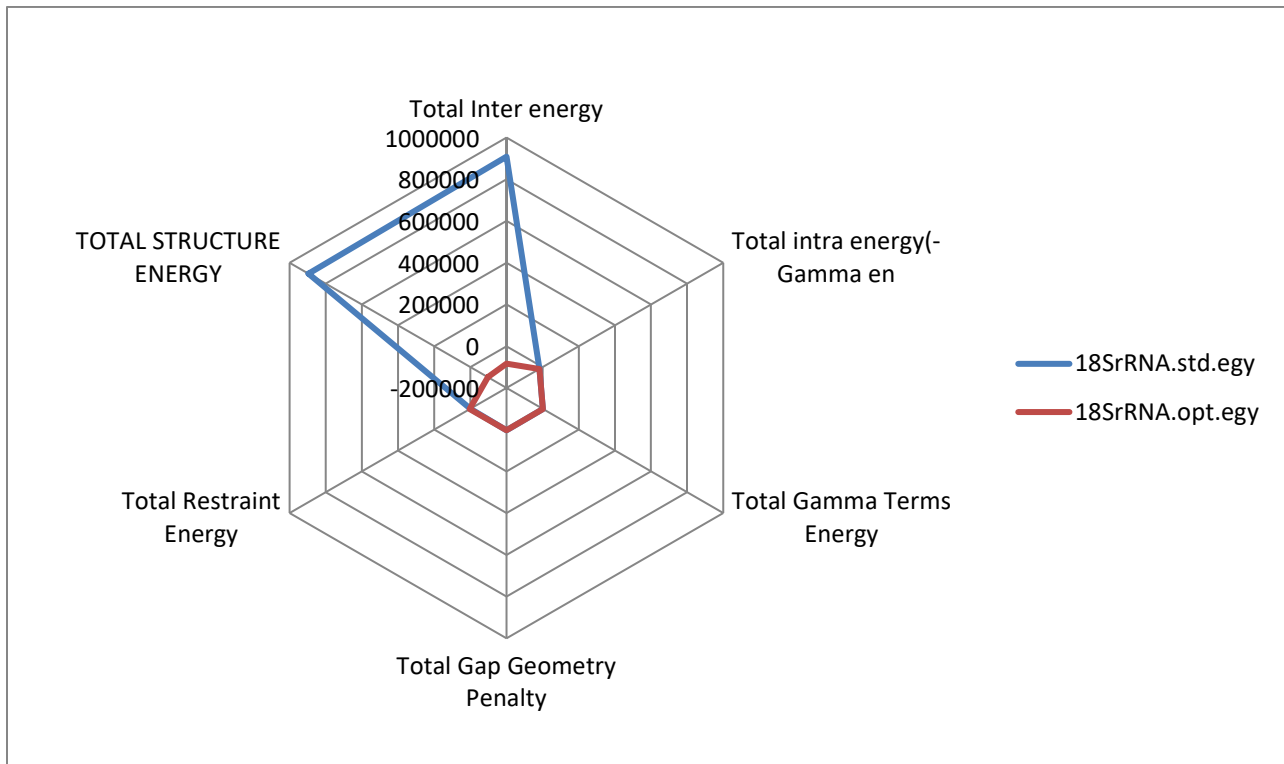
Figure 24 *Leishmania major* 18S rRNA Energy optimization differences from the positive before energy minimization to a more favorable negative structure

5.4.2 *Trypanosoma brucei* 18S rRNA Three dimensional image
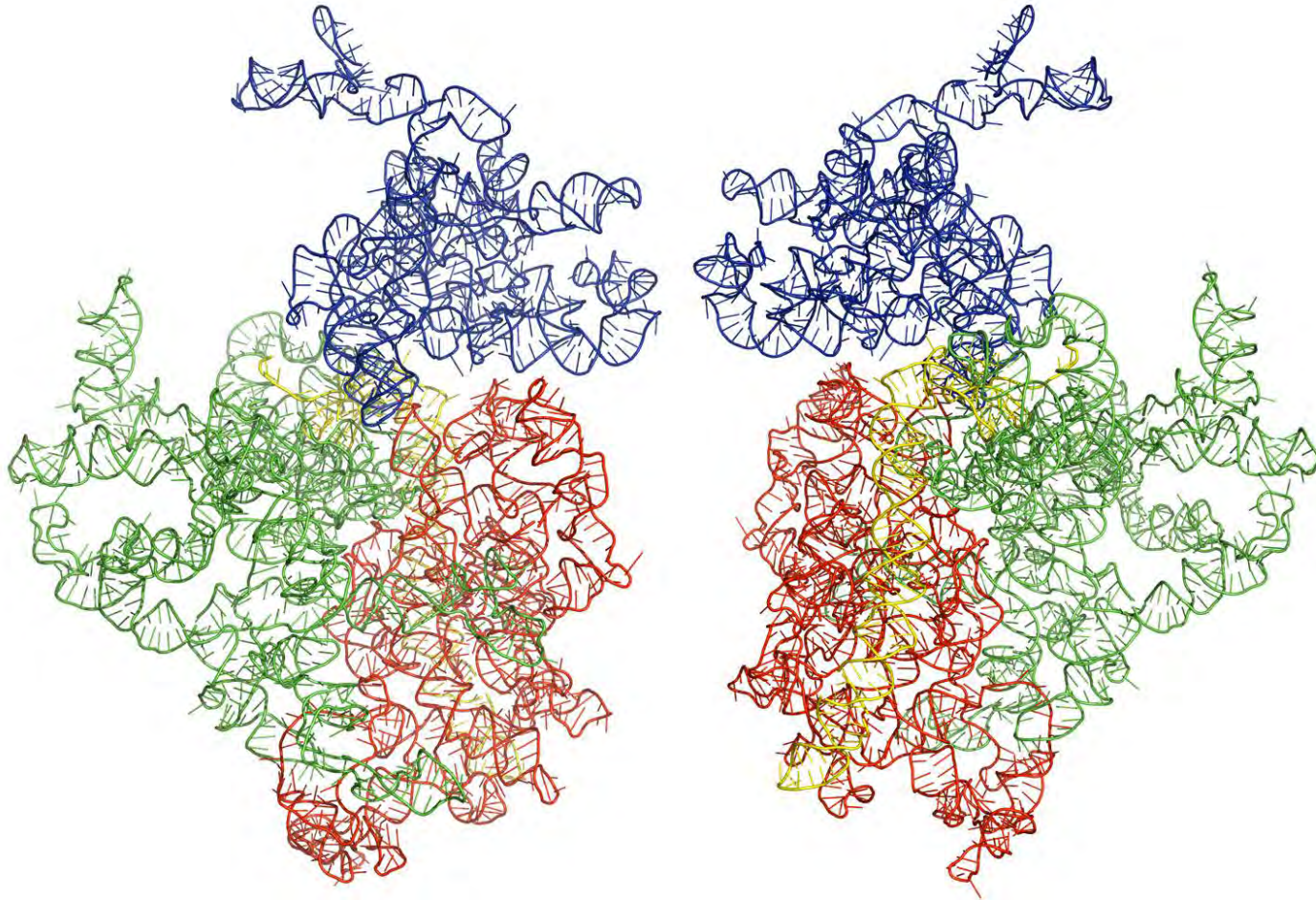
Figure 25 Architectural tertiary structure of *Trypanosoma brucei* 18S rRNA front and back view. Shown is the 18S rRNA colored differently depending on domains (5'major –red, Central-green, 3'major-blue, and 3'minor-yellow).

5.4.2.1 *Typanosoma brucei* **18S rRNA Energy Minimization**

Name	18SrRNA.std.egy	18SrRNA.opt.egy
Total Inter energy	2021190.53226	-102281.51122
Total intra energy(-Gamma en	71145.86365	10625.04139
Total Gamma Terms Energy	2357.05629	2277.69707
Total Gap Geometry Penalty	24166.41849	9500.81691
Total Restraint Energy	0.00000	8078.41829
TOTAL STRUCTURE ENERGY	2118859.87069	-79877.95586

Table 7 *Trypanosoma brucei* 18S rRNA Energy optimization Table obtained from results of RNA 123that helps minimize the energy from a large positive figure to a more acceptable negative figure that is biologically functional.
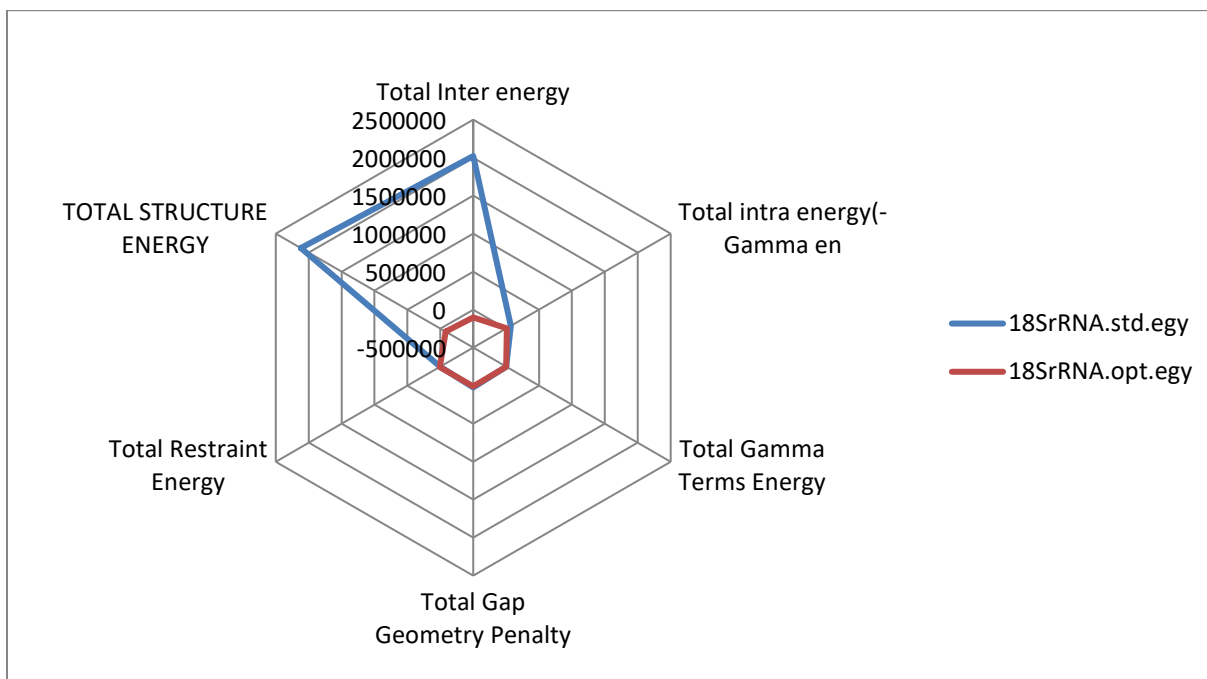
Figure 26 *Trypanosoma brucei* 18S rRNA Energy optimization differences from the positive before energy minimization to a more favorable negative structure.

5.4.3 *Trypanosoma cruzi* 18S rRNA Three dimensional image
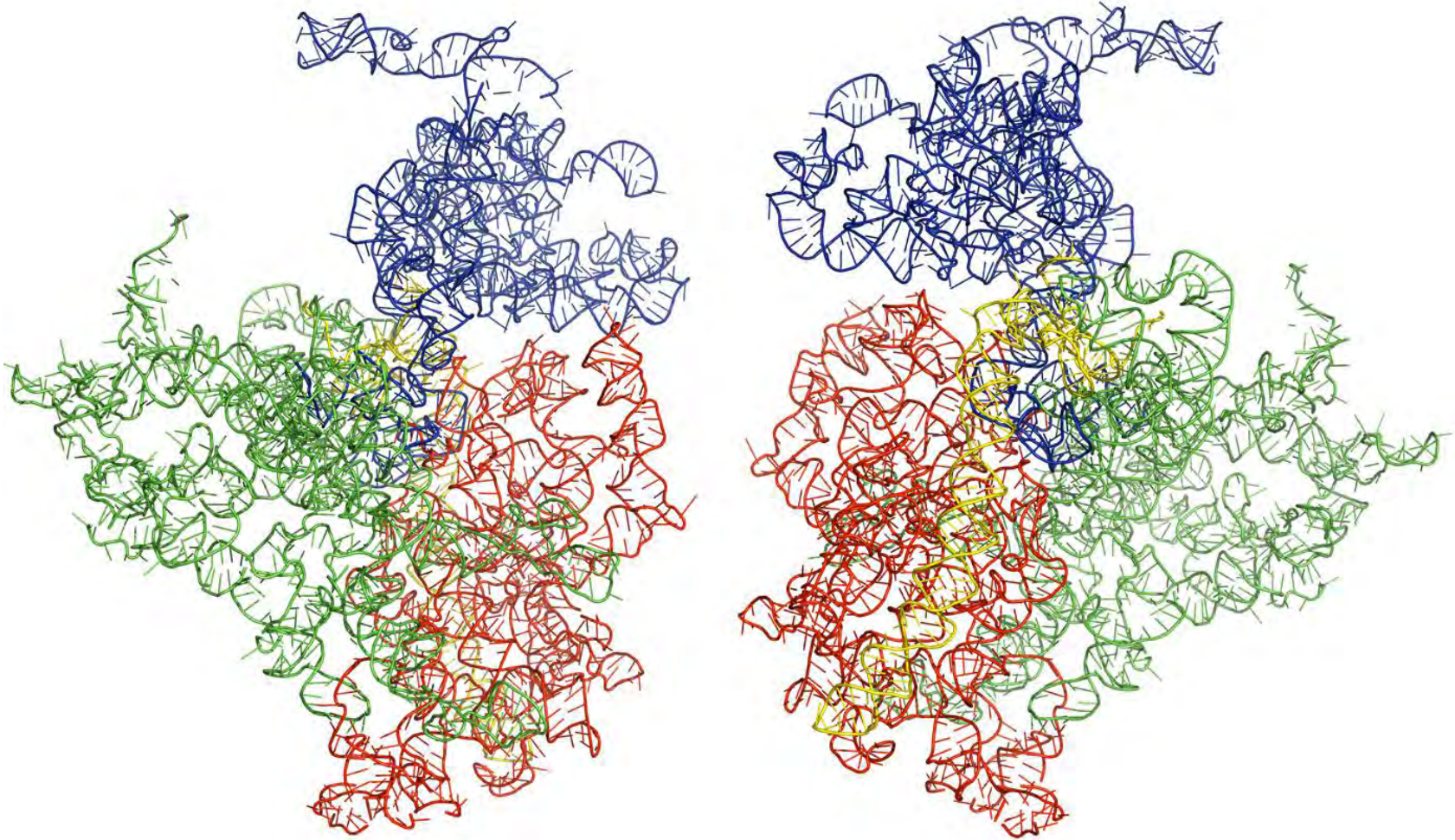
Figure 27 Architectural tertiary structure of Trypanosoma cruzi 18S rRNA front and back view. Shown is the 18S rRNA colored differently depending on domains (5'major –red, Central-green, 3'major-blue, and 3'minor-yellow).

5.4.3.1 *Trypanosoma cruzi* 18S rRNA Energy Minimization

Name	18SrRNA.std.egy	18SrRNA.opt.egy
Total Inter energy	7208497.219	-98209.94034
Total intra energy(-Gamma en	208432.083	-7806.99519
Total Gamma Terms Energy	2458.82781	2440.13835
Total Gap Geometry Penalty	35870.86855	12018.7965
Total Restraint Energy	0	10017.37844
TOTAL STRUCTURE ENERGY	7455258.998	-91558.00067

Table 8 *Trypanosoma cruzi* 18S rRNA Energy optimization Table obtained from results of RNA 123 that helps minimize the energy from a large positive figure to a more acceptable negative figure that is biologically functional.
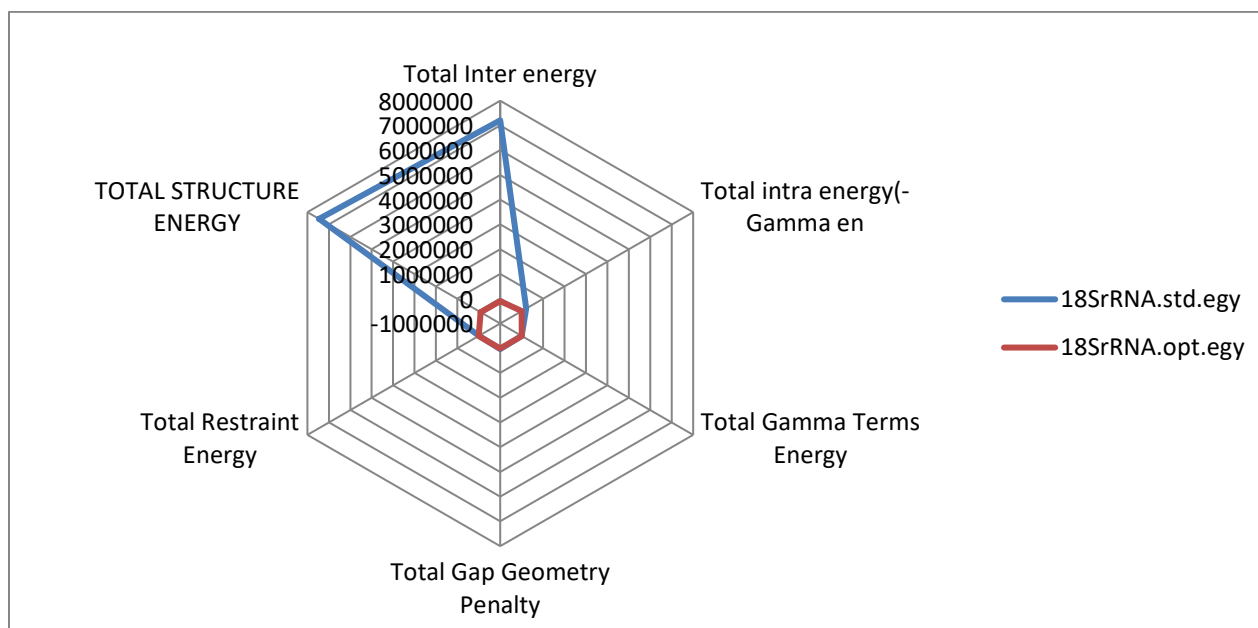
Figure 28 *Trypanosoma cruzi* 18S rRNA Energy optimization differences from the positive before energy minimization to a more favorable negative structure

5.5 Myxobacterial and Selected Pathogen Box Metabolites

5.5.1 Pathogen Box Compounds that show activity on the selected kinetoplastids

| Rack	SMILES	Molecular Formula	Total Molecular Weight	commercial supplier	ChEMBL ID	Trivial name	Compound ID
PathogenBox_PlateB	CC(C)n1c(C)ncc1c2ccnc(Nc3ccc(cc3)S(=O)(=O)C)n2	C18H21N5O2S	371.46	yes	CHEMBL488436	AZD5438	MMV676604
PathogenBox_PlateC	NCCN(C(=O)c1ccc(Cl)cc1Cl)c2ccc(OCc3ccc(Cl)cc3)cc2	C22H19N2O2Cl3	449.76	no	CHEMBL3637896		MMV688371
PathogenBox_PlateC	CN1CCN(CC1)S(=O)(=O)c2ccc(cc2)c3ccc4c(Nc5ncc(Cl)cn5)ccnc4c3	C24H23N6O2ClS	495.00	no	CHEMBL3637895		MMV688283
PathogenBox_PlateD	COc1cc(ccc1n2cc(nn2)c3cccc(c3)C4=NCCN4)C5=NCCN5	C21H21N7O	477.47	no	CHEMBL1198787		MMV688474
PathogenBox_PlateC	COc1ccc(cc1c2cn(nn2)c3cccc(c3)C4=NCCN4)C5=NCCN5	C21H21N7O	613.47	no	CHEMBL1197423		MMV688547
PathogenBox_PlateE	CN(CCc1ccccc1)c2nc(nc(C)c2Cl)c3ccccn3	C19H19N4Cl	338.84	yes	CHEMBL472881		MMV658988
PathogenBox_PlateD	Cc1nc(nc(N2CCC(Cc3ccccc3)C2)c1Cl)c4cccn4	C21H21N4Cl	364.88	yes	CHEMBL474899		MMV659004
PathogenBox_PlateE	CO\N=C(/C(=O)OC)\c1ccccc1CO\N=C(/C)\c2cccc(c2)C(F)(F)F	C20H19N2O4F3	408.37	yes	CHEMBL1897483/CHEMBL3183941/CHEMBL3637899	Trifloxystrobin	MMV688754
PathogenBox_PlateC	CC1=C(F)C=CC=C1CN2CCCC(CC3=NC(C4=CC=C(NC(C)=O)C=C4)=NO3)C2	C24H27N4O2F	422.50	no	CHEMBL3431191		MMV689061
PathogenBox_PlateE	Clc1cccc(CO[C@H]2CCc3ccccc3[C@@H]2n4ccnn4)c1	C19H18N3OCl	339.82	no	CHEMBL1683016		MMV676159
PathogenBox_PlateE	Nc1ccc2ncnc(Nc3ccc(OCc4cccc(F)c4)c(Cl)c3)c2c1	C21H16N4OClF	394.83	no	CHEMBL3637893		MMV688273
PathogenBox_PlateD	FC(F)(F)c1cc(NCCc2ccccc2)nc(n1)c3ccccn3	C18H15N4F3	344.34	yes	CHEMBL511692		MMV659010
PathogenBox_PlateD	Cc1nc(nc(NCCc2ccc(cc2)S(=O)(=O)C)c1Cl)c3cccn3	C19H19N4O2ClS	402.90	no	CHEMBL472854		MMV658993
PathogenBox_PlateA	CCc1nn(C)c(C(=O)NCc2ccc(Oc3ccc(C)cc3)cc2)c1Cl	C21H22N3O2Cl	383.87	yes	CHEMBL2229099	Tolfenpyrad	MMV688934
PathogenBox_PlateC	CC(C)(C)OC(=O)N1CCN(CC1)c2cc(n[nH]2)c3ccc(Oc4ccccc4)cc3	C24H28N4O3	420.50	no	CHEMBL394241		MMV676057
PathogenBox_PlateD	COc1cc(cc2cc(oc12)c3ccc(cc3)C4=NCCN4)C5=NCCN5	C21H20N4O2	360.41	no	CHEMBL520654		MMV688362
PathogenBox_PlateD	COc1cc(ccc1c2cn(nn2)c3cccc(c3)C4=NCCN4)C5=NCCN5	C21H21N7O	387.44	no	CHEMBL1197424		MMV688407
PathogenBox_PlateD	OC(=O)c1cccc2oc(nc12)c3cccc(O)c3NC(=O)c4cc(on4)c5ccccc5	C24H15N3O6	441.39	no	CHEMBL519040		MMV676008
PathogenBox_PlateD	OC(=O)c1cccc2oc(nc12)c3ccccc3NC(=O)c4ccccc4	C21H14N2O4	358.35	no	CHEMBL458799		MMV676162
PathogenBox_PlateD	COC(=O)c1cccc2oc(nc12)c3cccc(OC)c3NC(=O)c4cc(on4)c5ccccc5	C26H19N3O6	469.45	no	CHEMBL3637892		MMV676186

Table 9: showing the Compounds of pathogen box plate rack, their smiles, molecular formulae, molecular weight the commercial supplier the chemical id, those that have trivial names and the compound id from mmv.
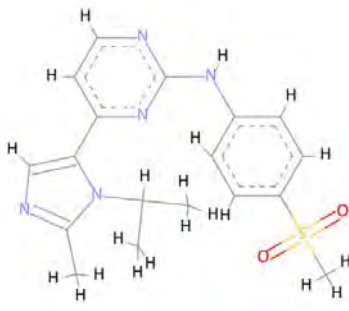
MMV676604
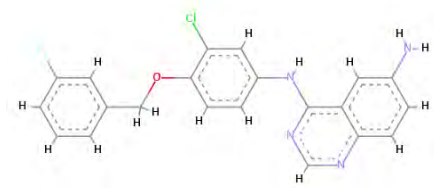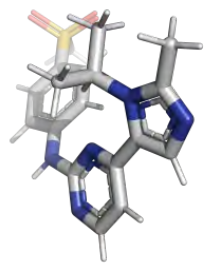
MMV688273
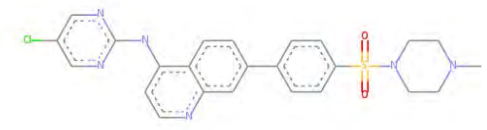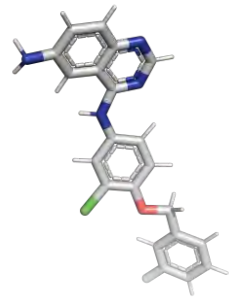
MMV688283
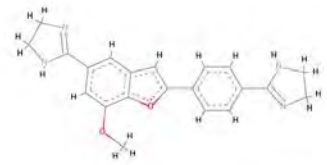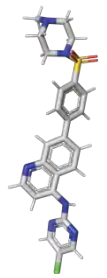
MMV688362
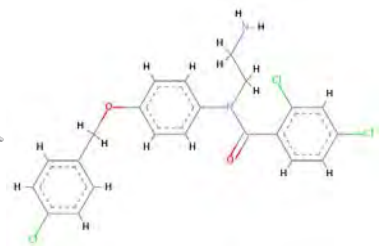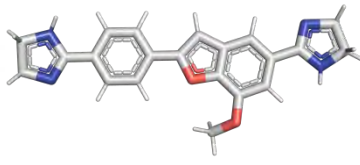
MMV688371
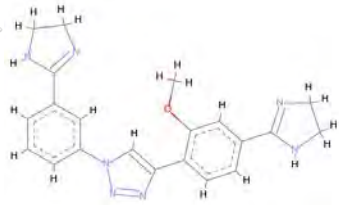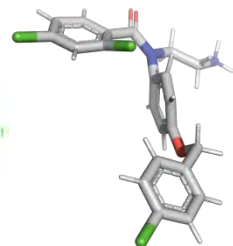
MMV688407
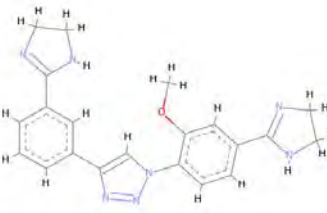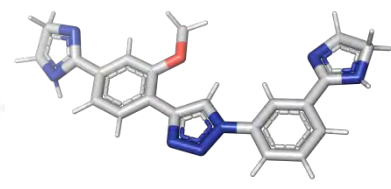
MMV688474

MMV688547

MMV688754

MMV688934

MMV689061

Figure 29: both the two and three-dimensional structures of the 20 metabolites from the pathogen box that were selected for docking and binding with the three selected kinetoplastids

The molecule labels in the figure are:

MV658988 MMV658993 MMV659004

MMV659010 MMV676008 MMV676057

MMV676159 MMV676162 MMV676186

5.5.2 Structural diversity Myxobacteria Compounds that show activity on the selected kinetoplastids

| No	SMILES	Molecular Formula	Total Molecular Weight	commercial supplier	ChEMBL ID	Trivial name
1	[H]\C(=C(\C)[C@]1([H])C[C@]2([H])O[C@]2(C)CCC[C@]([H])(C)[C@]([H])(O)[C@@]([H])(C)C(=O)C(C)(C)[C@@]([H])(O)CC(=O)O1)C1=CSC(C)=N1	$C_{27}H_{41}NO_6S$	507.683		Patupilone	Epothilone B
2	C[C@H]1/C=C/[C@H]([C@H](CCCC[C@H](OC(=O)[C@H]([C@@]2([C@@H]([C@H]([C@@H]([C@H]1O2)C)O)OC)O)C)C3=CC=CC=C3)OC)OC	$C_{29}H_{44}O_8$	520.663	Chemieliva Pharmaceutical Co., Ltd	5281897	Solaphen A
3	CC/C=C\C=C/C(=O)N/C=C/C[C@H]1C[C@H]2C[C@@H](C[C@@H](O2)CC3=C(C(=CC=C3)O)C(=O)O1)O[C@H]4[C@@H]([C@H]([C@@H]([C@H](O4)CO)O)O)NC(=O)C	$C_{33}H_{44}N_2O_{11}$	644.718		10627958	Apicularen
4	C[C@H]([C@H](/C=C/C1=CSC(=N1)C2=CSC(=N2)C(C)/C=C/C=C/C(C)C)OC)/C(=C\C(=O)N)/OC	$C_{25}H_{33}N_3O_3S_2$	487.677		6437357	Myxothiazol A
5	C/C/1=C\CC2=CC=C(N2)C3=NC(C(O3)COC(=O)[C@H](CC(=O)OC4C1OC(=O)C4(C)CO)CC(C)C)/C=C/5\C(OC\C5=C(\C)/CCCOC)CC(C)C	$C_{41}H_{58}N_2O_{10}$	738.919		10908784	Leupyrin
6	CC1=C(C(=O)C2=CC=CC=C2N1O)C/C=C(\C)/CC/C=C(\C)/CCC=C(C)C	$C_{25}H_{33}NO_2$	379.544	MuseChem	6439171	Aurachin A
7	C[C@H]1/C=C/C=C(/[C@H]([C@H]([C@@H](/C=C/C(=C\C(=C/[C@@H]([C@@H](/C=C(/C/C=C(/C(=O)O[C@@H]1C2=CSC(=N2)[C@H](CC(C)C)OC(=O)NC)\C)\C)O)C)\C)/C)O)C)OC)\C	$C_{42}H_{62}N_2O_7S$	739.025		16680454	Archazolid A
8	CC/C=C(\C)/C[C@H](C)[C@@H]([C@@H](C)[C@@H]1[C@@H]([C@@H](C[C@@]2(O1)[C@H](C[C@@H]([C@H](O2)[C@@H](C)[C@H]([C@@H](C)/C=C/C=C/C=C/C=C/[C@@H](CC(=O)O)OC)O)C)O)OC)C)O	$C_{42}H_{68}O_9$	716.997		86576036	Spiragien B
9	C1=CC=C2C(=C1)C(=CN2)C/C=N/O	$C_{10}H_{10}N_2O$	174.203	Chembase.cn	5371769	3-indolylacetaldoxin
10	C1=CC(=C(C(=C1)Cl)[N+](=O)[O-])C2=CNC=C2Cl	$C_{10}H_6Cl_2N_2O_2$	257.07	AHH Chemical co.,ltd MolCore BioPharmatech	13916	Pyrroinitrin
11	CC1CCC(CC(=O)C(C(CCC/C=C/C=C/C=C/C=C/C(=O)OC(C(/C=C/C=C/C1=O)\C)CC(C)C)OS(=O)(=O)O)O)O	$C_{34}H_{50}O_{10}S$	650.824		101578998	Sulfangolid A
12	C[C@H]1[C@H]2C/C=C/[C@H]3[C@@H]([C@H]4C[C@@H](O3)[C@H](O4)/C=C/C=C\C=C/C(=O)O[C@H]5C=C[C@H](C/C=C/CC/C=C/[C@@H]([C@@H]([C@H](O2)C[C@@H]1O)O)O)O[C@H]5/C(=C/[C@H](C)CCCCC(=O)O)/C)C	$C_{47}H_{66}O_{11}$	807.034		657059	Sorangicin A
13	C/C=C/C/C=C(\C)/C1CCC(O1)(C)/C=C/C(C)C(=O)C2=C(C=C(OC2=O)C(C)CC/C=C/NC(=O)OC)O	$C_{30}H_{41}NO_7$	527.658		90477824	Corallopyronin C
14	CC1=C(C2=C([C@@H]3[C@@H]4CC5=C([C@@H](N4[C@H]([C@H]([C@H]2OC)N3C)O)CNC(=O)[C@H](C)N)C(=O)C(=C(C5=O)C)OC)C(=C1OC)O)O	$C_{29}H_{38}N_4O_9$	586.642		159538	Saframycin MX1
15	CCC(C)[C@H]1C(=O)NC2CC3=CC(=C(C(=C3)C4=C(C=CC(=C4)CC(C(=O)N1)N)OC)O)OC5=CC=C(CC(NC2=O)C(=O)O)C=C5	$C_{34}H_{38}N_4O_8$	630.698		101859253	Cittilin A
16	C/C=C/C1=CC(=O)C(C1O)O	$C_8H_{10}O_3$	154.165	Ambinter Chem-Space.com	5372499	Lipoterrein A
17	CC(CC(C)C(=O)C(C)(C)C(C(C)C(CC(=O)O)O)OC)C(C(C)(C(C(C)C=C(C)C)O)O)O	$C_{26}H_{48}O_8$	488.662		10097023	Gephyronic Acid A
18	CCC(=C)C(=O)C1=C(C(=C(C=C1)OCC(=O)O)Cl)Cl	$C_{13}H_{12}Cl_2O_4$	303.135	Sigma-Aldrich	3278	Phenoxan
19	C/C=C(\C)/C=C/C=C/C(C(C)C(C(C)CCC1=C(C(=O)C2=C(O1)C(=C(C=C2OC)OC)O)C)OC)OC	$C_{30}H_{42}O_7$	514.659	3B Scientific (Wuhan) Corp	5353970	Stigmatellin
20	CC1CCC(OC(=O)/C(=C/C(C(C(C(=O)CCC(NC(=O)CC1=O)C)C)O)C)/C)/C=C/C(C)CCC(/C=C\C=C/)C)O	$C_{34}H_{53}NO_7$	587.798		6438827	Angiolam A

Table 10 Showing the Compounds of Myxobacterial, their smiles, molecular formulae, molecular weight the commercial supplier the chemical id, those that have trivial names and the compound id from PubChem

Epothilone B

SORAPHEN A

Apicularen B

Myxothiazol

Leupyrin

Aurachin C

Archazolid A

Spirangien B

Pyrrolnitrin

Sulfangolid A Sorangicin A Corallopyronin C Saframycin Mx1

Terrein Gephyronic Acid A Ethacrynic Acid Stigmatellin

Indole-3-acetaldoxime Cittilin A

Figure 30 : both the two and three dimensional structures of the 20 metabolites from the Myxobacteria compound database that were selected for docking and binding with the three selected kinetoplastids

5.6   Docking Results

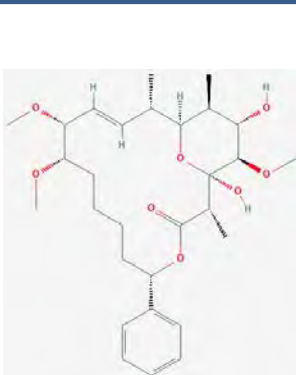|	compound	Score	Area	ACE	Transformation
18S Rrna *T. Brucei*	**Angiolam A**	7538	960.1	-491.7	0.90 0.16 2.72 -2.97 73.29 -20.71
	Apicularen B	7782	1024.5	-549.58	-2.31 1.27 -2.86 -31.34 2.00 26.58
	Archazolid A	8090	995.9	-516.32	-0.47 -0.67 2.98 -37.96 48.52 41.13
	Aurachin C	6112	688.5	-452.88	-0.26 -0.72 -3.13 -14.04 87.70 -19.63
	Cittilin A	7320	835.8	-495.42	1.27 0.72 -0.71 -21.12 68.88 -19.66
	Corallopyronin C	6642	787.4	-374.61	-2.91 1.12 0.27 -77.90 18.40 37.53
	Epothilone B	6556	786.7	-573.04	0.06 0.90 0.70 -13.29 88.18 -17.77
	Ethacrynic Acid	4544	578.1	-299.12	-2.10 -0.13 -1.74 -10.01 77.16 -18.32
	Gephyronic Acid A	7210	762.4	-313.9	1.19 -0.55 0.24 -95.90 4.16 -21.45
	Indole-3-acetaldoxime	3372	401.7	-236.32	1.73 -0.07 -1.38 -10.99 75.60 -18.28
	Leupyrin	7968	1164.3	-598.53	1.62 0.22 -1.68 -78.41 22.38 33.11
	Myxothiazol	7764	907.7	-595.18	-0.67 0.08 2.17 -12.11 84.09 -19.73
	Pyrrolnitrin	4042	447.8	-197.57	1.23 -0.87 -1.31 -30.29 2.01 27.63
	Saframycin Mx1	6666	760.5	-411.7	-3.00 -0.23 -0.71 -78.86 19.74 38.70
	Sorangicin A	8372	1019.7	-466.93	1.26 -0.36 2.35 -40.66 0.88 52.68
	Soraphen A	6664	726.6	-326.43	-1.88 0.07 -1.07 -21.83 69.61 -19.09
	Spirangien B	8606	1135.1	-503.52	1.71 -0.01 1.76 -77.33 2.71 33.47
	Stigmatellin	7300	905.4	-391.6	-2.76 0.29 -2.96 -11.44 105.84 -31.50
	Sulfangolid A	8134	986.9	-613.53	0.54 0.78 2.91 -22.28 79.52 8.58
	Terrein	3248	344.7	-170.2	-0.99 -0.34 1.90 -11.14 75.61 -18.89
	MMV658988	5596	624.8	-364.06	2.52 -0.15 1.18 -15.32 87.98 -15.33
	MMV658993	6008	734.3	-461.11	-2.51 -1.44 -1.22 -35.45 5.85 25.70
	MMV659004	5584	695.4	-379.48	-0.71 0.29 -1.29 -6.93 78.03 -22.58
	MMV659010	5710	632.6	-292.53	-2.59 1.04 -0.53 -14.28 86.01 -18.79
	MMV676008	5884	699.6	-403.76	2.52 0.48 0.57 -31.14 -4.00 19.34
	MMV676057	6436	788.4	-376.63	-2.81 -0.88 2.57 -76.91 14.39 39.56
	MMV676159	5354	655.8	-392.28	1.84 -0.30 -2.05 -33.82 10.65 27.19
	MMV676162	5238	629.5	-415.7	1.29 0.57 -2.23 -31.57 8.03 23.87
	MMV676186	6522	778.4	-527.13	-1.94 -0.20 -3.09 -8.31 84.46 -23.09
	MMV676604	5436	655	-446.72	2.33 0.32 -2.36 -33.95 6.78 20.17
	MMV688273	5846	683.3	-365.05	0.10 -0.88 1.59 -19.72 62.12 -14.25
	MMV688283	6424	771	-402.87	-0.24 -0.88 2.15 -78.93 16.95 44.10
	MMV688362	5566	644	-408.49	-2.87 0.02 -0.99 -35.57 9.04 20.84
	MMV688371	6222	742.7	-426.76	-1.13 -1.08 -1.41 -16.61 89.54 -16.36
	MMV688407	5972	710.2	-443.86	-2.21 0.23 0.51 -41.59 4.19 16.90
	MMV688474	5766	703	-475.43	-0.85 0.61 -1.56 -36.66 4.24 15.59
	MMV688547	5956	702	-450.88	-3.04 0.55 1.10 -14.60 61.99 -23.00
	MMV688754	6210	742.5	-315.15	2.74 0.75 1.85 -11.57 61.57 -26.16
	MMV688934	6324	728.7	-421.33	0.10 -0.02 -2.97 -32.46 8.40 17.03
	MMV689061	6558	723.7	-381.28	-0.22 1.10 -1.48 -33.32 5.48 15.80

Table 11 showing docking and binding results of the best pose of *T.Brucei* and the compounds from pathogen box and myxobacteria databases

18S rRNA Receptor	compound	Score	Area	ACE	Transformation
18S Rrna *T. cruzii*	**Angiolam A**	8270	997.10	-673.71	-2.52 0.18 -1.33 -13.02 84.56 -18.90
	Apicularen B	8222	974.60	-529.41	1.59 -0.09 2.62 -8.71 57.25 125.25
	Archazolid A	8320	984.30	-470.74	2.21 0.39 2.50 -19.59 69.42 -21.95
	Aurachin C	6024	690.30	-445.79	-0.49 -0.80 2.64 -14.40 87.92 -19.33
	Cittilin A	6960	933.90	-529.71	-0.22 -0.41 -2.87 -78.15 84.97 88.37
	Corallopyronin C	7182	875.10	-390.99	-0.39 0.17 2.05 -132.85 -12.76 21.96
	Epothilone B	6548	799.90	-513.65	0.39 0.02 -0.77 -11.85 85.74 -18.70
	Ethacrynic Acid	4794	518.60	-248.66	2.05 1.08 2.53 -119.65 -49.12 68.43
	Gephyronic Acid A	7338	962.00	-443.97	-1.68 0.32 1.64 -10.32 60.89 -22.59
	Indole-3-acetaldoxime	3414	407.60	-235.62	-1.23 -0.06 1.87 -10.93 75.20 -19.02
	Leupyrin	8276	1102.10	-648.66	-0.51 0.45 2.04 -127.03 -12.70 22.84
	Myxothiazol	7552	989.80	-573.36	1.29 -0.40 -1.75 -10.20 70.88 -19.66
	Pyrrolnitrin	4194	491.10	-253.11	0.75 -0.35 -1.46 -10.11 75.13 -18.00
	Saframycin Mx1	6960	933.90	-529.71	-0.22 -0.41 -2.87 -78.15 84.97 88.37
	Sorangicin A	8372	1019.70	-466.93	1.26 -0.36 2.35 -40.66 0.88 52.68
	Soraphen A	6664	726.60	-326.43	-1.88 0.07 -1.07 -21.83 69.61 -19.09
	Spirangien B	8964	1093.20	-576.45	-2.52 0.30 -0.65 -31.08 75.12 5.59
	Stigmatellin	7300	905.40	-391.60	-2.76 0.29 -2.96 -11.44 105.84 -31.50
	Sulfangolid A	8134	986.90	-613.53	0.54 0.78 2.91 -22.28 79.52 8.58
	Terrein	3264	377.00	-148.67	1.55 0.18 -1.48 -10.47 74.37 -18.75
	MMV658988	5712	658.40	-305.82	2.47 -1.00 -0.05 -82.02 -1.02 -7.22
	MMV658993	5976	771.90	-482.12	-0.83 0.39 -2.59 -6.79 70.51 -22.19
	MMV659004	6116	744.70	-436.29	2.70 0.73 2.72 -5.57 70.81 -21.74
	MMV659010	5824	692.40	-298.28	-0.96 0.14 -3.07 -84.35 -3.18 -9.79
	MMV676008	6234	749.40	-456.56	-0.41 -0.56 -1.51 -32.99 7.92 24.60
	MMV676057	6660	837.10	-507.32	-1.31 0.47 0.46 -11.25 62.87 -28.30
	MMV676159	5596	656.30	-408.51	-2.14 0.85 1.65 -3.68 56.13 116.73
	MMV676162	5362	663.00	-366.83	-0.81 -0.59 -0.26 -80.05 91.77 85.16
	MMV676186	6434	733.30	-368.50	-0.35 0.83 1.25 -32.38 43.54 41.41
	MMV676604	5598	684.10	-495.09	0.23 -0.15 1.49 -37.00 2.71 17.93
	MMV688273	6062	686.10	-375.72	-0.39 0.18 -2.00 -25.85 37.76 118.82
	MMV688283	6956	863.90	-620.09	-1.42 -1.38 1.01 -13.18 75.28 -11.91
	MMV688362	5744	735.50	-385.98	-1.95 -0.89 -2.26 -10.56 65.89 -17.66
	MMV688371	6272	772.00	-369.69	-2.35 -0.84 -1.56 -41.56 37.13 127.62
	MMV688407	6194	742.70	-428.66	-2.09 0.43 -2.62 -3.00 56.03 120.69
	MMV688474	6080	770.30	-451.01	2.34 -0.13 3.05 -9.44 71.35 -16.24
	MMV688547	6346	800.20	-477.69	-3.06 0.78 -0.39 -12.83 68.79 -24.17
	MMV688754	6740	795.70	-333.81	1.57 1.16 -1.75 -13.47 84.52 -24.93
	MMV688934	6224	712.00	-441.40	0.29 -0.27 -2.83 -32.20 8.32 17.16
	MMV689061	6708	819.20	-375.30	-1.15 -1.15 -1.54 -79.64 -2.41 -8.42

Table 12 showing docking and binding results of the best pose of *T.cruzi* and the compounds from pathogen box and myxobacteria databases

18S rRNA Receptor	compound	Score	Area	ACE	Transformation
18S Rrna *L.Major*	**Angiolam A**	7682	1034.30	-550.93	2.92 0.90 -2.52 -80.03 -14.11 -2.62
	Apicularen B	7878	1012.30	-585.93	-1.51 -0.99 -2.18 -30.85 74.22 -3.84
	Archazolid A	8520	1035.70	-413.53	0.17 -0.92 1.18 0.27 73.02 -18.49
	Aurachin C	5558	637.00	-294.22	0.16 -0.14 2.88 -92.73 27.50 -13.98
	Cittilin A	7158	864.20	-520.78	-1.85 -0.30 -1.97 -30.25 74.93 -0.05
	Corallopyronin C	7036	830.70	-493.45	-0.59 0.67 1.74 -22.61 72.31 -19.02
	Epothilone B	6712	730.70	-346.85	2.61 0.59 1.99 -23.10 74.50 -17.98
	Ethacrynic Acid	4770	542.60	-259.51	-0.03 -0.88 -1.93 -31.44 72.71 -7.17
	Gephyronic Acid A	7230	903.20	-435.94	2.53 -0.66 -1.97 -27.09 83.02 0.66
	Indole-3-acetaldoxime	3670	401.20	-240.98	2.45 -1.35 0.36 -81.54 -16.52 0.90
	Leupyrin	8202	899.80	-393.82	-0.78 0.06 -0.70 -13.48 81.52 13.85
	Myxothiazol	7014	812.00	-449.90	-2.26 -0.88 2.05 52.20 114.66 40.02
	Pyrrolnitrin	4176	468.70	-257.02	-1.84 -0.45 -2.66 -31.22 72.13 -9.03
	Saframycin Mx1	6770	789.50	-369.01	1.54 -0.14 -1.14 -79.65 17.54 38.02
	Sorangicin A	8550	992.80	-456.49	-2.68 1.07 1.15 51.70 102.16 5.45
	Soraphen A	6856	794.90	-432.27	2.80 -0.63 2.73 -22.02 71.55 -19.16
	Spirangien B	8660	1142.20	-516.62	-1.17 -0.32 -0.03 -103.76 -5.69 -15.28
	Stigmatellin	7148	887.00	-439.63	2.12 0.83 1.02 -31.15 74.14 -5.99
	Sulfangolid A	7974	982.70	-643.25	1.79 1.20 1.11 -30.83 70.84 -8.48
	Terrein	3110	356.90	-178.85	0.51 -0.79 0.51 17.69 109.81 3.88
	MMV658988	5632	680.40	-340.63	-0.02 0.08 -0.76 -45.35 5.38 14.62
	MMV658993	6458	769.50	-407.70	-2.60 -0.27 2.01 -76.43 -17.86 -2.69
	MMV659004	6100	715.90	-429.23	0.32 0.37 2.93 -76.33 -17.89 -3.13
	MMV659010	5786	703.00	-322.78	-2.13 1.02 -0.05 -42.88 3.20 15.36
	MMV676008	6054	720.90	-418.32	-0.40 1.34 1.11 -48.25 4.11 13.56
	MMV676057	6678	813.90	-453.23	1.76 0.06 1.84 16.68 104.98 5.43
	MMV676159	5458	643.10	-424.10	2.88 0.75 -0.24 -47.02 -0.20 12.56
	MMV676162	5498	640.90	-318.07	2.02 0.08 0.54 2.07 102.84 -7.15
	MMV676186	6514	783.90	-407.26	-1.16 0.95 1.41 -49.66 6.81 13.52
	MMV676604	5908	672.50	-379.47	-2.36 0.26 -0.29 20.49 114.77 -1.16
	MMV688273	6224	735.80	-437.21	-1.49 0.18 -0.58 -85.86 -13.83 -3.59
	MMV688283	6808	845.00	-477.80	-2.73 -0.92 1.77 -32.27 65.91 -5.63
	MMV688362	5658	699.30	-411.60	-0.16 0.61 -2.09 -82.72 -12.58 -1.63
	MMV688371	6326	716.40	-414.71	2.46 -0.17 -2.94 -26.47 71.67 3.80
	MMV688407	5994	738.70	-452.27	-2.22 -0.22 0.68 -44.49 1.19 14.00
	MMV688474	6264	737.20	-454.60	-1.48 0.61 -1.24 -42.88 2.96 11.86
	MMV688547	6170	743.40	-391.76	-1.30 0.68 0.49 -1.93 105.61 -13.62
	MMV688754	6520	789.90	-282.72	-0.84 0.28 -1.11 20.45 121.14 -0.01
	MMV688934	6378	757.70	-419.02	0.17 0.41 -2.97 -37.96 6.66 10.92
	MMV689061	6778	798.50	-316.48	0.07 1.09 0.42 -82.22 -13.68 -4.04

Table 13 Showing docking and binding results of the best pose of *L.major* and the compounds from pathogen box and myxobacteria databases

5.6.1 Summary Atomic Contact Energies of best pose complexes of 18S rRNA and various compounds

Compound Name	Kinetoplastids 18S rRNA Atomic Contact Energies(ACE) Values (kcal/mol)		
	T.Brucei	*T. Cruzi*	*L.Major*
Angiolam A	-491.7	-673.71	-550.93
Apicularen B	-549.58	-529.41	-585.93
Archazolid A	-516.32	-470.74	-413.53
Aurachin C	-452.88	-445.79	-294.22
Cittilin A	-495.42	-529.71	-520.78
Corallopyronin C	-374.61	-390.99	-493.45
Epothilone B	-573.04	-513.65	-346.85
Ethacrynic Acid	-299.12	-248.66	-259.51
Gephyronic Acid A	-313.9	-443.97	-435.94
Indole-3-acetaldoxime	-236.32	-235.62	-240.98
Leupyrin	-598.53	-648.66	-393.82
Myxothiazol	-595.18	-573.36	-449.90
Pyrrolnitrin	-197.57	-253.11	-257.02
Saframycin Mx1	-411.7	-529.71	-369.01
Sorangicin A	-466.93	-466.93	-456.49
Soraphen A	-326.43	-326.43	-432.27
Spirangien B	-503.52	-576.45	-516.62
Stigmatellin	-391.6	-391.60	-439.63
Sulfangolid A	-613.53	-613.53	-643.25
Terrein	-170.2	-148.67	-178.85
MMV658988	-364.06	-305.82	-340.63
MMV658993	-461.11	-482.12	-407.70
MMV659004	-379.48	-436.29	-429.23
MMV659010	-292.53	-298.28	-322.78
MMV676008	-403.76	-456.56	-418.32
MMV676057	-376.63	-507.32	-453.23
MMV676159	-392.28	-408.51	-424.10
MMV676162	-415.7	-366.83	-318.07
MMV676186	-527.13	-368.50	-407.26
MMV676604	-446.72	-495.09	-379.47
MMV688273	-365.05	-375.72	-437.21
MMV688283	-402.87	-620.09	-477.80
MMV688362	-408.49	-385.98	-411.60
MMV688371	-426.76	-369.69	-414.71
MMV688407	-443.86	-428.66	-452.27
MMV688474	-475.43	-451.01	-454.60
MMV688547	-450.88	-477.69	-391.76
MMV688754	-315.15	-333.81	-282.72
MMV688934	-421.33	-441.40	-419.02
MMV689061	-381.28	-375.30	-316.48

Table 14 Selected kinetoplastids 18S rRNA and various natural compounds docked complexes best pose Atomic Contact Energy (ACE)
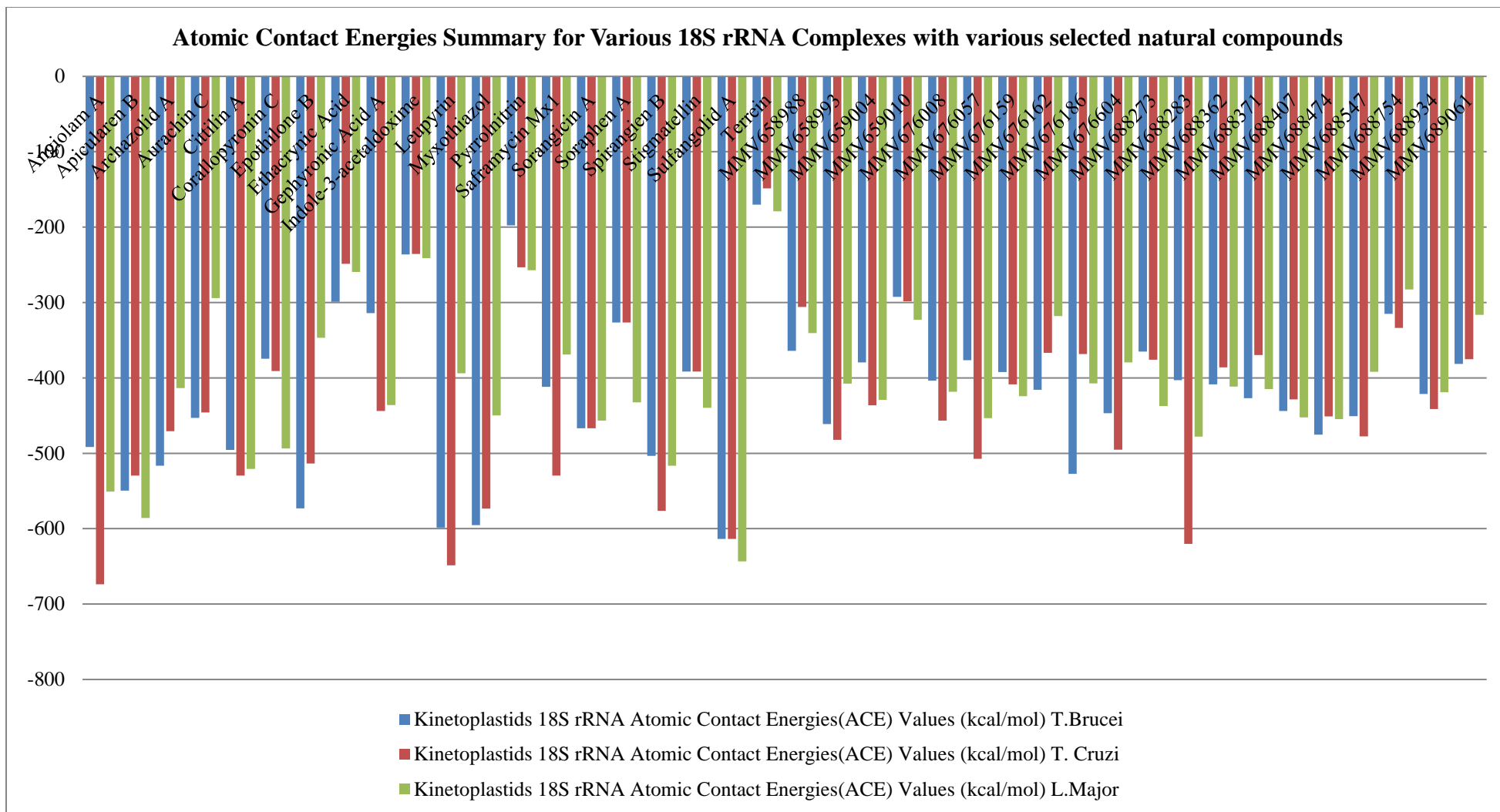
Figure 31 A 2D Graph showing the various Atomic Contact Energy (ACE) values in kcal/mol of the various selected kinetoplastid 18S rRNA docked to selected natural compounds
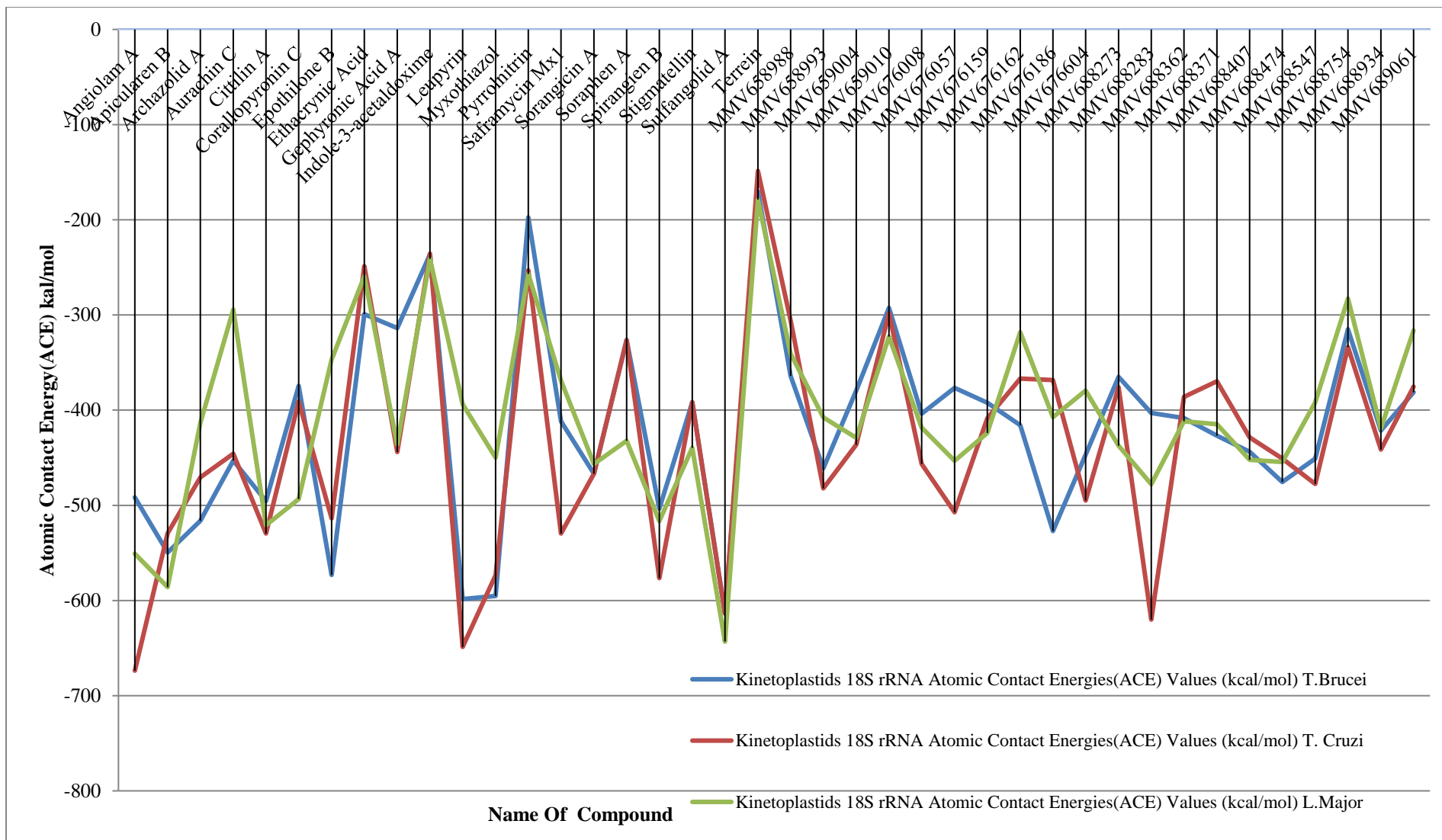
Figure 32 A Line Graph showing the various Atomic Contact Energy (ACE) values in kcal/mol of the various selected kinetoplastid 18S rRNA docked to selected natural compounds and how they correlate.

5.6.2 Atomic Contact Energies Summary for Various 18S rRNA Complexes kinetoplastids more negative ACE (kcal/mol) -400

Compound Name	Compounds with activity on all the kinetoplastids more negative ACE (kcal/mol) -400		
	T.Brucei	*T. Cruzi*	*L.Major*
Angiolam A	-491.7	-673.71	-550.93
Apicularen B	-549.58	-529.41	-585.93
Archazolid A	-516.32	-470.74	-413.53
Cittilin A	-495.42	-529.71	-520.78
Epothilone B	-573.04	-513.65	-346.85
Leupyrin	-598.53	-648.66	-393.82
Myxothiazol	-595.18	-573.36	-449.9
Sorangicin A	-466.93	-466.93	-456.49
Spirangien B	-503.52	-576.45	-516.62
Sulfangolid A	-613.53	-613.53	-643.25
MMV658993	-461.11	-482.12	-407.7
MMV676008	-403.76	-456.56	-418.32
MMV688283	-402.87	-620.09	-477.8
MMV688407	-443.86	-428.66	-452.27
MMV688474	-475.43	-451.01	-454.6
MMV688934	-421.33	-441.4	-419.02

Table 15 Line Graph showing the various Atomic Contact Energy (ACE) values in kcal/mol of the various selected kinetoplastid 18S rRNA docked Compounds with activity on all the kinetoplastids more negative ACE (kcal/mol) -400
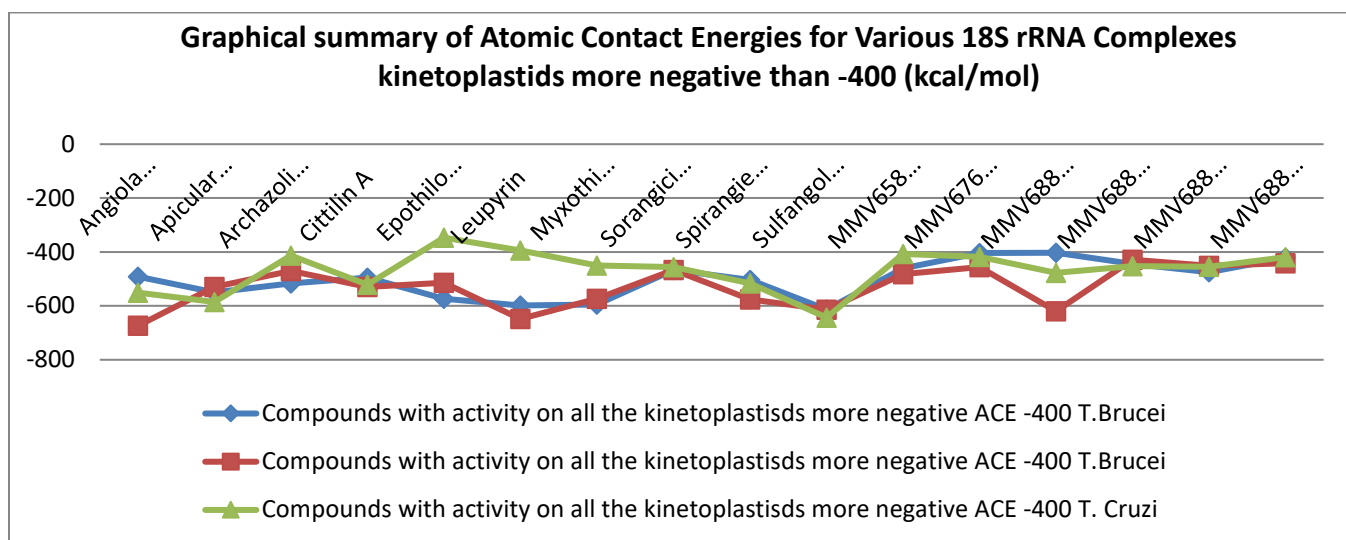
Figure 33 Graphical summary of compounds docked on specific positions of selected kinetoplastids with ACE below -400(kcal/mol)

5.6.3   Natural compounds binding sites on Various Kinetoplastids

Figure 34 *T. Brucei* bound to Angiolam at helix
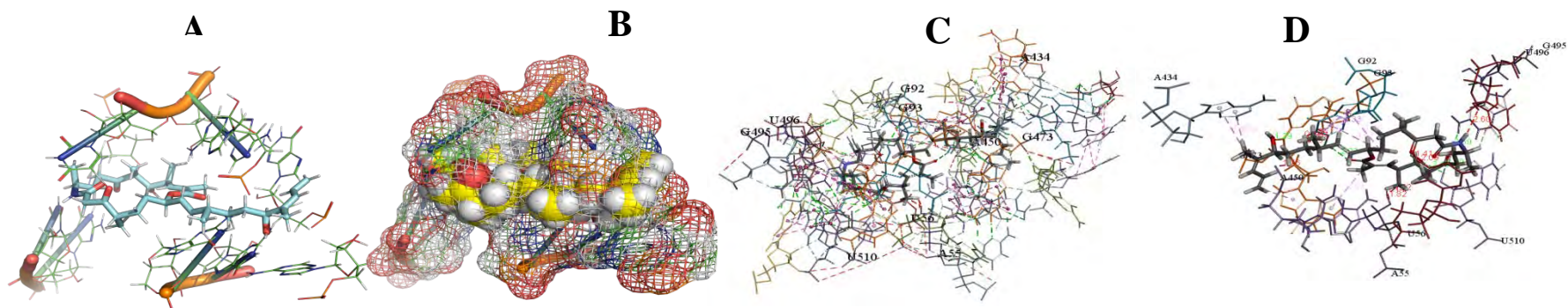
Figure 35 *T Cruzi* bound to Angiolam at helix
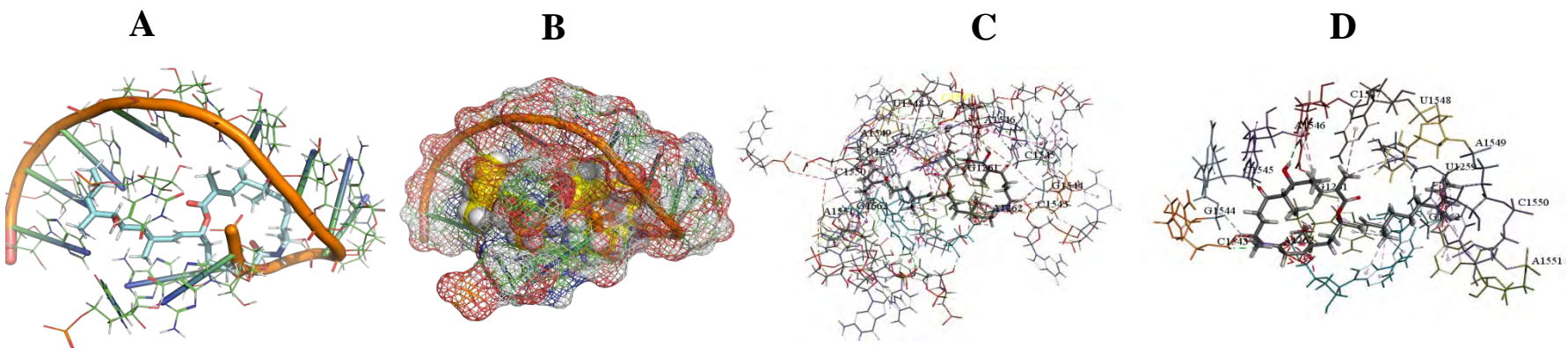
Figure 36 *L.Major* bound to Angiolam at helix

5.6.4 Residues at the active site binding to the natural compounds

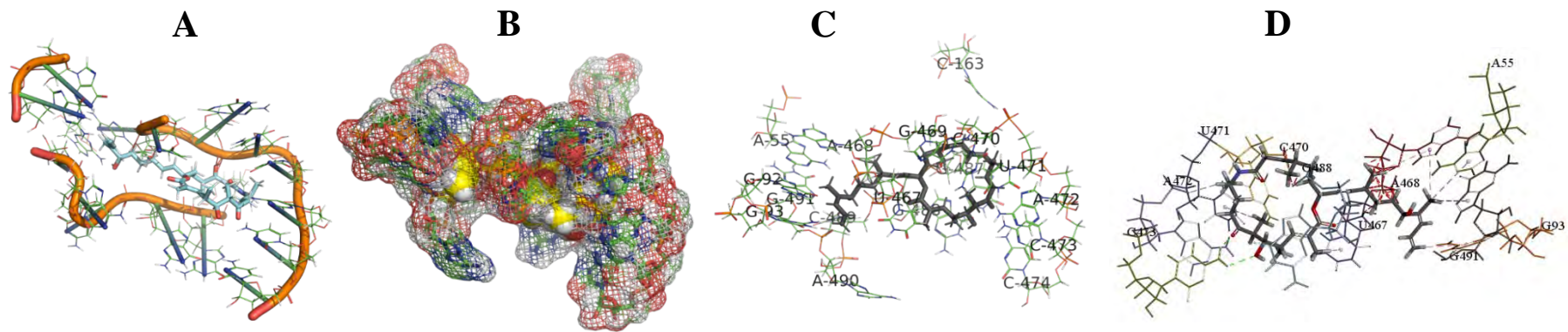| Compound Name	Compounds with activity on all the kinetoplastisds more negative ACE -400					
	T.Brucei	Contact residues	*T. Cruzi*	Contact residues	*L.Major*	Contact residues
Angiolam A	-491.7	G92,G93,A434,A450,G470,G473,G495,U496,U510	-673.71	A55,U56,G92,G93,A434,A450,G473,G495,U496,U510	-550.93	U1259,G1261,A1262,C1543,G1544,C1545,A1546,C1547,U1548,A1549,C1550,A1551,G1662
Apicularen B	-549.58	G1253,A1254,C1255,A1257,U1258,G1260,U2230,G2231	-529.41	G1109,U1110,A1134,C1135,U1150,G1151,U1152,C1153	-585.93	U27,A28,A40,G41,G407,A421,U422,U423,A813
Archazolid A	-516.32	G690,U691,U692,A693,G1281,A1282,C1283,A1284,G1460,A1461,A1470,G1471,G1472,U1473,G1478	-470.74	C94,U427,A472,C474,A475,G476,G477,C478,A485	-413.53	C164,G165,U445,C448,U449,A450,G465,G466,
Cittilin A	-495.42	A43,A47,G48,C94,U95,C492,A493,G494,C496TTTT	-529.71	U716,G719,G738,U740,G741,A742,C1051,U1052	-520.78	A26,U27,A40,C419,G420,A421,U422,U423,
Epothilone B	-573.04	A55,U56,A90,U91,G92,A468,G513,U514,C515,A527,U528,A530	-513.65	U56,A90,G92,A450,U496,C497,A512	-346.85	G42,C50,A471,G472,G473,C474,A481
Leupyrin	-598.53	G1532,C1533,A1534,U1663,U1683,G1686,A1690,U1691,A2092,U2093	-648.66	C1804,A1807,U1809,A1810,A1813,U1884,U1887	-393.82	A103,C105,G107,A108,A347,C349,U350
Myxothiazol	-595.18	A55,U56,G92,U467,A468,C489,G513,U514,C515,A527,U528,A530	-573.36	C49,C50,A434,G435,U449,A450,G470,C471,A472,G473,	-449.9	C203,A205,G206,C218,U224,C225,U227,G228,
Sorangicin A	-466.93	A1240,A1241,G1253,C1255,C1256,A1257,U1258,U1259,U2220	-466.93		-456.49	C235,C236,A237,A304,U306,
Spirangien B	-503.52	C59,U60,A64,G79,G80,A520,A521,C525,G526	-576.45	A100,G106,G407,G409,C423,G424,A864,A902	-516.62	A1294,C1535,C1536,A1538,A1549,A1642,U1644,U1645,A1689
Sulfangolid A	-613.53	A105,G373,C374,G428,A885,U919,A920,C921,A922	-613.53		-643.25	A26,U27,G41,G407,G408,U422,U423,A813
MMV658993	-461.11	A1244,G1253,A1254	-482.12	C49,G92,G93,A434,A450,G470,C471,A472,G473	-407.7	U1259,G1260,G1261,A1262,C1547,U1548,C1550
MMV676008	-403.76	G1243,A1244,G1253,A1254	-456.56	G1328,A1329,G1330,A1331,C1332,C2297	-418.32	A1092,A1094,G1241,A1833,G1839,U1840,C1855
MMV688283	-402.87	G1532,C1533,U1663,G1664,G1665,U1683,G1689,A1690,U1691	-620.09	C49,A55,G93,A434,G435,U449,A450,C458,G470,A472, G473	-477.8	U27,A28,A40,G41,A43,C95,U96,A421,U422,U423,A517
MMV688407	-443.86	G1251,A1252,A1254,A2214,A2221,G2222,A2233,C2235,U2236,	-428.66	G1099,U1100,G1109,U1111,A1134,C1135,A1147	-452.27	A1092,G1093,A1833,A1834,U1840,G1842,C1854,C1855
MMV688474	-475.43	G1251,A1252,G1253,A1254,C1255,U2230,G2231,A2232,A2233,C2235	-451.01	C50,G435,G470,C471,A472,G473	-454.6	A1092,G1093,A1094,A1833,A1834,G1839,U1840,G1842,C1854,C1855
MMV688934	-421.33	G1252,G1253,A2214,G2222,A2233,C2235,	-441.4	G1328,A1329,G1330,A1331	-419.02	A1092,A1834,U1840,G1842,C1855

A   B   C   D

T Brucei + Angiolam

A   B   C   D

T cruzi + Angiolam

A   B   C   D

L.major+Angiolam

A

B

C

D

T brucei + Apicuralen

A

B

C

D

T cruzi +Apicuralen

A

B

C

D

L major Apicuralen

A

B

C

D

T Brucei+ Archazolid

A

B

C

D

T cruzi+ Archazolid

A

B

C

D

L major+ Archazolid

A T brucei + Cittilin

B

C

D

A T cruzi + cittilin

B

C

D

A L major+ cittilin

B

C

D

79
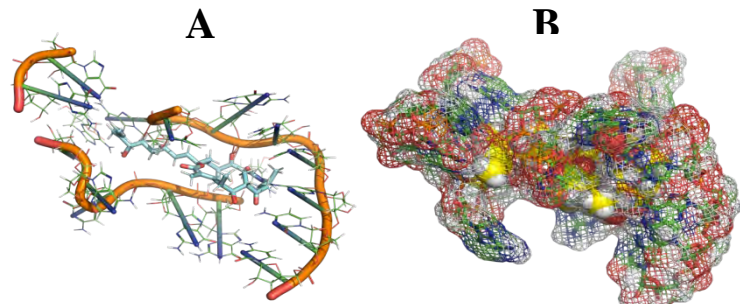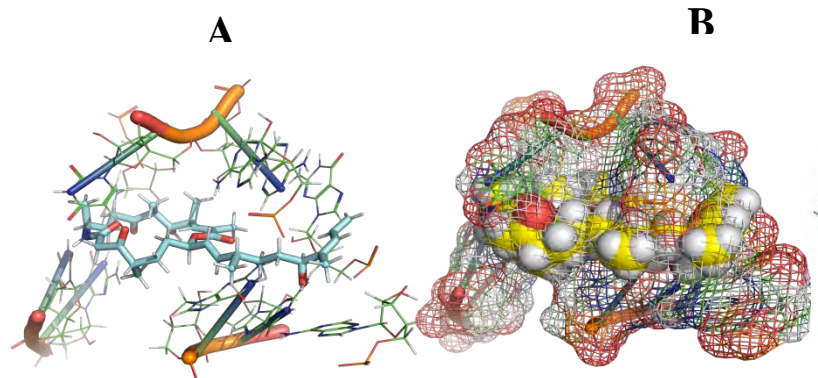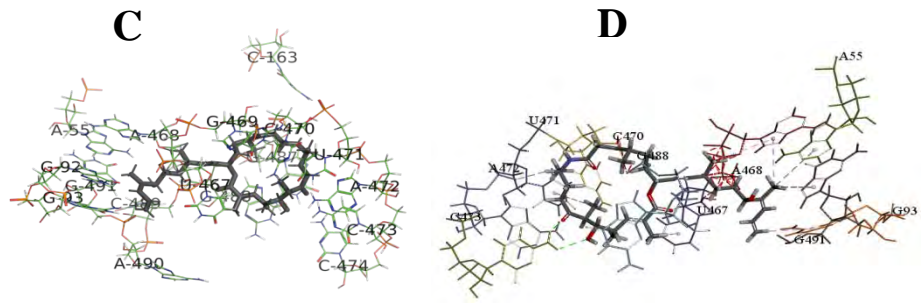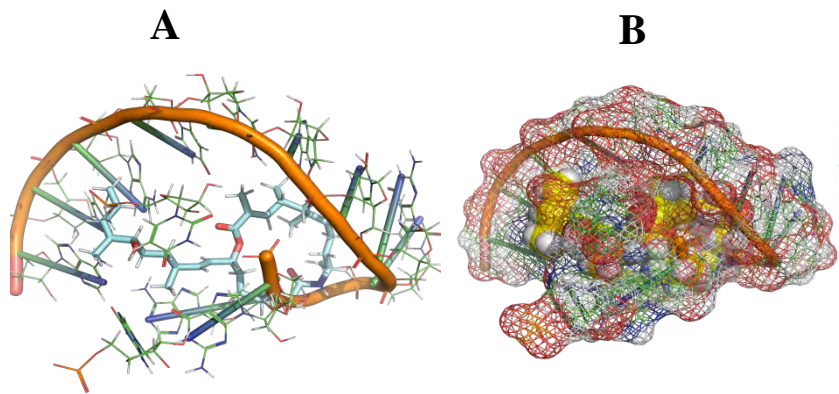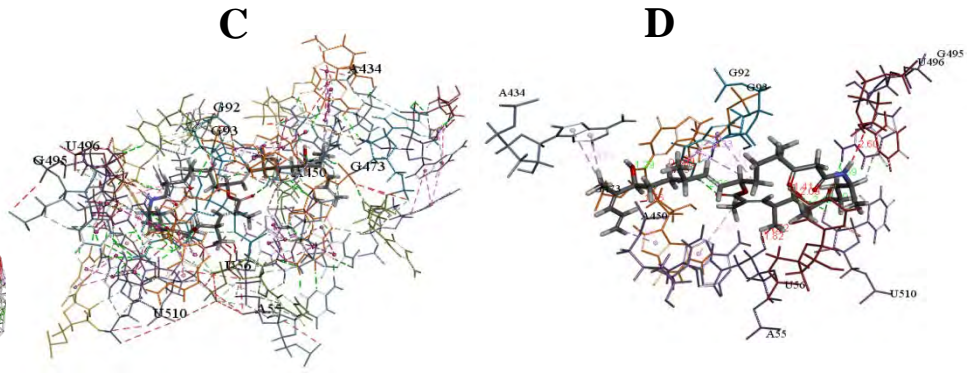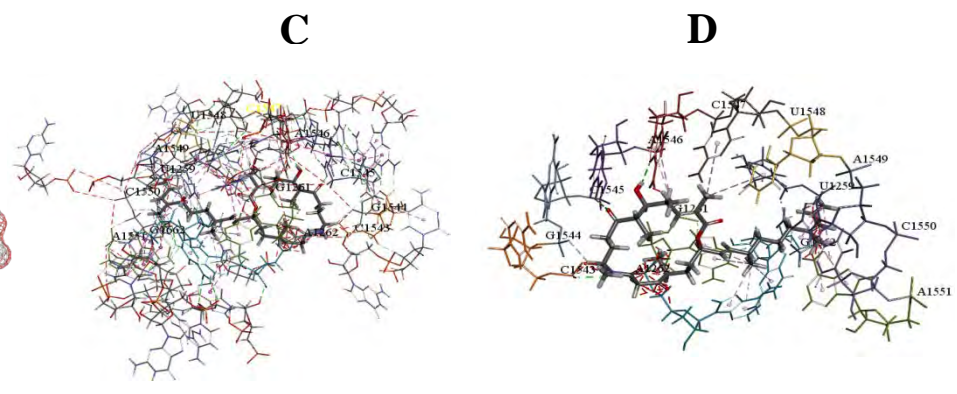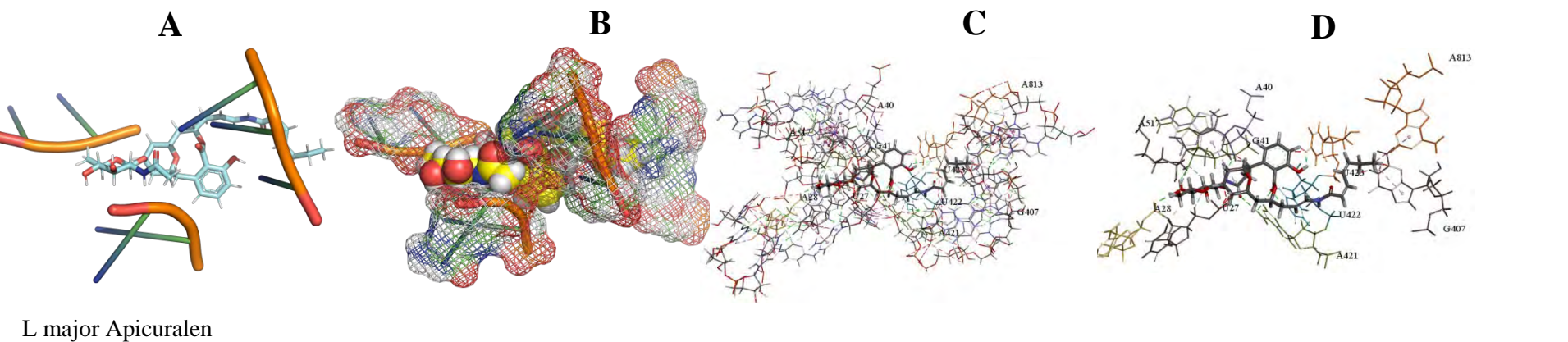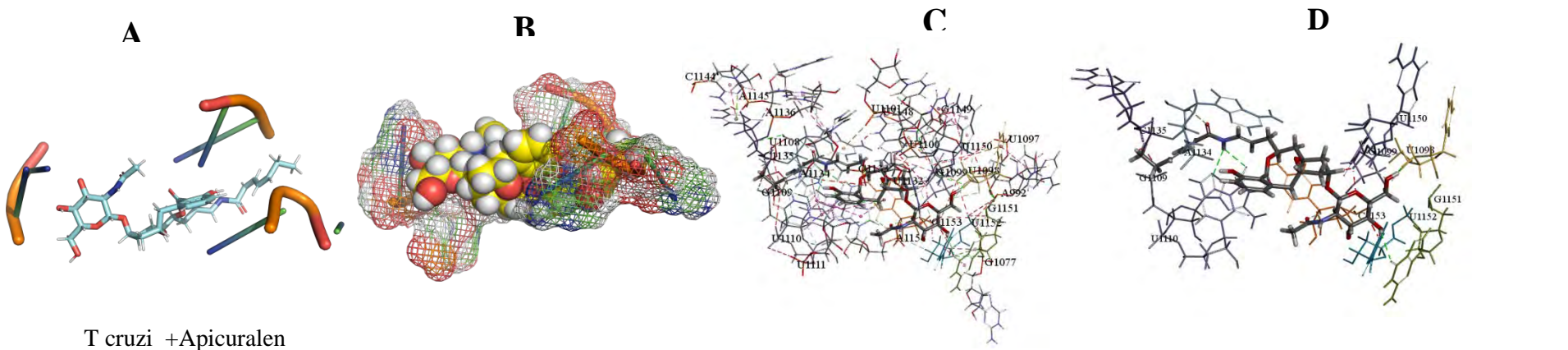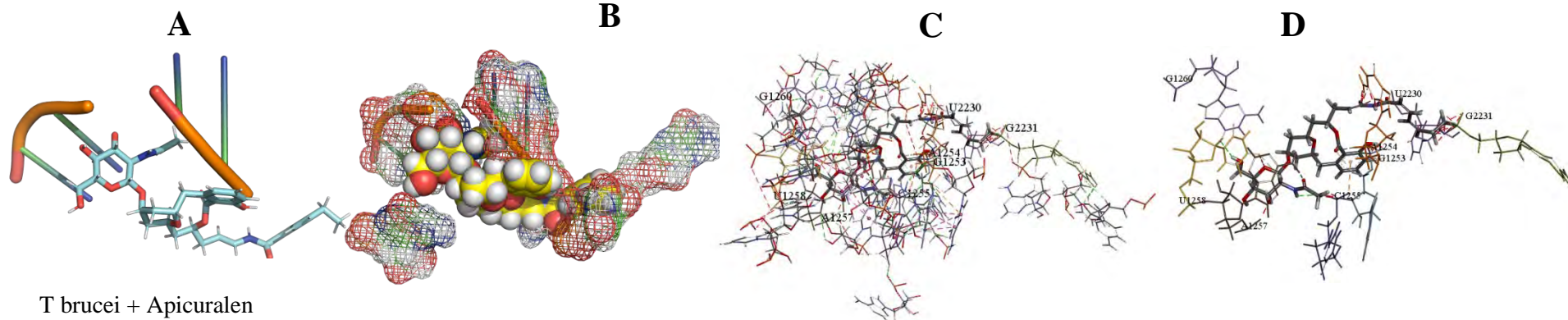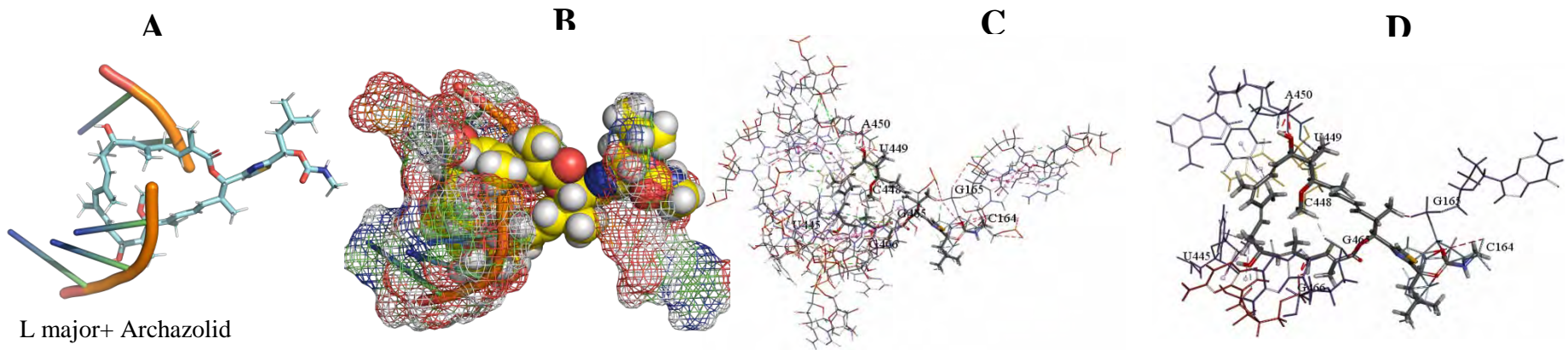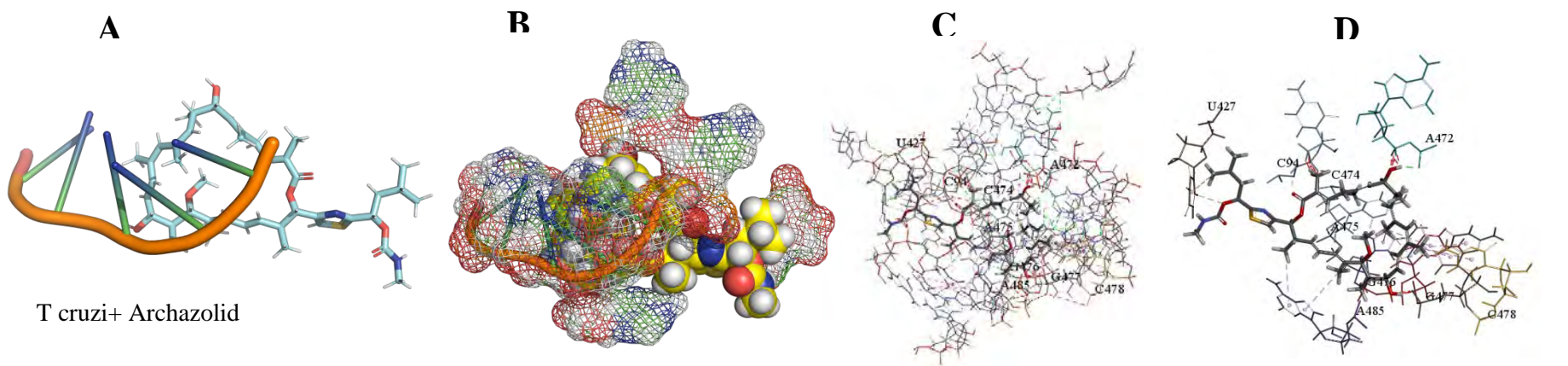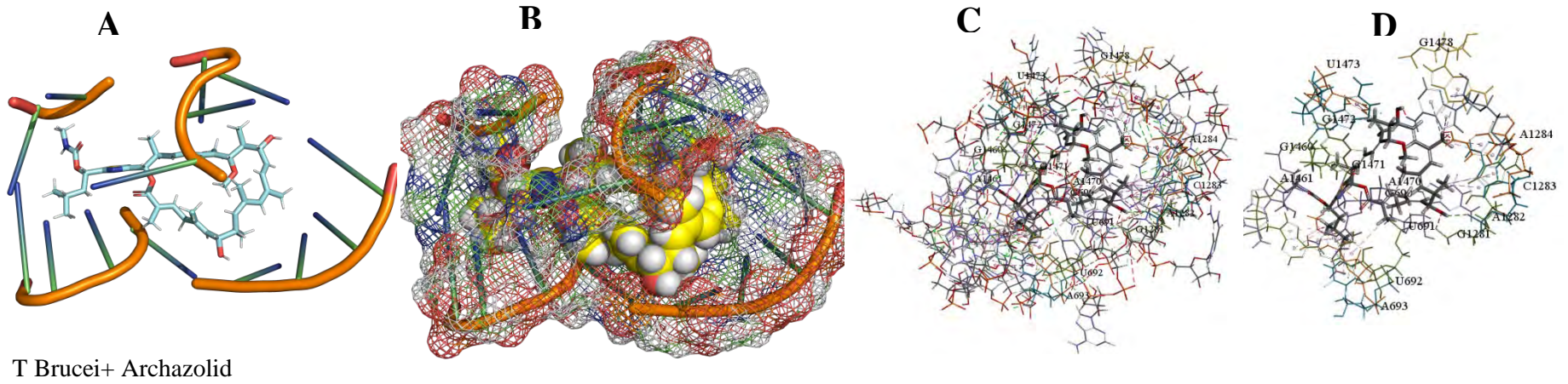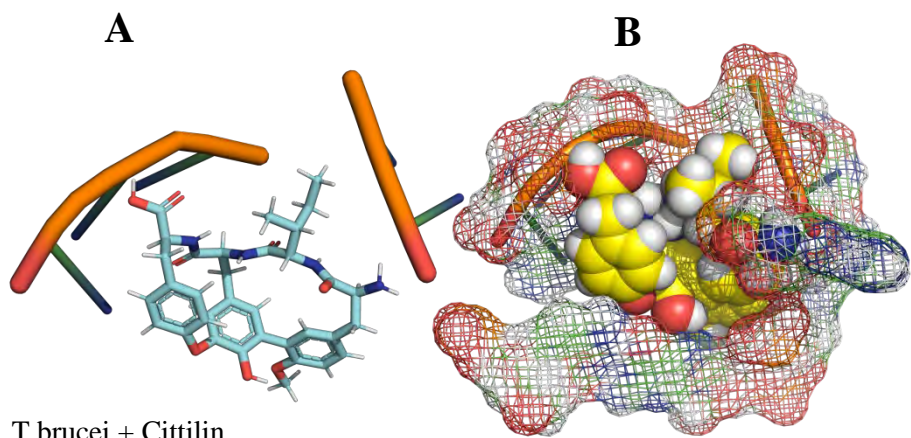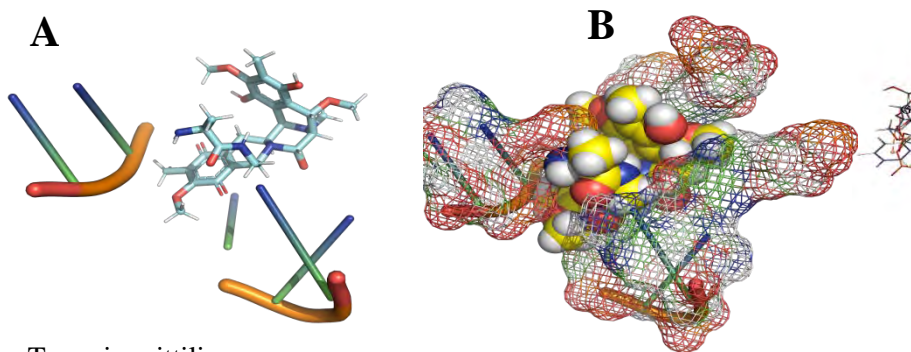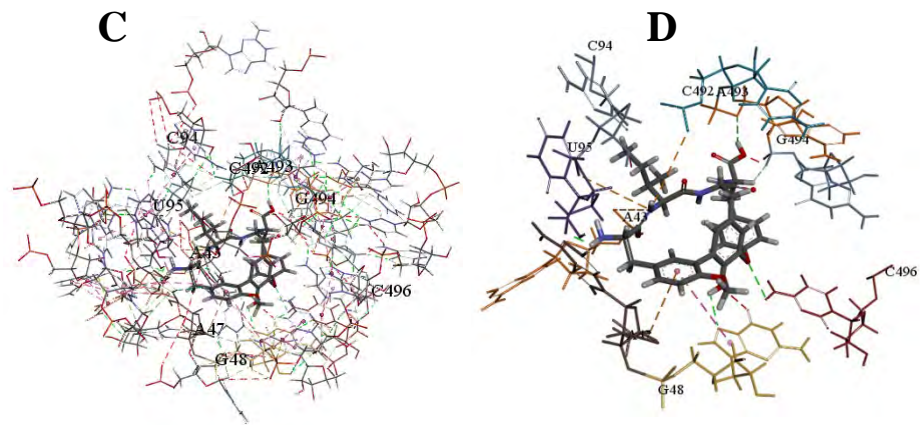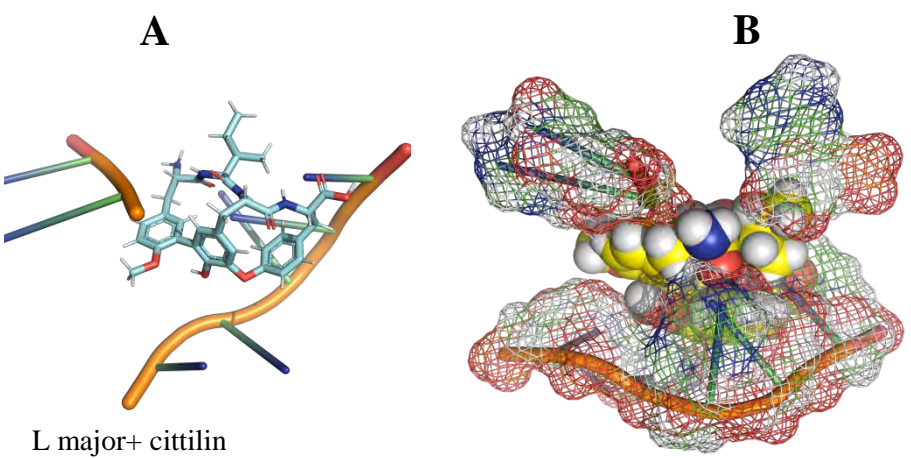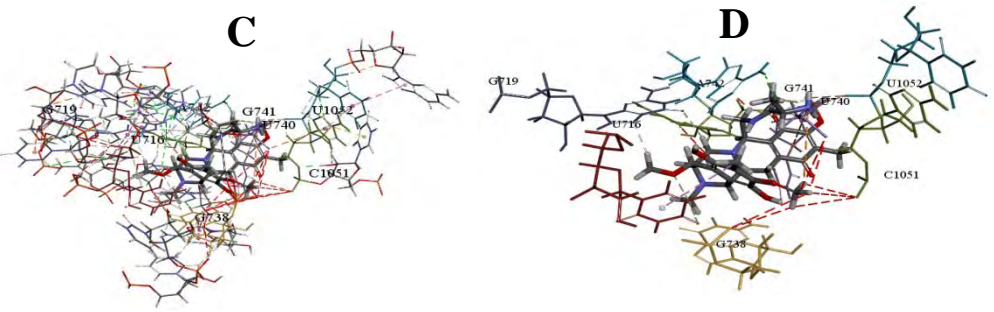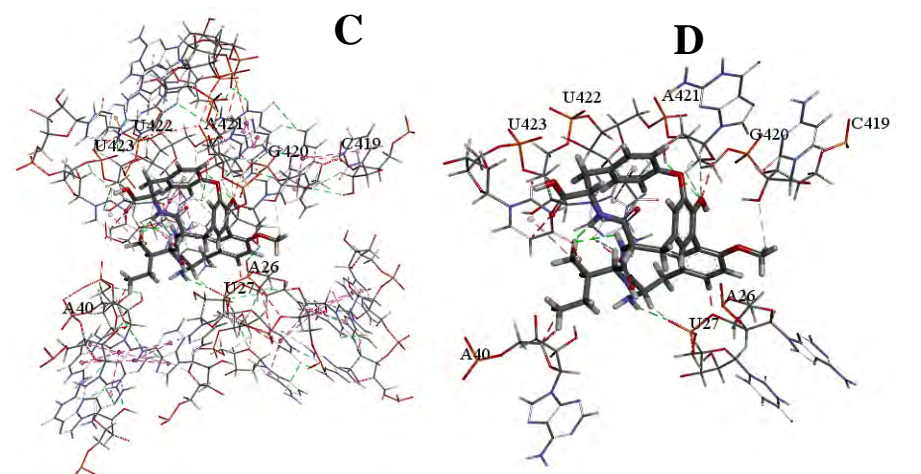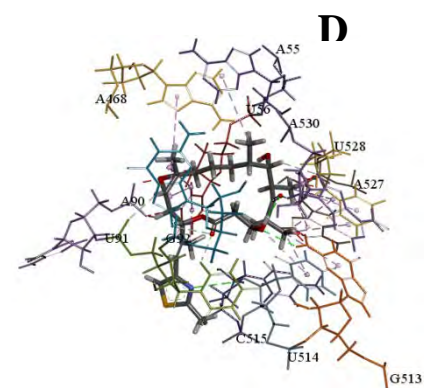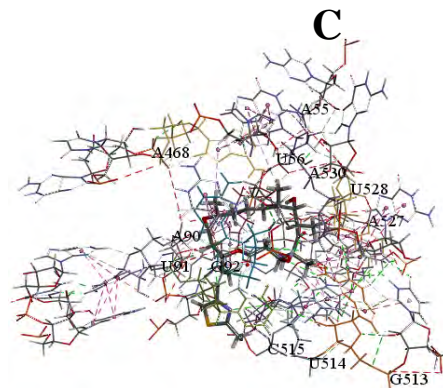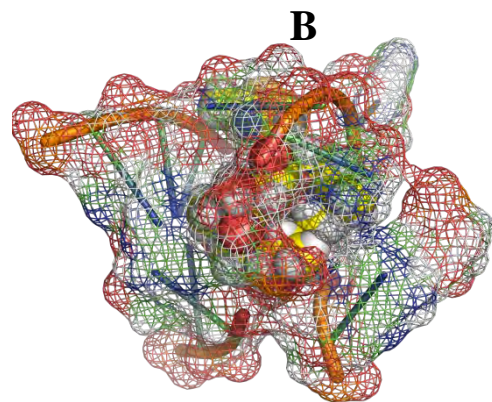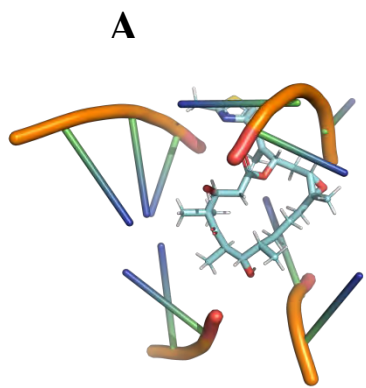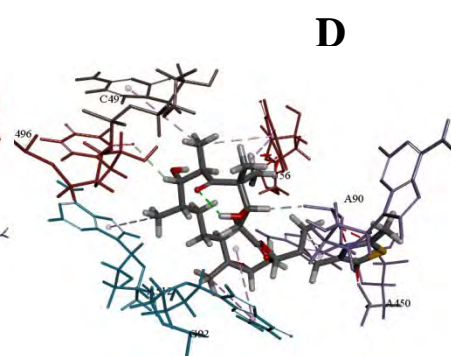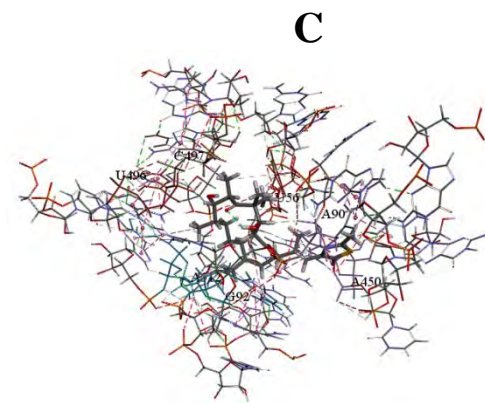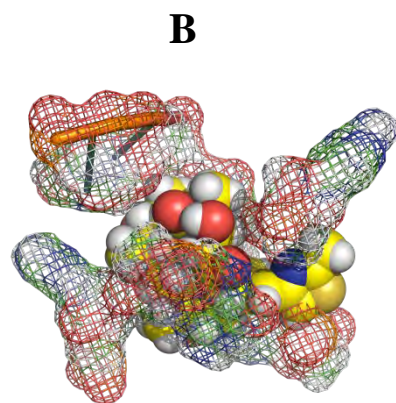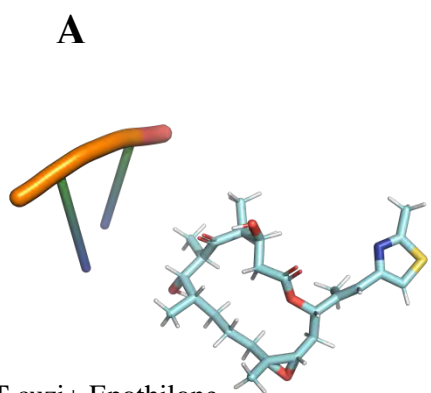
T Brucei+ Epothilone

T cuzi+ Epothilone

L major + Epothilone

A

T cruzi+ leupyrin

A

T Brucei+ leupyrin

A

L major+ leupyrin

A   B   C   D
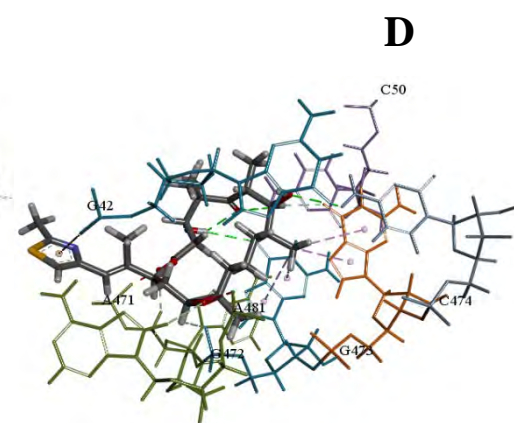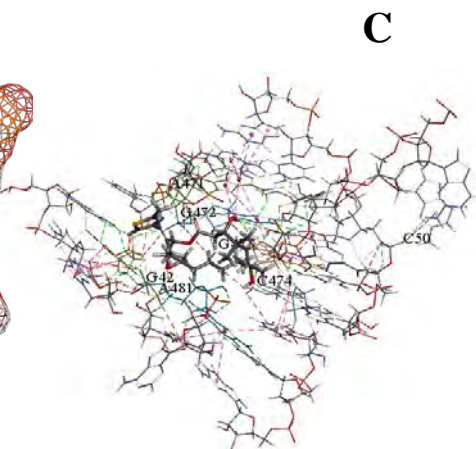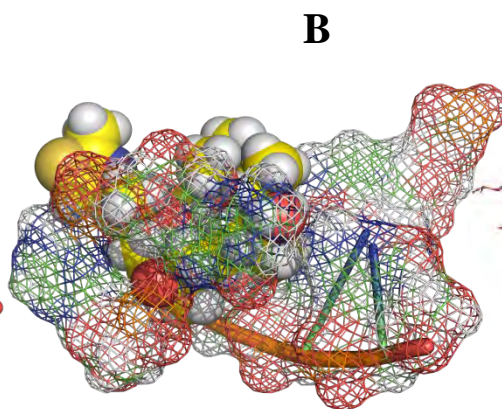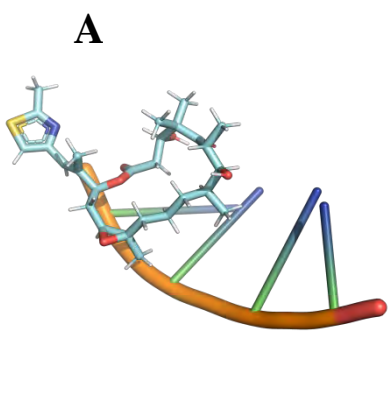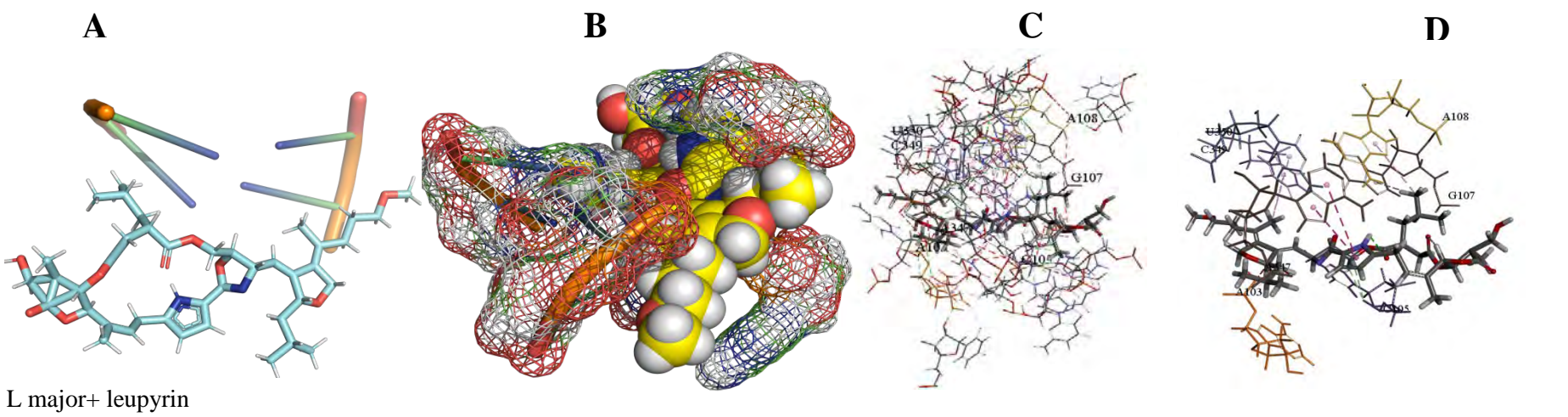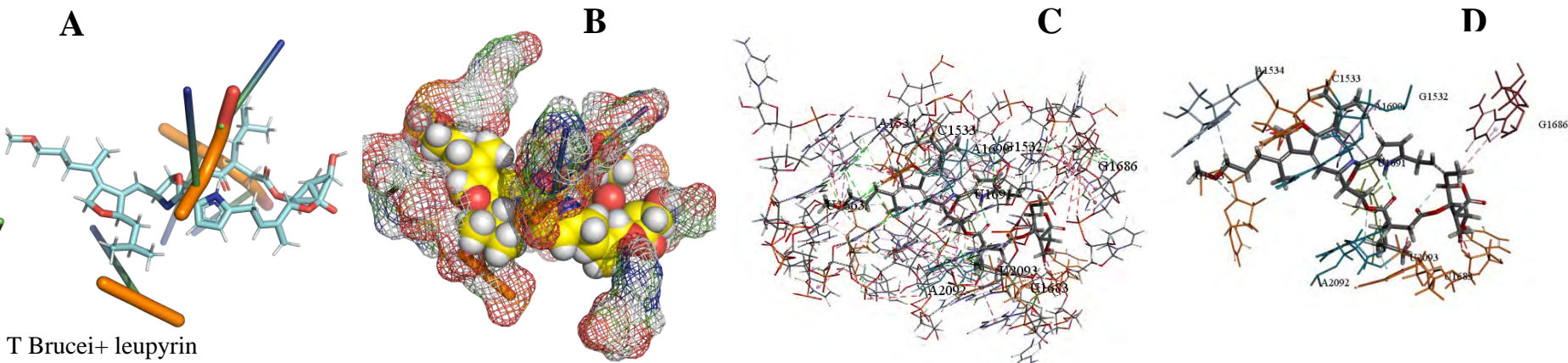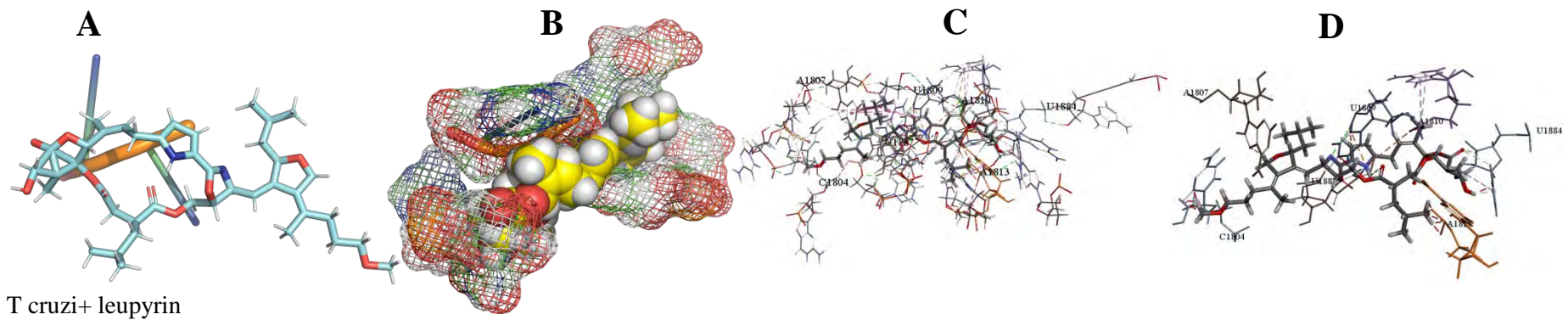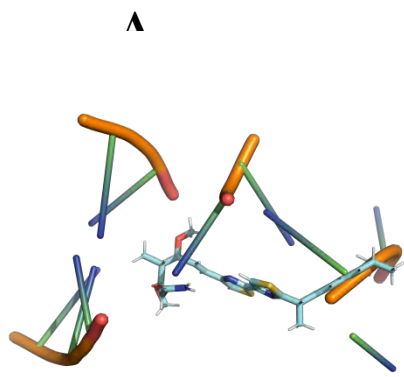
T Brucei+ Myxothiazol

A   B   C   D

T Cruzi+ Myxothiazol

A   B   C   D

L Major+ Myxothiazol

A

B

C

D

T Brucei+

A

B

C

D

L major + Sorangicin A

A

T Brucei+ Sulfangolid A

B

C

D

A

L Major + Sulfangolid A

B

C

D

A

B

C

D

T brucei+ mmv658993

A

B

C

D

T cruzi+ mmv658993

A

B

C

D

L Major + mmv658993

85

T brucei+ mmv676008

T cruzi+ mmv676008

L Major + mmv676008

A   B   C   D

T Brucei+ mmv688283

A   B   C   D

T cruzi+ mmv688283

A   B   C   D

L Major + mmv688283

T brucei + mmv688407

T Cruzi + mmv688407

L Major + mmv688407

A

B

C

D

T Brucei + mmv688474

A

B

C

D

T cruzi + mmv688474

A

B

C

D

L Major + mmv688474

6   Discussion

6.1   Sequence selection and analysis of the 18S rRNA of the selected kinetoplastids *Leishmania Major*, *Trypanosoma brucei,* and *Trypanosoma cruzi*

For one to select the important sequences from the gene bank to aid in the process of homology and de novo modeling the status of all the sequences were analyzed. These meant that all the sequences collected in various databases were measures in terms of completeness (Cannone et al., 2002, 2015) of each of the 18S rRNA sequences of each kinetoplastid selected. The three best-ranked sequences from verified peer-reviewed research databases were selected. In section 5.1.1 table 3 *leishmania major* sequences that were best-curated gene bank sequences were M81427 which had 882 nucleotides which were 40% in terms of completeness. This was followed by X53915 which had 2137 nucleotides and 97% complete. Finally was AC005806 which had 2203 nucleotides and was 100% complete. Sequence alignments for these three were performed to check on their similarity index and variations as shown in figure 13 using multi-align (Corpet, 1988; Edgar, 2004). The consensus sequence indicated the similarity index of the sequences but from the completeness, this gave the insight to pick the most complete to guide through the whole process of structure prediction. It's important to note that the 18S rRNA structure is very conserved with very minimal allowed changes in the expansion segments and since this is a single species not much variation is expected. AC005806 which had 2203 nucleotides and was 100% complete was selected after the rigorous analysis for the secondary structure prediction and eventual three-dimensional structure predictions.

6.2   *Trypanosoma brucei, Trypanosoma cruzi* and *Leishmania Major* Sequence Selection and alignment

The process performed in section 5.1.1 and explained in section 6.1 was repeated for section 5.1.2. For *Trypanosoma brucei* four sequences deposited in the gene bank and curated in other peer-reviewed databases were selected. These gene bank sequences in table 4 are M12676 which has 2251 nucleotides and is 100% complete, AJ009142 with 2188 nucleotides and 97% complete, AC005806 which has also 2188 nucleotides and 97% complete, and finally, AL929605 which is 100% complete and 2252 nucleotides. These four sequences were aligned using MEGA (Edgar, 2004) and MULTIALIGN (Corpet, 1988) to show the similarity and variation index. This again is from the understanding that the 18S rRNA for this species won't change much but it's great to see areas of changes within the sequences that would change translational conformations. Figure 14

shows *Trypanosoma brucei* alignments with the consensus which is indicative that although there are very minimal changes in some areas evolution changes could account for the single nucleotide polymorphism within the sequences. M12676 which has 2251 nucleotides and is 100% complete was again selected for this batch to further go to the process of secondary and tertiary structure prediction.

Table 5 shows the *Trypanosoma cruzi* sequences selected to undergo sequence alignments to determine the best sequence to be modeled. This kinetoplastid had more sequence data compared to the other two selected. Nine sequences were selected from the gene bank after they were analyzed and curated in various databases. Gene bank sequences AF228685 with 2234 nucleotides and 97% complete, X53917 with 2240 nucleotides and also 97% complete, AF245383 with 2240 nucleotides and 97% complete, AF239980 with 2246 nucleotides and 97% complete, AF239981 with 2253 nucleotides and 97% complete, AF245381 with 2246 nucleotides and 97% complete, AF245380 with 2248 nucleotides and 97% complete, AF245382 with 2315 nucleotides and 100% complete, and finally M31432 with 2319 nucleotides and 100% complete. From the alignments again there were very minimal changes showing that the sequences are highly similar with small differences because of the single nucleotide polymorphism and the different sequence lengths. After all the analysis AF245382 with 2315 nucleotides and 100% complete was selected to predict the secondary and the tertiary structure of their respective 18S rRNA.
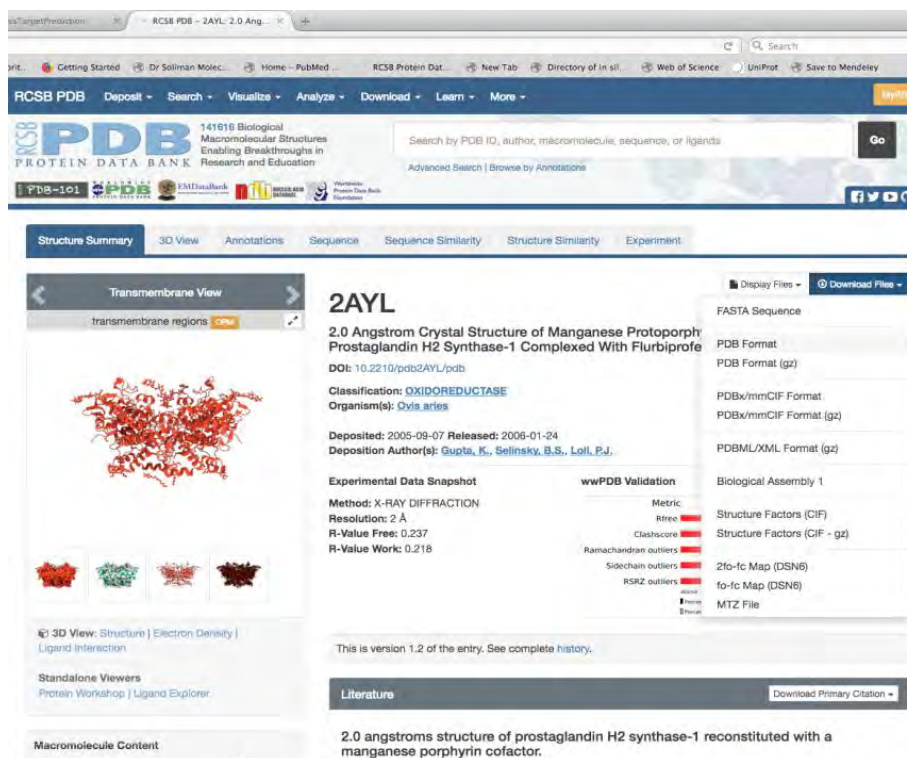
6.3   18S rRNA Secondary structure

The kinetoplastid three species sequences selected were taken through a rigorous exercise of determining their secondary structure using a software known as **Varna** (Darty, Denise, & Ponty, 2009); (version 3.93 http://varna.lri.fr/), **RNAstructure** (Reuter & Mathews, 2010); (Version 6.1) https://rna.urmc.rochester.edu/RNAstructure.html), **RNAComposer** (Popenda et al., 2012) (http://rnacomposer.cs.put.poznan.pl/), **RNApdbee** (Zok et al., 2018) (http://rnapdbee.cs.put.poznan.pl/), **xRNA** (http://rna.ucsc.edu/rnacenter/xrna/), **RNA2D3D** (Martinez, Maizel Jr, & Shapiro, 2008) (https://binkley2.ncifcrf.gov/users/bshapiro/rna2d3d/rna2d3d.html). Figure 17 shows the secondary structure of the 18S rRNA of *leishmania major* AC005806 as predicted and verified in the Comparative RNA Web (CRW) Site (Cannone et al., 2002). This is followed by the next secondary structure of *Trypanosoma brucei* M12676 which has 2251 nucleotides and 100% complete figure 18. Lastly is figure 19 which shows the secondary structure of *Trypanosoma cruzi* AF245382 with 2315 nucleotides and 100% complete.

6.4   Three Dimensional structures of the modeled kinetoplastids

6.4.1   Crystal Structures of Biological Targets

Just as with compound libraries, there are freely available libraries for 3D crystal structures of biological targets. These do not comprise strictly of proteins, as there are DNA structures, nanoparticles and peptides that have also been crystalized and are also available.

When searching for a target, the search box allows you to add a name of a compound, when you hit search, a large number of hits may be identified. To filter this hits and to make sure you have the correct target, the following criteria and filters should be met:

- The correct organism/taxonomy should be selected

- The correct strain (if organism)

- The experimental method should always be x-ray diffraction (however, NMR can be used if desperate).

- The resolution structure optimally should be below 2, this increases the accuracy of analyzing target.

- The target should be deposited onto PDB recently (date should be most recent)

These criteria will allow you to have an updated and accurate 3D crystal structure of the target, which can now be used for molecular docking and/or molecular dynamic simulations. As with most chemical reactions, there are many factors that permit a stable system. In protein data bank, there are crystalized targets with other small molecules at the active site and surrounded by solvent. The protein or nueclotide can then be taken to a software various softwares, where all modifications can be made.

6.4.2   Homology & de novo modelling and Active Site Identification

The initial step in molecular modeling and drug design is having a valid 3D structure, from X-ray crystallography, Nuclear Magnetic Resonance (NMR) or computational design using homology modeling. The aim of homology and de novo modeling is to predict a three-dimensional (3D) model of a biological structure from a template sequence based on the structure of one or more homologous proteins or Nucleic acids of which crystal assemblies have or have been not reported.

Homology and de novo modeling has played influential roles in many research areas and has aided in drug design by giving insights into spatial conformations and providing a structural template to construct novel drugs that are both specific and effective.

In order to generate a 3D model of a biological target, a general procedure is followed, with validation at each step as shown in the methodology

The Protein data bank holds information on 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease. The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and more.

The entire process of modeling each of the three kinetoplastid structure due to the complexity took some time and involved a number of steps. The first step involved finding the correct template from the growing number of probable available experimentally identified three-dimensional structures from curated databases. These databases include, but not limited to the RSCB Protein Data Bank (Berman et al., 2000). Templates for all the three structures kinetoplastids were obtained by blasting their sequences and checking if there was a most probable homolog of the query sequence used. For *Trypanosoma brucei* blast search using our query to be **M12676** which has 2251 nucleotides the best hit was **4V8M** template High-resolution cryo-electron microscopy

structure of the *Trypanosoma brucei* (Hashem et al., 2013). This structure was obtained from electron microscopy experimental technique with a Resolution of **5.57Å**. The multi percentile validation for the structure can be shown below indicating the challenges of choosing a template just because it has been obtained experimentary. This structure from the below result is not that accurate but still could be used as a template by firstly minimizing its energies further to obtain a more optimum structure that was used as the final template.
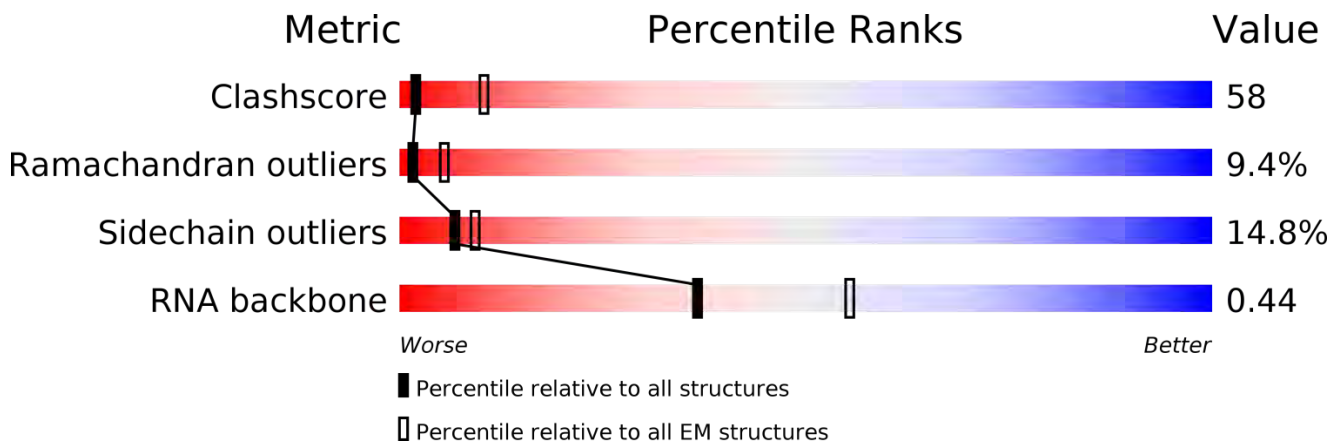
Figure 37 showing the graphic 4v8m_multipercentile_validation metrics ranging between 0-100 (Hashem et al., 2013).

The next in line was *Trypanosoma cruzi (*AF245382 with 2315 nucleotides and 100% complete) which was our second query sequence to try finding a suitable template. From our blast result, the best template structure was **5OPT** from the study **"**Structure of KSRP in context of Trypanosoma cruzi the 40S (https://www.rcsb.org/structure/5OPT) (Querido et al., 2017). This is a structure was obtained again from electron microscopy experimental technique with a resolution of **4 Å**. The overall validation results of the template are shown in the Figure below. Of keen interests, it's ok to choose the template as it is but to note that some of the parameters fall below a very good result. This might be because of some unresolved regions in the primary data that was obtained by the group. This leads to breaks in the eventual structure which are only shown in the three-dimensional structure by gaps in various domains. Importantly this structure has a very good clash score summary compared to the other two templates and this can be attributed to even the resolution.
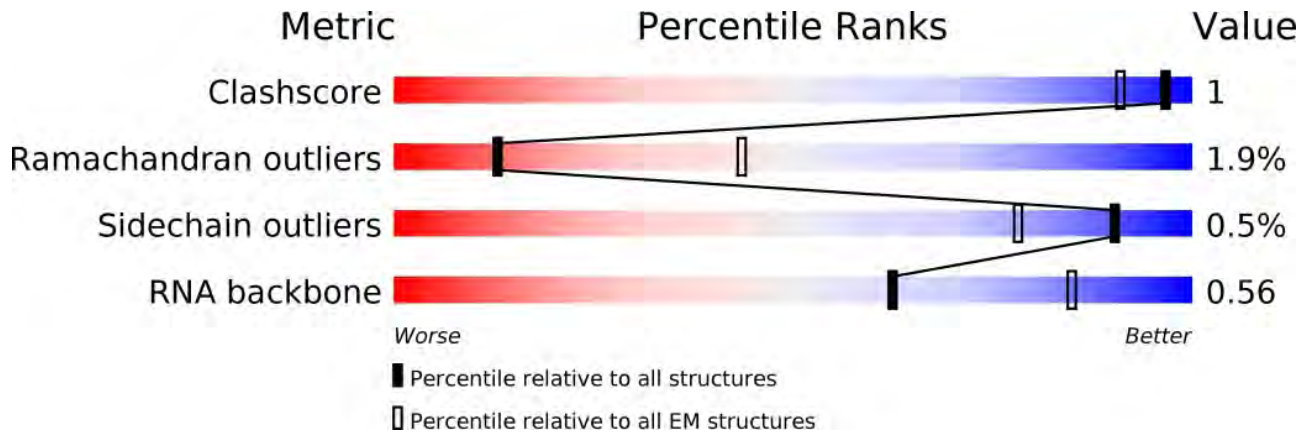
Figure 38 showing a graphic of the **5OPT** global percentile scores validation metrics ranging between 0-100 (Young et al., 2017).

Finally, *Leishmania major* (AC005806 which had 2203 nucleotides and was 100% complete) query against a blast search was performed. After a blast searches and structure analysis, the best template chosen was **5T2A** form the study "CryoEM structure of the Leishmania donovani 80S ribosome at **2.9** Angstrom resolution" (https://www.rcsb.org/structure/5T2A). This structure was obtained from an experimental Electron microscopy technique with a resolution of **2.9** Angstrom. This template is much better than the other two compared to the resolution but still, there were still challenges that appear on to it due to the Ramachandran outliers. From the template analysis, it had serious gaps and many regions that needed to be addressed before any other work was to be done.
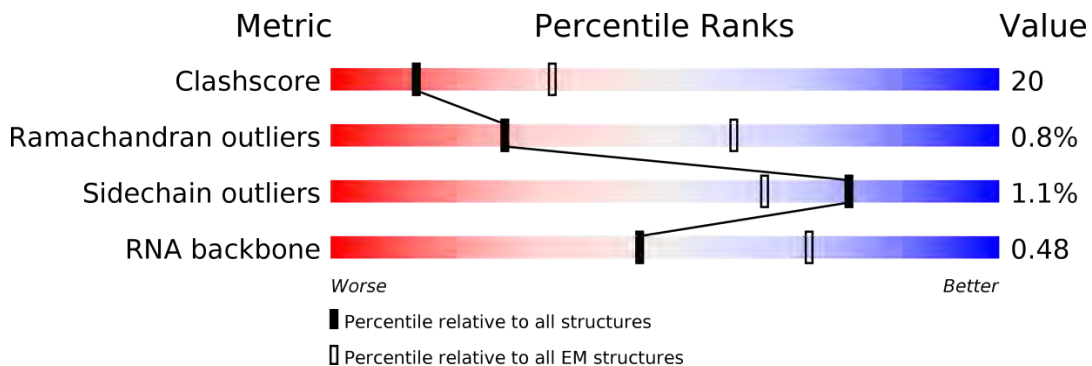
Figure 39 showing a graphic of the **5T2A** global percentile scores validation metrics ranging between 0-100

Nucleic acids and protein crystal structure collected data quality measure is referred to as its resolution (Berman et al., 2000; Wlodawer, Minor, Dauter, & Jaskolski, 2008). Identical aligned nucleotides within a crystal means, a precise flawless crystal, meaning similar X-rays scatter for all the nucleotides, and the crystal fine details are shown by the diffraction pattern (Berman et al., 2000). Alternatively, a slight difference in all nucleotides in the crystal, attributed to local

flexibility or motion, the pattern of diffraction will not comprise as ample fine data. Therefore, the measure of the level of detail that will be seen when the electron density map is calculated and the level of detail present in the diffraction configuration is its resolution (S. Chen et al., 2013). Structures with high resolutions of values of <1 Å or so are highly well-arranged and it's easy to appreciate every atom in the electron density map (Berman et al., 2000). Structures with >4 Å are considered to be lower resolutions structures, presenting only rudimentary contours of the nucleotides chain and the atomic structure is requisite to be inferred (Berman et al., 2000; S. Chen et al., 2013). Most nucleotides structures defined crystallographically fall in the midst of these two extremes. As a universal rule of thumb, we have more assurance in the position of atoms in structures with resolution values that are small, called "high-resolution structures" which is not the case to our template when it comes to all the template that was used.

Due to the complexity of the ribosome with regard to the 18S rRNA, it is highly conserved due to its conserved roles during translation. However, this may differ between species in specific areas known as the expansion segments(Panek, Kolář, Vohradský, & Shivaya Valášek, 2013). From experimental evidence shown in various researches conducted the way ribosomal proteins interact with the 18S rRNA segments, and the interactions with the mRNA, these areas are involved in the regulatory process of translation (Mwangi et al., 2017a; Panek et al., 2013; Scaiola et al., 2018). This difference could be mean that during the translation of various messengers RNA some species could translate them to protein others cannot leading to the uniqueness of diseases caused by this class of eukaryotes known as kinetoplastids. As shown in figure 17, 18 and 19 the secondary structures of all the three chosen kinetoplastid have very minimal differences even to the expected varying regions.

The three kinetoplastids were modeled by segmenting them into the four domains which ensure more accuracy compared to modeling as single units. These four units are thereafter combined to single structures of a complex 18S rRNA. It's of importance to note the reason we modeled each of the domains independently was that each of the domains folds differently (Mwangi, Gitonga, Wagacha, Sijenyi, & Mulaa, 2018; Mwangi et al., 2017a). The successful process of modeling involved achieving a similarity index between the template secondary structures with the query structures. Presence of mismatches regions between the queries and the templates was addressed by picking the mismatched region in the query and fitting them to the template matches. This was a more logical step in the process as it ensured that the process of modeling is comparative. After that process, both the template and the queries were cut at similar points. This segments cut into domains were modeled where the template helped develop coordinates for the queries for the new

models. This process ensured that during the whole process of modeling using the *RNA123* suite did not have gaps that would bring problems in the entire process of homology modeling. In figure 20(*Leishmania major*), 22(*Trypanosoma brucei*) and 24(*Trypanosoma cruzi*) Architectural tertiary structure of the 18S rRNA front and back view are shown colored differently depending with domains (5' major –red, Central-green, 3'major-blue, and 3'minor-yellow). From the analysis, the superimposed new models with their template appeared to have some major differences which are expected due to the differences between them and their respective templates and the difference in their primary sequence alignments shown in figure 13, 14 and 15. *RNA 123* had the ability to perform deletions, insertions and also cater for an increase in nucleotide base number compared to the templates crystal structures which ensured successful modeling was achieved. Further evaluations for the model such as calculations additional force fields that ensured minimal energy structures that could infer their biological functions. The total combination of all optimized domains of every kinetoplastid was combined still using *RNA123* successfully and again the energies were further optimized by running different force fields algorithms that reduce energies of the structure to the bare minimum to allow it to conform to its functional biological morphology. Tables 6, 7, and 8 show the18S rRNA Energy optimization, standard energies at various points compared to the optimized energies of the optimized structure. Notably, there are significant differences between the energies inter and the standard total energies of the structures. Finally figures 21, 23, 25, show a graphical representation of energy minimization through an optimization process in RNA123 a large positive figure is compressed to an acceptable negative figure that is biologically functional (Mwangi et al., 2018; Mwangi, Wagacha, Mathenge, Sijenyi, & Mulaa, 2017b).

From the tables 6, the three-dimensional structure of standard versus optimized structure *Leishmania major* shows a very significant change in the total inter energy from a positive to a negative figure, the total intra energy has a very slight change, the gap penalty is slightly reduced and the restraint energy increased from initial zero. For *Trypanosoma brucei* table 7 a big change of the total inter energies, the total intra energies changes from a high to a very low figure and lastly the restrain energy increases significantly from no initial value for the two structures. Lastly for figure 8 standard and optimized structure of *Trypanosoma cruzi* we see significant changes between the higher energy to a minimized negative figure, the total intra energy moves from a higher figure to a negative figure whereas the restrain energy again increases to help optimize the structure. The gap penalty of all is slightly reduced significantly. From all this data we achieve a

more negative structure for all the three kinetoplastids which undergo a further process of validation so as to be used for the next objective of screening.

The homology and de novo structures of the kinetoplastids selected in this study provide invaluable insights into the evolutionary process of this eukaryotic disease-causing organism and into their functions. The 18S rRNA structure being the architectural component of the 40S subunit that is key in the process of translation performs various multiple roles. This process involves regulation and expression of genes, coding, and decoding within the cells. Just from understanding the important components of the 40S which are made up of single-stranded nucleic acids chain fold alongside itself and then highly complexed onto itself forming the 18S rRNA, it's important to see deeper and its functional insights. The 18S rRNA is synthesized in the nucleolus and is the catalysis component of the ribosome. Recent studies have shown that the ribosomal RNA could be exploited to exhaustively characterize the process of synthesis, but still more work in structure determination has been done (Mwangi et al., 2018, 2017). More has been suggested involving exploiting the 18S rRNA structure as a drug target for specific organisms such as the selected kinetoplastid. With the good document conserved loops that could be used to dock and bind specific ligands that would stop the basic process of messenger RNA being translated (Mwangi et al., 2018). Just by having high resolutions structures of this selected kinetoplastids the specific determined of loops were used in the next process of screening. This was done to try finding newer leads towards the treatment of this tropical neglected disease-causing agents that have a major economical implication towards human and animals.

6.5   Compounds Libraries

Information about different metabolites from plants, fungus and bacteria were collected from the two main literature databases. The myxobacterial metabolites databases (https://hzi.openrepository.com/handle/10033/122086) and the pathogen box databases (https://www.mmv.org/mmv-open/pathogen-box/about-pathogen-box). Thanks to the Medicines for Malaria Venture (MMV) science group and reports available for neglected disease compounds and the Helmholtz Institute for Pharmaceutical Research and Department of Pharmaceutical Biotechnology in Germany for Myxobacterial secondary metabolites.

6.5.1   Pathogen Box Metabolites

MMV provides a pathogen box database available for free for 400 diverse, drug-like molecules active against neglected diseases of interest. This pathogen box database offers the 400 compounds upon request to researchers interested in drug development. They come in 96 plate wells, which

contain each unique compound in 10μL of a 10mM dimethyl sulfoxide (DMSO) solution. This offers a collaborative front towards developing newer intervenes to the neglected diseases. The only requirement from this group is to make available data generated from research using the database compounds with the public within two years. This in its eventuality creates a forum towards collaborative efforts to solve the neglected disease eradication in tropical regions where they are a menace affecting growth.

This study performed several screening experiments using the three modeled structure of kinetoplastids to try evaluating their activity on the process of the kinetoplastid protein synthesis. This suggests a possible mode of actions towards the formation of the small ribosomal subunit assembly towards its role when it comes to initiation of the translation process on the mRNA.

Total of 400 compounds listed from the library was used against the three 18S rRNA kinetoplastid structures. From the results provided in table 9 above the 20 best compounds are provided. Their literature and SMILES string files of these compounds were obtained from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/). **Online SMILES Translator and Structure File Generator** (https://cactus.nci.nih.gov/translate/) server was used to convert the SMILES strings to three-dimensional structures shown in figure 26 above. This figure shows the SMILE file of each compound and its three-dimensional structures that were used in the next process of docking.

6.5.2   Myxobacterial Secondary Metabolites Databases

This is a growing source of secondary metabolites that were obtained from gram-negative proteobacteria (Weissman & Müller, 2010).This bacteria have a wide range of known habitats that include the decaying plant material, soil, tree barks, marine environment and herbivore dung (Weissman & Müller, 2010). These naturally occurring microbes have several distinct characteristic behaviors such as moving in solid surfaces by creeping and gliding as amoebas do to differentiate them from other bacteria's(Weissman & Müller, 2010; Whitworth, 2008). In addition to this, they are known rich producers of natural secondary metabolites products by virtue of their metabolism (i.e., *Bacillus species*, actinomycetes, Pseudomonads, and fungi) (Dehhaghi, Mohammadipanah, & Guillemin, 2018; Wenzel & Mueller, 2009).

Closely to 7500 strains of myxobacterial have yielded at least 100 distinct core structures to date only having a portion of this (67) reported in primary literature (Hoffmann et al., 2018; Weissman & Müller, 2010) and over 500 chemical derivatives (Garcia, Krug, & Müller, 2009; Hoffmann et al., 2018). Most Myxobacterial metabolites structures are non-ribosomal polypeptides, their

hybrids, polyketides, phenylpropanoids, alkaloids and terpenoids (Dehhaghi et al., 2018; Nett & König, 2007; Weissman & Müller, 2010). Many strains of Myxobacterial metabolites belong to multiple structural classes in addition to the number of chemical variations on each scaffold (Dehhaghi et al., 2018; Weissman & Müller, 2010). Furthermore, many of the natural products reveal distinctive structural topographies comparative to compounds known from other microorganisms (Dehhaghi et al., 2018).

This study screened a total of 400 compounds using the three highly resolved structures of the kinetoplastids selected to show their probability to be candidate compounds towards solving the neglected tropical diseases. Presented in table 10 are the best twenty compounds from the binding and docking results. This table 10 shows the Compounds of myxobacteria, their smiles, molecular formulae, molecular weight the commercial supplier the chemical id, those that have trivial names and the compound ID from PUBCHEM.

As mentioned in the other database the literature and SMILES string files of these 20 compounds were obtained from the PUBCHEM database (https://pubchem.ncbi.nlm.nih.gov/). **Online SMILES Translator and Structure File Generator** (https://cactus.nci.nih.gov/translate/) server was used to convert the SMILES strings to three dimensional structures shown in figure 27 above. This figure shows the SMILE file of each compound the two dimensional and it's three dimensional structures that were used in the next process of docking.

6.5.3   Predictive Pharmacokinetic Analysis

When designing a drug, it is in some studies very important to understand the pharmacokinetic and physiochemical properties of the drug prior to experimental testing. This, again, assists in saving both time and money and allows the researcher to optimize a drug is the above-mentioned characteristics are not favorable.

SWISS-ADME is a validated, free-available, online tool that is utilized in the prediction of both pharmacokinetic and physiochemical properties of a compound, thus measuring the drug-likeness of molecules still in the design process.

The online tool, which takes into account the "Boiled egg hypothesis": *Brain or IntestinaL EstimateD permeation* method (BOILED-Egg), an accurate predictive model that works by computing the lipophilicity and polarity of small molecules. Concomitant predictions for both brain and intestinal permeation are obtained from the same two physicochemical descriptors and

straightforwardly translated into molecular design, owing to the speed, accuracy, conceptual simplicity and clear graphical output of the model.

6.5.4   Identifying Potential Targets of a Compound

In the process of ligand-based drug design, the biological target of the compound is, in most instances, unknown. It is therefore vital to identify the biological target prior to understanding the mechanism of action of the molecule.

The most popular cases for the use of this technique is the study of natural products. In Africa, traditional medicine is widely used on a daily basis. Researchers are now teaming up with traditional healers to institutionalize herbal medicine and elaborate on the scientific evidence of certain therapeutic plants or plant combinations.

If the herbal remedy is said to contain anticancer properties, further evaluation can be done to identify the specific phytochemical that portrays these inhibitory characteristics as well as its biological target. There are many plant extracts that may have multiple therapeutic effects, including anticancer, antihypertensive and antidiabetic properties. In this instance, the phytochemicals of the extract may be extracted, the chemical structures determined, and the biological target may be predicted. This computer-aided drug design technique may take a few minutes at no cost, compared to protein analysis of each enzyme within a metabolic pathway. With this being said, it is always important to validate these predictive techniques with experimental analysis.

6.6   Binding and Docking Results

6.6.1   Selected Kinetoplastid 18S rRNA structure preparation

The present study used the predicted structures of the 18S rRNA of the selected kinetoplastids *leishmania major* (M12676)*, Trypanosoma brucei* (M12676)*, Trypanosoma cruzi* (AF245382)

The molecular docking results gave several consensus scoring value functions, which estimate the binding energies of study substances (Myxobacteria metabolites & Pathogen Box compounds) with the 18S rRNA target obtained from both schrodinger and Accerlys discovery suite as shown in table 2. The binding affinities in terms of the binding energies is shown as the atamic contact energies (ACE). The low values for the binding energies as per the software suites of the compounds docked to the 18S rRNA active motif sites gives a ligand pose in the actual active

bindidng site.the binding site is an area where the hydrophobic fragment of the compound is engrossed as shown in the images of the various poses below.

6.6.2 Binding Site Identification

In some studies, the drug biding pocket on the biological target is unknown. As a result of this, it becomes impossible to dock compounds and further progress through the rational drug design process.

There are a huge number of online tools that are available to identify an active site from aprotein, however, one of the most validated and popular tools is Metapocket (Further reading in publications). Metapocket uses 8 different algorithms to identify ligand binding sites by computing interactions between a chemical probe and a protein structure. The input is a PDB file of a protein structure, the output is a list of "interaction energy clusters" corresponding to putative binding sites

6.6.3 Computer-aided drug design can be broadly classified into 2 main subgroups:

(1) Structure-based drug design: this method assumes that the structure of a biological target is known (the protein/DNA has been crystallized or a 3D model of the target is built). Compounds are then designed/screened to fit the structural characteristics of the target, thus rendering strong molecular interactions that stabilize the compound at the targets binding site. This technique is the most widely used in computational chemistry and yields a plethora of potential compounds that may then be screened for activity.

(2) Ligand-based drug design: this method assumes that only the structure of the drug is known and there is absence of the 3D biological target. Optimized compounds are then designed based on the knowledge the drug's chemical analogs and their biological activity. Quantitative structure activity relationship (QSAR) features are designed based on physiochemical attributes of a set of chosen analogs and their biological activity with a target molecule. These QSAR features are then used as a template to screen for potential compounds with more favorable characteristics. Computational tools are now also available to predict potential targets of a compound prior to QSAR analysis.

Pharmacophore-based drug design, which implements aspects of both structure and ligand-based design, is an optimized and more accurate method of identifying optimized lead molecules. This method requires the 3D crystal structure of the target, as well as the structure activity relationship of the compound at the binding site of the target to be known. Once the intermolecular forces

between the compound and target has been established, a pharmacophore model can be generated (Pharmacophore- minimum number of atoms in a compound that is required to induce a biological response).

This pharmacophoric model/ scaffold, containing only vital molecular moieties, is then used to screen chemical libraries to identify potential lead molecules.

flow chart describing the stages of pharmacophore-based drug design is depicted below:
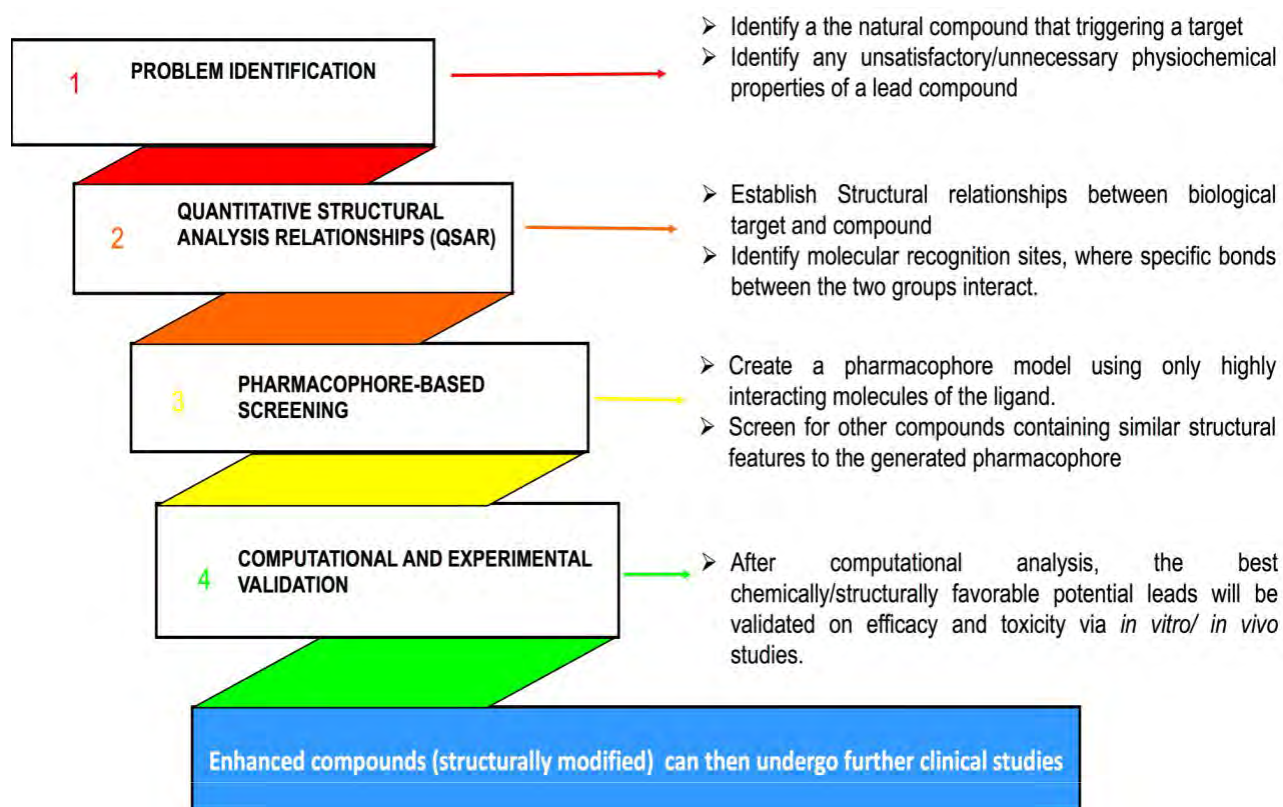
Over recent decades, the pharmaceutical industry has implemented computational drug design to save on both time and cost when researching optimized drugs. Most of the available drugs on the market originate from natural sources and most research has circled back to identifying active compounds from nature's biological organisms. Computational-aided drug design has proven to be a vital component of the discovery of bioactive compounds from natural sources:
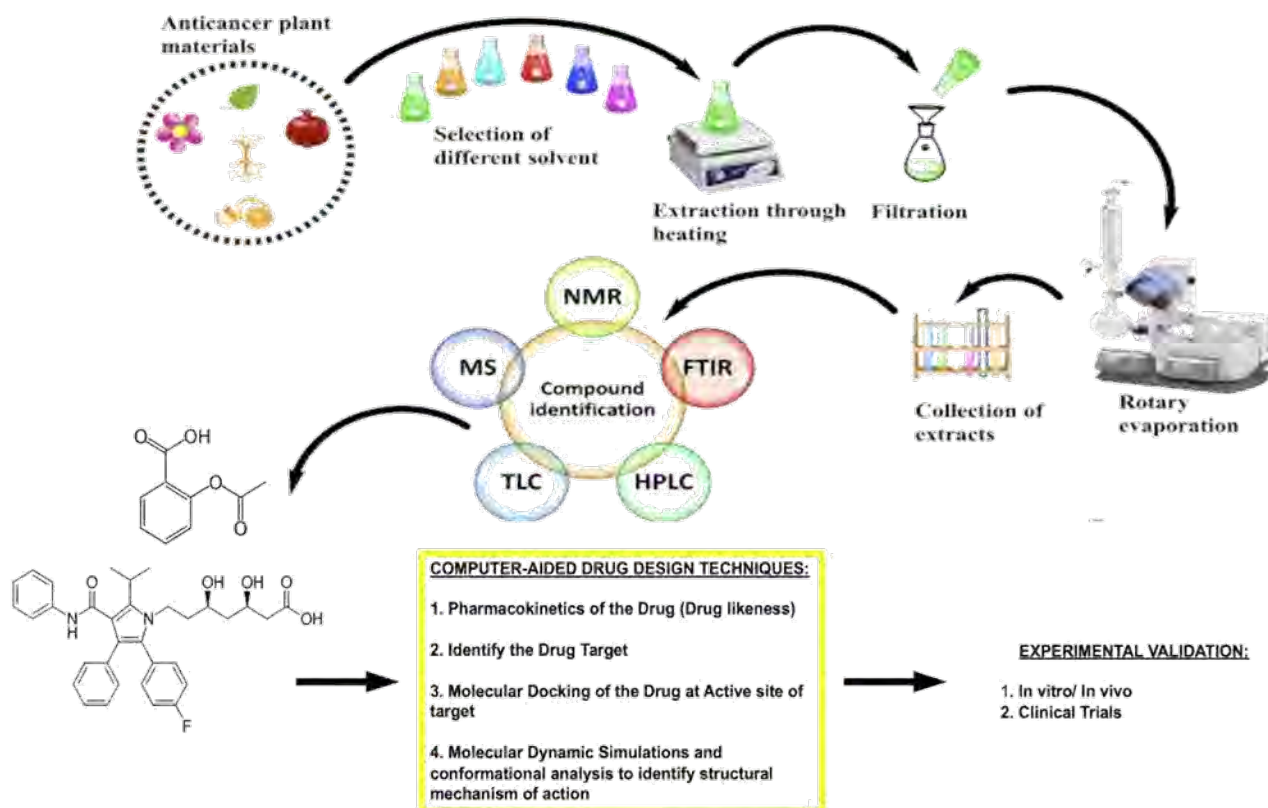
In the our study, we elaborated further on these computational tools by identifying software and procedures that will allow a researcher to progress through the rational drug design timeline.

6.6.4 Molecular modelling and optimization

Molecular modeling is an investigation of structures and molecular properties by using techniques of computational chemistry and graphic visualization. Knowledge on specific stereochemistry of target active sites justifies the conformational study of drugs, which can interact with these sites. The drug-target interaction produces a conformational change that is ultimately observed as a pharmacological response. A target site can only bind a single compound conformation from many possible geometric fits. This pharmacophoric conformation has all the atoms of the drug correctly positioned in space to fit the target.

Conformational isomerism is defined as the non-identical spatial arrangement of the atoms in the molecule. It is the result of rotation through one or more single bonds. In conformational analysis of molecules, the bond rotation changes the dihedral angles and, consequently, the corresponding steric energy due to spatial overlaying of non-linked atoms and rotation torsion barriers. The

molecules drawn three-dimensionally are not necessarily in the most stable conformation. Generating a certain structure causes molecular distortions with unfavorable lengths, angles, and dihedral angles. Non-linked atoms also interreact in the same spatial regions generating steric and electrostatic repulsion. Correction of the molecule distortions may be achieved by energy minimization through the use molecular mechanics forcefields. These forcefields arrange the atoms within a molecule to be "geometrically comfortable", thus minimizing steric hinderance and unfavorable electrostatic energy.

There are 3 main forcefields used to optimize drugs:

Ghemical: Ghemical is a forcefield that performs energy optimizations by providing accurate geometries to organic-like molecules.

MMFF94 and MMFF94s: The MMFF94 and MMFF94S were designed by Merk and optimize geometries, bond lengths and angles, as well as electrostatic and hydrogen bonding. The MMFF94s (static) varies slightly different out-of-plane bending and dihedral torsion parameters. This provides a better match to the time-average molecular geometry in solution or crystal structures. The MMFF94 and MMFF94S methods provide good accuracy across a range of organic and *drug-like* molecules thus allowing it to give better final energy optimizations compared to that of the chemical forcefield.

6.6.5   Molecular Docking

One of the most popular tools utilized in computational drug design is molecular docking. The technique of molecular docking makes use of a multiple methods in the prediction of binding affinity and configuration of a complex. Ligand-receptor complexes exemplify the most general use of docking, although there are numerous studies that demonstrate protein-protein complex or drug delivery complexes such as nanoparticles or aptamers.

There are two main steps involved in docking:

3. Sampling conformations of a ligand in the active site of protein or nucleotide- different algorithms may be used when sampling the numerous conformations of the docked complex: the "lock and key" model which describes the ligand and receptor as rigid

structures, or the ligand may be flexible either through random or simulation-based methods. The latter algorithm is the most commonly used method as it allows for a more realistic fit of the ligand to the protein.

4. Ranking the different conformations by scoring function- the scoring function may be based on statistically preferred contacts, MM force fields or pre-existing protein-ligand binding affinities.

Over the past decade, there has been flood of molecular docking related publications and although these papers may add to the structural information about a biological target or new lead compound, there are still many inconsistencies that arise. Frequent criticism associated with docking includes incorrect binding sites, choice of docked complex (conformational pose) and choice of small molecule (inhibitor or agonist).

Due to these concerns, all docked complexes should be verified with MD simulations or experimental analysis.

There are many different tools that can be used for molecular docking with popular examples being Autodock, Shrodinger Maestro with Glide, GOLD and Discovery studio.

6.7   Visualization of Ligand-Protein Interaction

Molecular visualization is a key aspect of the analysis and communication of modeling studies. It enables a mechanistic understanding of a molecule's structure to be visualized, so that key insights can be shared between computational modeling experts and collaborating team members. There are many software's for viewing Nucleic acids/protein-ligand interactions. In this study we used Biovia Discovery studio and Schrodinger's tool.

CHAPTER SEVEN

7   Conclusions and Recommendations

Advances in biomedical research specifically in the area of tropical neglected diseases have delivered critical opportunities to improve their avoidance, recognition, and treatment. Insights of knowledge about the genomic and sub-atomic level mechanisms of illness have empowered fundamental researchers to recognize new remedial targets and grow new agents that are changing the worldview of neglected tropical diseases research from non-specific, comprehensively harmful chemotherapies to profoundly focused combination treatments. Be that as it may, the capacity to translate biomedical discoveries into progress in care for patients with these NTDs stays reliant on the clinical trials framework. Clinical trials give a fundamental connection between logical revelation and clinical practice. These preliminaries are critical to the interpretation of new information into tangible advantages for patients and societies, and the information picked up in clinical trials can likewise advise and control further future investigation into the science of the diseases.

Numerous clinical trials are undertaken by pharmaceutical and biotechnology ventures, whose essential goals are to create novel therapeutic agents and gain approvals from various regulatory bodies such as Food and Drug Administration (FDA) for endorsement for clinical use. These innovative work endeavors involve huge costs and are basic to advance in not only NTDs treatment but other diseases treatments. Clinical trials publicly funded, additionally assume an indispensable job and are complementary to industry trials in propelling science and patient care, especially by tending to questions that are imperative to patients however are less inclined to be top needs of the industry. For instance, organizations may have less impetus to

- conduct clinical preliminaries to look at the viability of various treatment choices that are as of now endorsed for clinical use,
- combine novel treatments created by various supporters,
- develop treatments for uncommon illnesses,
- determine ideal term and portion of treatment with drugs in clinical use,
- test multimodality treatments, for example, radiation treatment, medical procedure, or gadgets in mix with drugs,
- study screening and counteraction procedures, or

- focus on recovery and personal satisfaction following treatment.

Production of negative research discoveries about the treatments utilized practically, which are underreported in the writing yet which are fundamental in setting the standard of care, is additionally a significant part of openly subsidized research.

The process of obtaining a drug lead in biomedical research is tedious and takes a long process. Most often researchers are demoralized by the endless possibilities they have to search through. Fortunately, pharmaceuticals research lately has permeated the use of computational tools to offer much-required assistance. Successful incorporation of molecular modeling techniques such as homology and de novo modeling from only sequence data to identify the structure and function of genes within an organism has helped speed up the process. A deeper understanding of chemical and biological systems coupled up with both computational and experimental strategies have played a pivotal role in offering rationalized drug discovery processes. As shown from the study performed in silico molecular modeling, molecular docking and binding experiments which explore the ligand conformation adopted within the binding sites of macromolecular targets. These processes have swiftly gained a valuable position in modern techniques of finding intervenes for growing ailments both global pandemics, together with majorly neglected diseases.in our study, we have successfully demonstrated our ability to obtain macromolecular structures from sequence data only, with a proven concept from classical techniques of structure-based design and X-ray crystallography, coupled with drug docking and binding in the process of drug discovery. We have shown the successes of using our methodology to not only save up on time and cost but also the ability to offer patients a better therapeutic strategy that is predictable. In addition to this, we still could use some of the used drugs as possible intervenes in other diseases by repurposing their previous intended use.

To date, a variety of algorithms are available to help in this process but more success in the future can be expected. The surge of computational power by quantum computing to solve space-time complexity which could help in improving biological simulation to generate more accurate and precise results is unfathomable. With better and relevant results we presume a future where we could have tailored intervention due to the advancement in technologies in sequencing, genome analytics toxicity and reaction simulations integrated with quick experimental strategies. A final word from our work is our capacity to offer any intervention soon lies within our imagination

CHAPTER EIGHT

8   References

Akiyoshi, B. and K. Gull (2014). "Discovery of unconventional kinetochores in kinetoplastids." Cell **156**(6): 1247-1258.

Alvar, J., S. Yactayo, et al. (2006). "Leishmaniasis and poverty." Trends in parasitology **22**(12): 552-557.

Avila, H., J. B. Pereira, et al. (1993). "Detection of Trypanosoma cruzi in blood specimens of chronic chagasic patients by polymerase chain reaction amplification of kinetoplast minicircle DNA: comparison with serology and xenodiagnosis." Journal of Clinical Microbiology **31**(9): 2421-2426.

Babokhov, P., A. O. Sanyaolu, et al. (2013). "A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis." Pathogens and global health **107**(5): 242-252.

Bao, G., W. J. Rhee, et al. (2009). "Fluorescent probes for live-cell RNA detection." Annual review of biomedical engineering **11**: 25.

Bern, C. (2015). "Chagas' disease." New England Journal of Medicine **373**(5): 456-466.

Blakeley, B. D., S. M. DePorter, et al. (2012). "Methods for identifying and characterizing interactions involving RNA." Tetrahedron **68**(43): 8837-8855.

Brown, R. A., V. Diemer, et al. (2013). "End-to-end conformational communication through a synthetic purinergic receptor by ligand-induced helicity switching." Nature Chemistry **5**(10): 853-860.

Brun, R. and J. Blum (2012). "Human african trypanosomiasis." Infectious disease clinics of North America **26**(2): 261-273.

Brun, R., J. Blum, et al. (2010). "Human african trypanosomiasis." The Lancet **375**(9709): 148-159.

Buguet, A., R. Cespuglio, et al. (2015). African Sleeping Sickness. Sleep Medicine, Springer**:** 159-165.

Burri, C. and R. Brun (2003). "Human African trypanosomiasis." Manson's tropical diseases **21**: 1303-1323.

Cavalli, A. and M. L. Bolognesi (2009). "Neglected tropical diseases: multi-target-directed ligands in the search for novel lead candidates against Trypanosoma and Leishmania." Journal of medicinal chemistry **52**(23): 7339-7359.

Chang, P. and R. S. Bray (1986). Leishmaniasis, Elsevier Science Publishers BV (Biomedical Division).

Chappuis, F., S. Sundar, et al. (2007). "Visceral leishmaniasis: what are the needs for diagnosis, treatment and control?" Nature reviews microbiology **5**(11): 873-882.

Control, C. f. D. and Prevention (2006). "Chagas disease after organ transplantation--Los Angeles, California, 2006." MMWR. Morbidity and mortality weekly report **55**(29): 798.

Coura, J. R. (2013). "Chagas disease: control, elimination and eradication. Is it possible?" Memorias do Instituto Oswaldo Cruz **108**(8): 962-967.

Cruz, J. A., M.-F. Blanchet, et al. (2012). "RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction." RNA **18**(4): 610-625.

Dawit, G., Z. Girma, et al. (2013). "A review on biology, epidemiology and public health significance of leishmaniasis." Journal of Bacteriology & Parasitology **2013**.

De Gaudenzi, J., A. C. Frasch, et al. (2005). "RNA-binding domain proteins in Kinetoplastids: a comparative analysis." Eukaryotic cell **4**(12): 2106-2114.

de Silanes, I. L., M. Zhan, et al. (2004). "Identification of a target RNA motif for RNA-binding protein HuR." Proceedings of the National Academy of Sciences of the United States of America **101**(9): 2987-2992.

Desjeux, P. (2004). "Leishmaniasis: current situation and new perspectives." Comparative immunology, microbiology and infectious diseases **27**(5): 305-318.

Desjeux, P., R. S. Ghosh, et al. (2013). "Report of the post kala-azar dermal leishmaniasis (PKDL) consortium meeting, New Delhi, India, 27–29 June 2012." Parasites & vectors **6**(1): 1.

Dorywalska, M., S. C. Blanchard, et al. (2005). "Site-specific labeling of the ribosome for single-molecule spectroscopy." Nucleic acids research **33**(1): 182-189.

Dumas, M. and F. Boa (1988). "Human african trypanosomiasis." Handbook of clinical neurology **8**: 339-344.

Dumas, M., B. Bouteille, et al. (2013). Progress in human African trypanosomiasis, sleeping sickness, Springer Science & Business Media.

Elsheikha, H. M. (2014). "The future of parasitology: challenges and opportunities." Frontiers in Veterinary Science **1**.

Erlanson, D. A., R. S. McDowell, et al. (2004). "Fragment-based drug discovery." Journal of medicinal chemistry **47**(14): 3463-3482.

Gehrke, S. S. (2012). Small Molecules with Anti-trypanosomal and Anti-leishmanial Activity, University of East Anglia.

Goto, H. and J. A. L. Lindoso (2012). "Cutaneous and mucocutaneous leishmaniasis." Infectious disease clinics of North America **26**(2): 293-307.

Guiguemde, W. A., A. A. Shelat, et al. (2010). "Chemical genetics of Plasmodium falciparum." Nature **465**(7296): 311-315.

Gull, K. (2001). "The biology of kinetoplastid parasites: insights and challenges from genomics and post-genomics." International journal for parasitology **31**(5): 443-452.

Haddad, S. J., R. Herbert, et al. (2013). "New world cutaneous leishmaniasis." BMJ **346**: f1431.

Hailu, A., D. A. Dagne, et al. (2016). Leishmaniasis. Neglected Tropical Diseases-Sub-Saharan Africa, Springer**:** 87-112.

Irwin, J. J. and B. K. Shoichet (2016). "Docking screens for novel ligands conferring new biology."

Ivens, A. C., C. S. Peacock, et al. (2005). "The genome of the kinetoplastid parasite, Leishmania major." Science **309**(5733): 436-442.

Jordan, A. M. (1986). Trypanosomiasis control and African rural development, Longman London; New York.

Kariuki, M., R. Ogwang, et al. (1998). "Signal transduction in malaria gametocytes." South African journal of science **94**(6): 271-272.

Kariuki, M. M., J. Kiaira, et al. (1998). "Plasmodium falciparum: purification of the various gametocyte developmental stages from in vitro-cultivated parasites." The American journal of tropical medicine and hygiene **59**(4): 505-508.

Kennedy, P. G. (2008). "The continuing problem of human African trypanosomiasis (sleeping sickness)." Annals of neurology **64**(2): 116-126.

Kennedy, P. G. (2013). "Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness)." The Lancet Neurology **12**(2): 186-194.

Kramer, P. M., C. M. Weber, et al. (2007). Optical Immunosensor and ELISA for the Analysis of Pyrethroids and DDT in Environmental Samples. ACS symposium series, Oxford University Press.

Krull, I. S., C. Gendreau, et al. (2000). Capillary electrophoresis in combinatorial library analysis, Marcel Dekker, New York.

Lake, J. A., V. F. De La Cruz, et al. (1988). "Evolution of parasitism: kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences." Proceedings of the National Academy of Sciences **85**(13): 4779-4783.

Lent, H. and P. Wygodzinsky (1979). "Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease." Bulletin of the American Museum of Natural History **163**(3): 123-520.

Lutumba, P., E. Matovu, et al. (2016). Human African Trypanosomiasis (HAT). Neglected Tropical Diseases-Sub-Saharan Africa, Springer**:** 63-85.

Mathenge, P., H. N. u. Mwangi, et al. (2015). "In silico structure of the 40s ribosomal subunit from plasmodium falciparum as drug targetby homology and de novo modeling."

Mathenge, P. G. (2013). Homology model of 30S ribosomal subunit from Mycobacteria tuberculosis, University of Nairobi.

Matthews, K. R. (2005). "The developmental cell biology of Trypanosoma brucei." Journal of cell science **118**(2): 283-290.

McDonald, A. and N. rH Stone (2016). "Human African Trypanosomiasis (Sleeping Sickness)." The Travel and Tropical Medicine Manual: 382.

Mulaa, F. J. and P. M. Krämer (2011). "Biosensors." Handbook of Food Safety Engineering: 313-351.

Murillo, J., L. M. Bofill, et al. (2016). "Congenital Chagas' disease transmission in the United States: Diagnosis in adulthood." IDCases **5**: 72-75.

Murray, H. W., J. D. Berman, et al. (2005). "Advances in leishmaniasis." The Lancet **366**(9496): 1561-1577.

Murray, M., P. Murray, et al. (1977). "An improved parasitological technique for the diagnosis of African trypanosomiasis." Transactions of the Royal Society of Tropical Medicine and Hygiene **71**(4): 325-326.

Mwangi, H. N. u. (2013). Structure of the 40S ribosomal subunit from Plasmodium falciparum By Homology and De novo modeling, University of Nairobi.

Mwangi, H. N. U., Wagacha, P., Mathenge, P., Sijenyi, F., & Mulaa, F. (2017). Structure of the 40S ribosomal subunit of Plasmodium falciparum by homology and de novo modeling. *Acta Pharmaceutica Sinica. B*, *7*(1), 97.

Nagle, A. S., S. Khare, et al. (2014). "Recent developments in drug discovery for leishmaniasis and human African trypanosomiasis." Chemical reviews **114**(22): 11305-11347.

Oberdörster, G., A. Maynard, et al. (2005). "Principles for characterizing the potential human health effects from exposure to nanomaterials: elements of a screening strategy." Particle and fibre toxicology **2**(1): 1.

Organization, W. H. (2013). "Control and surveillance of human African trypanosomiasis." World Health Organization Technical Report Series(984): 1.

Organization, W. H. (2014). "Chagas disease (American trypanosomiasis)." World Health Organ Fact Sheet **340**.

Organization, W. H. (2015). "Human African trypanosomiasis: update of the methodological framework for clinical trials: report of the first meeting of the Development of New Tools subgroup, Geneva, 24 September 2014."

Pavli, A. and H. C. Maltezou (2010). "Leishmaniasis, an emerging infection in travelers." International Journal of Infectious Diseases **14**(12): e1032-e1039.

Pepin, J. (1993). "The Treatment of Human African." Advances in parasitology **33**: 1.

Perryman, A. L., W. Yu, et al. (2015). "A virtual screen discovers novel, fragment-sized inhibitors of Mycobacterium tuberculosis InhA." Journal of chemical information and modeling **55**(3): 645-659.

Pirard, M., N. Iihoshi, et al. (2005). "The validity of serologic tests for Trypanosoma cruzi and the effectiveness of transfusional screening strategies in a hyperendemic region." Transfusion **45**(4): 554-561.

Rassi, A. and J. M. de Rezende (2012). "American trypanosomiasis (Chagas disease)." Infectious disease clinics of North America **26**(2): 275-291.

Rassi, A. and J. A. Marin-Neto (2010). "Chagas disease." The Lancet **375**(9723): 1388-1402.

Ready, P. D. (2014). "Epidemiology of visceral leishmaniasis." Clin Epidemiol **6**: 147-154.

Ruiz-Carmona, S., D. Alvarez-Garcia, et al. (2014). "rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids." PLoS Comput Biol **10**(4): e1003571.

Schrödinger, L. (2013). "Small-Molecule Drug Discovery Suite 2013–3: Glide, version 6.1." New York: Schrödinger, LLC.

Scott, W. R., P. H. Hünenberger, et al. (1999). "The GROMOS biomolecular simulation program package." The Journal of Physical Chemistry A **103**(19): 3596-3607.

Shikanai-Yasuda, M. A. and N. B. Carvalho (2012). "Oral transmission of Chagas disease." Clinical Infectious Diseases: cir956.

Siegfried, N. A., S. Busan, et al. (2014). "RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP)." Nature methods **11**(9): 959-965.

Sijenyi, F., P. Saro, et al. (2012). The RNA folding problems: different levels of sRNA structure prediction. RNA 3D structure analysis and prediction, Springer**:** 91-117.

Simmons, D. M., J. L. Arriza, et al. (2013). "A complete protocol for in situ hybridization of messenger RNAs in brain and other tissues with radio-labeled single-stranded RNA probes." Journal of Histotechnology.

Simpson, A. G., J. Lukeš, et al. (2002). "The evolutionary history of kinetoplastids and their kinetoplasts." Molecular biology and evolution **19**(12): 2071-2083.

Steinmann, P., C. M. Stone, et al. (2015). "Contemporary and emerging strategies for eliminating human African trypanosomiasis due to Trypanosoma brucei gambiense: review." Tropical Medicine & International Health **20**(6): 707-718.

Stich, A. (2013). "Human african trypanosomiasis." Principles of Medicine in Africa: 415.

Stuart, K., R. Brun, et al. (2008). "Kinetoplastids: related protozoan pathogens, different diseases." The Journal of clinical investigation **118**(4): 1301-1310.

Vickerman, K. and G. H. Coombs (1999). "Protozoan paradigms for cell biology." J Cell Sci **112**(17): 2797-2798.

Vickerman, K. and T. Preston (1976). "Comparative cell biology of the kinetoplastid flagellates." Biology of the Kinetoplastida.

Warrier, T., M. Martinez-Hoyos, et al. (2015). "Identification of Novel Anti-mycobacterial Compounds by Screening a Pharmaceutical Small-Molecule Library against Nonreplicating Mycobacterium tuberculosis." ACS Infectious Diseases **1**(12): 580-585.

Weedall, G. D. and N. Hall (2015). "Sexual reproduction and genetic exchange in parasitic protists." Parasitology **142**(S1): S120-S127.

Woo, P. T. (2001). "Cryptobiosis and its control in North American fishes." International journal for parasitology **31**(5): 566-574.

Young, C., P. Losikoff, et al. (2007). "Transfusion-acquired Trypanosoma cruzi infection." Transfusion **47**(3): 540-544.

Altschul, S. F., et al. (1990). "Basic local alignment search tool." Journal of molecular biology **215**(3): 403-410.

Berman, H., et al. (2006). "The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data." Nucleic acids research **35**(suppl_1): D301-D303.

Cannone, J. J., et al. (2002). "The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs." BMC bioinformatics **3**(1): 2.

Coordinators, N. R. (2016). "Database resources of the national center for biotechnology information." Nucleic acids research **44**(Database issue): D7.

Corpet, F. (1988). "Multiple sequence alignment with hierarchical clustering." Nucleic acids research **16**(22): 10881-10890.

Das, R. and D. Baker (2007). "Automated de novo prediction of native-like RNA tertiary structures." Proceedings of the National Academy of Sciences **104**(37): 14664-14669.

Flohil, J., et al. (2002). "Completion and refinement of 3-D homology models with restricted molecular dynamics: application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis." Proteins: Structure, Function, and Bioinformatics **48**(4): 593-604.

Harvey, M. J., et al. (2009). "ACEMD: accelerating biomolecular dynamics in the microsecond time scale." Journal of chemical theory and computation **5**(6): 1632-1639.

Leontis, N. B. and E. Westhof (2001). "Geometric nomenclature and classification of RNA base pairs." Rna **7**(4): 499-512.

Magnus, M., et al. (2016). "SimRNAweb: a web server for RNA 3D structure modeling with optional restraints." Nucleic acids research **44**(W1): W315-W319.

Mathews, D. H. and D. H. Turner (2006). "Prediction of RNA secondary structure by free energy minimization." Current opinion in structural biology **16**(3): 270-278.

Mwangi, H. N. u., et al. (2017). "Structure of the 40S ribosomal subunit of Plasmodium falciparum by homology and de novo modeling." Acta pharmaceutica sinica B **7**(1): 97-105.

Nelson, M. T., et al. (1996). "NAMD: a parallel, object-oriented molecular dynamics program." The International Journal of Supercomputer Applications and High Performance Computing **10**(4): 251-268.

Rother, M., et al. (2011). "ModeRNA: a tool for comparative modeling of RNA 3D structure." Nucleic acids research **39**(10): 4007-4022.

Sijenyi, F., et al. (2012). The RNA folding problems: different levels of srna structure prediction. RNA 3D structure analysis and prediction, Springer**:** 91-117.

Studio, D. (2013). "Accelrys Inc." San Diego, CA, USA.

Watson, J. D. and F. H. Crick (1953). "Molecular structure of nucleic acids." Nature **171**(4356): 737-738.

\AAqvist, J., & Kamerlin, S. C. (2015). Exceptionally large entropy contributions enable the high rates of GTP hydrolysis on the ribosome. *Scientific Reports*, *5*, 15817.

ABDI, N. M., & FREDRICK, K. (2005). Contribution of 16S rRNA nucleotides forming the 30S subunit A and P sites to translation in Escherichia coli. *Rna*, *11*(11), 1624–1632.

Aksoy, S. (2003). Control of tsetse flies and trypanosomes using molecular genetics. *Veterinary Parasitology*, *115*(2), 125–145.

Alcântara, L. M., Ferreira, T. S., Gadelha, F. R., & Miguel, D. C. (2018). Challenges in drug discovery targeting TriTryp diseases with an emphasis on leishmaniasis. *International Journal for Parasitology: Drugs and Drug Resistance*.

Bates, P. A. (2018). Revising Leishmania's life cycle. *Nature Microbiology*, *3*(5), 529–530.

Ben-Shem, A., de Loubresse, N. G., Melnikov, S., Jenner, L., Yusupova, G., & Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 \AA resolution. *Science*, *334*(6062), 1524–1529.

Ben-Shem, A., Jenner, L., Yusupova, G., & Yusupov, M. (2010). Crystal structure of the eukaryotic ribosome. *Science*, *330*(6008), 1203–1209.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, *28*(1), 235–242.

Bern, C. (2015). Chagas' disease. *New England Journal of Medicine*, *373*(5), 456–466.

Brown, A., Rathore, S., Kimanius, D., Aibara, S., Bai, X., Rorbach, J., … Ramakrishnan, V. (2017). Structures of the human mitochondrial ribosome in native states of assembly. *Nature Structural and Molecular Biology*, *24*(10), 866.

Brun, R., Blum, J., Chappuis, F., & Burri, C. (2010). Human african trypanosomiasis. *The Lancet*, *375*(9709), 148–159.

Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., … Müller, K. M. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, *3*(1), 2.

Cannone, J. J., Sweeney, B. A., Petrov, A. I., Gutell, R. R., Zirbel, C. L., & Leontis, N. (2015). R3D-2-MSA: the RNA 3D structure-to-multiple sequence alignment server. *Nucleic Acids Research*, *43*(W1), W15–W23.

Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., & Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, *407*(6802), 340.

Castillo, E., A Dea-Ayuela, M., Bolás-Fernández, F., Rangel, M., & E Gonzalez-Rosende, M. (2010). The kinetoplastid chemotherapy revisited: current drugs, recent advances and future perspectives. *Current Medicinal Chemistry*, *17*(33), 4027–4051.

Charette, M., & Gray, M. W. (2000). Pseudouridine in RNA: what, where, how, and why. *IUBMB Life*, *49*(5), 341–351.

Chastain, M., & Tinoco Jr, I. (1991). Structural elements in RNA. In *Progress in nucleic acid research and molecular biology* (Vol. 41, pp. 131–177). Elsevier.

Chen, B. (2015). *Cryo-EM and Time-Resolved Cryo-EM Studies on Translation*. Columbia University.

Chen, B., Kaledhonkar, S., Sun, M., Shen, B., Lu, Z., Barnard, D., … Frank, J. (2015). Structural dynamics of ribosome subunit association studied by mixing-spraying time-resolved cryogenic electron microscopy. *Structure*, *23*(6), 1097–1105.

Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H., & Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, *135*, 24–35.

Clos II, L. J. (2007). *NMR solution studies of 16S rRNA A-site functional mutants and binding of novel ligands*. Wayne State University.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, *16*(22), 10881–10890.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561.

Croft, S. L. (2008). *Kinetoplastida: new therapeutic strategies*. EDP Sciences.

Croft, Simon L., Barrett, M. P., & Urbina, J. A. (2005). Chemotherapy of trypanosomiases and leishmaniasis. *Trends in Parasitology*, *21*(11), 508–512.

Cruz, I., Nieto, J., Moreno, J., Cañavate, C., Desjeux, P., & Alvar, J. (2006). *Leishmania/HIV co-infections in the second decade*.

Dale, T., & Uhlenbeck, O. C. (2005). Amino acid specificity in translation. *Trends in Biochemical Sciences*, *30*(12), 659–665.

Darty, K., Denise, A., & Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, *25*(15), 1974.

Davis, J. H., Tonelli, M., Scott, L. G., Jaeger, L., Williamson, J. R., & Butcher, S. E. (2005). RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop–receptor complex. *Journal of Molecular Biology*, *351*(2), 371–382.

de Loubresse, N. G., Prokhorova, I., Holtkamp, W., Rodnina, M. V., Yusupova, G., & Yusupov, M. (2014). Structural basis for the inhibition of the eukaryotic ribosome. *Nature*, *513*(7519), 517.

Decatur, W. A., & Fournier, M. J. (2002). rRNA modifications and ribosome function. *Trends in Biochemical Sciences*, *27*(7), 344–351.

Dehhaghi, M., Mohammadipanah, F., & Guillemin, G. J. (2018). Myxobacterial natural products: An under-valued source of products for drug discovery for neurological disorders. *Neurotoxicology*, *66*, 195–203.

Denny, S. K., & Greenleaf, W. J. (2018). Linking RNA Sequence, Structure, and Function on Massively Parallel High-Throughput Sequencers. *Cold Spring Harbor Perspectives in Biology*, a032300.

Dever, T. E. (2002). Gene-specific regulation by general translation factors. *Cell*, *108*(4), 545–556.

Dinger, M. E., Amaral, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., … Simons, C. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research*, gr–078378.

Doerfel, L. K., Wohlgemuth, I., Kubyshkin, V., Starosta, A. L., Wilson, D. N., Budisa, N., & Rodnina, M. V. (2015). Entropic contribution of elongation factor P to proline positioning at the catalytic center of the ribosome. *Journal of the American Chemical Society*, *137*(40), 12997–13006.

Dupuis-Sandoval, F., Poirier, M., & Scott, M. S. (2015). The emerging landscape of small nucleolar RNAs in cell biology. *Wiley Interdisciplinary Reviews: RNA*, *6*(4), 381–397.

Eddy, S. R. (2001). Non–coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, *2*(12), 919.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Fahlman, R. P., Dale, T., & Uhlenbeck, O. C. (2004). Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Molecular Cell*, *16*(5), 799–805.

Franco, J. R., Simarro, P. P., Diarra, A., & Jannin, J. G. (2014). Epidemiology of human African trypanosomiasis. *Clinical Epidemiology*, *6*, 257.

Frank, J., & Agrawal, R. K. (2000). A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature*, *406*(6793), 318.

Fredrick, K., & Noller, H. F. (2002). Accurate translocation of mRNA by the ribosome requires a peptidyl group or its analog on the tRNA moving into the 30S P site. *Molecular Cell*, *9*(5), 1125–1131.

Garcia, R. O., Krug, D., & Müller, R. (2009). Discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods. *Methods in Enzymology*, *458*, 59–91.

Geigenmüller, U., & Nierhaus, K. H. (1990). Significance of the third tRNA binding site, the E site, on E. coli ribosomes for the accuracy of translation: an occupied E site prevents the binding of non-cognate aminoacyl-tRNA to the A site. *The EMBO Journal*, *9*(13), 4527–4533.

Hammarton, T. C., Mottram, J., & Doerig, C. (2003). The cell cycle of parasitic protozoa: potential for chemotherapeutic exploitation. *PROGRESS IN CELL CYCLE RESEARCH.*, *5*, 91–102.

Hashem, Y., Des Georges, A., Fu, J., Buss, S. N., Jossinet, F., Jobe, A., … Bajaj, C. (2013). High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. *Nature*, *494*(7437), 385.

Hausser, J., & Zavolan, M. (2014). Identification and consequences of miRNA–target interactions—beyond repression of gene expression. *Nature Reviews Genetics*, *15*(9), 599.

He, L., & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, *5*(7), 522.

Hendrix, D. K., Brenner, S. E., & Holbrook, S. R. (2005). RNA structural motifs: building blocks of a modular biomolecule. *Quarterly Reviews of Biophysics*, *38*(3), 221–243.

Herman, J. D., Gallalee, J. V., & Best, J. M. (1987). Sodium stibogluconate (Pentostam) inhibition of glucose catabolism via the glycolytic pathway, and fatty acid β-oxidation in Leishmania mexicana amastigotes. *Biochemical Pharmacology*, *36*(2), 197–201.

Hernandez, N. (2001). Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *Journal of Biological Chemistry*, *276*(29), 26733–26736.

Hinnebusch, A. G. (2014). The scanning mechanism of eukaryotic translation initiation. *Annual Review of Biochemistry*, *83*, 779–812.

Hinnebusch, A. G., & Lorsch, J. R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harbor Perspectives in Biology*, a011544.

Hoffmann, T., Krug, D., Bozkurt, N., Duddela, S., Jansen, R., Garcia, R., … Müller, R. (2018). Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nature Communications*, *9*(1), 803.

Huang, L., Ashraf, S., & Lilley, D. M. (2018). The role of RNA structure in translational regulation by L7Ae protein in archaea. *RNA*, rna–068510.

Huang, L., & Lilley, D. M. (2016). The kink turn, a key architectural element in RNA structure. *Journal of Molecular Biology*, *428*(5), 790–801.

Iwasaki, Y. W., Siomi, M. C., & Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions. *Annual Review of Biochemistry*, *84*(1), 405–433. https://doi.org/10.1146/annurev-biochem-060614-034258

Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, *11*(2), 113.

Jády, B. E., & Kiss, T. (2001). A small nucleolar guide RNA functions both in 2′-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *The EMBO Journal*, *20*(3), 541–551.

Jobe, A., Liu, Z., Gutierrez-Vargas, C., & Frank, J. (2018). New Insights into Ribosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, a032615.

Katunin, V. I., Muth, G. W., Strobel, S. A., Wintermeyer, W., & Rodnina, M. V. (2002). Important contribution to catalysis of peptide bond formation by a single ionizing group within the ribosome. *Molecular Cell*, *10*(2), 339–346.

Kaur, G., & Rajput, B. (2014). Comparative analysis of the omics technologies used to study antimonial, amphotericin B, and pentamidine resistance in leishmania. *Journal of Parasitology Research*, *2014*.

Kazemi, M., Himo, F., & \AAqvist, J. (2016). Peptide release on the ribosome involves substrate-assisted base catalysis. *Acs Catalysis*, *6*(12), 8432–8439.

Kazemi, M., Sočan, J., Himo, F., & \AAqvist, J. (2018). Mechanistic alternatives for peptide bond formation on the ribosome. *Nucleic Acids Research*, *46*(11), 5345–5354.

Khatter, H., Myasnikov, A. G., Natchiar, S. K., & Klaholz, B. P. (2015). Structure of the human 80S ribosome. *Nature*, *520*(7549), 640.

Kim, V. N. (2006). Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes & Development*, *20*(15), 1993–1997.

Klein, D. J., Schmeing, T. M., Moore, P. B., & Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. *The EMBO Journal*, *20*(15), 4214–4221.

Koch, M., Clementi, N., Rusca, N., Vögele, P., Erlacher, M., & Polacek, N. (2015). The integrity of the G2421-C2395 base pair in the ribosomal E-site is crucial for protein synthesis. *RNA Biology*, *12*(1), 70–81.

Lee, K., Varma, S., SantaLucia Jr, J., & Cunningham, P. R. (1997). In vivo determination of RNA Structure-Function relationships: analysis of the 790 loop in ribosomal RNA1. *Journal of Molecular Biology*, *269*(5), 732–743.

Legros, D., Ollivier, G., Gastellu-Etchegorry, M., Paquet, C., Burri, C., Jannin, J., & Büscher, P. (2002). Treatment of human African trypanosomiasis—present situation and needs for research and development. *The Lancet Infectious Diseases*, *2*(7), 437–440.

Leontis, N. B., Stombaugh, J., & Westhof, E. (2002). The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Research*, *30*(16), 3497–3531.

Maarouf, M., Lawrence, F., Brown, S., & Robert-Gero, M. (1997). Biochemical alterations in paromomycin-treated Leishmania donovani promastigotes. *Parasitology Research*, *83*(2), 198–202. https://doi.org/10.1007/s004360050232

Paromomycin is used for the treatment of leishmaniasis in humans, but little is known about its mechanism of action. Investigating the effect of this antibiotic on promastigotes of Leishmania donovani, we showed that inhibition of the multiplication of these parasites could be related to its effect on RNA synthesis and to modifications of membranous polar lipids and membrane fluidity, leading to altered membrane permeability.

Martinez, H. M., Maizel Jr, J. V., & Shapiro, B. A. (2008). RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. *Journal of Biomolecular Structure and Dynamics*, *25*(6), 669–683.

Matamoros, M. C. (2016). *Identification of Leishmania spp. and Trypanosoma cruzi in sylvatic animals in El Paso County, Texas*. The University of Texas at El Paso.

Matsumura, S., Ikawa, Y., & Inoue, T. (2003). Biochemical characterization of the kink-turn RNA motif. *Nucleic Acids Research*, *31*(19), 5544–5551.

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, *15*(suppl_1), R17–R29.

Mbongo, N., Loiseau, P. M., Billion, M. A., & Robert-Gero, M. (1998). Mechanism of amphotericin B resistance inLeishmania donovani promastigotes. *Antimicrobial Agents and Chemotherapy*, *42*(2), 352–357.

McAdams, N. M., Ammerman, M. L., Nanduri, J., Lott, K., Fisk, J. C., & Read, L. K. (2015). An arginine-glycine-rich RNA binding protein impacts the abundance of specific mRNAs in the mitochondria of Trypanosoma brucei. *Eukaryotic Cell*, *14*(2), 149–157.

Melnikov, S., Ben-Shem, A., De Loubresse, N. G., Jenner, L., Yusupova, G., & Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nature Structural & Molecular Biology*, *19*(6), 560.

Mwangi, H. N., Gitonga, P. M., Wagacha, P. W., Sijenyi, F., & Mulaa, F. (2018). Integrating mechanism-based screening paradigm into homology and de novo modeling exemplified by Mycobacterium Tuberculosis 30S ribosomal structure and its potential application as a screening target. *International Journal of Scientific & Engineering Research*, *9*(2), 305–317.

Mwangi, H. N., Wagacha, P., Mathenge, P., Sijenyi, F., & Mulaa, F. (2017a). Structure of the 40S ribosomal subunit of Plasmodium falciparum by homology and de novo modeling. *Acta Pharmaceutica Sinica B*, *7*(1), 97–105. https://doi.org/10.1016/j.apsb.2016.10.003

Mwangi, H. N., Wagacha, P., Mathenge, P., Sijenyi, F., & Mulaa, F. (2017b). Structure of the 40S ribosomal subunit of Plasmodium falciparum by homology and de novo modeling. *Acta Pharmaceutica Sinica B*, *7*(1), 97–105.

Nagaswamy, U., Larios-Sanz, M., Hury, J., Collins, S., Zhang, Z., Zhao, Q., & Fox, G. E. (2002). NCIR: a database of non-canonical interactions in known RNA structures. *Nucleic Acids Research*, *30*(1), 395–397.

Nett, M., & König, G. M. (2007). The chemistry of gliding bacteria. *Natural Product Reports*, *24*(6), 1245–1261.

Nierhaus, K. H. (2004). tRNA Locations on the Ribosome. *Protein Synthesis and Ribosome Structure*, 207–217.

Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science*, *289*(5481), 920–930.

Noireau, F., Diosque, P., & Jansen, A. M. (2009). Trypanosoma cruzi: adaptation to its vectors and its hosts. *Veterinary Research*, *40*(2), 1–23.

Nok, A. J. (2003). Arsenicals (melarsoprol), pentamidine and suramin in the treatment of human African trypanosomiasis. *Parasitology Research*, *90*(1), 71–79.

Noller, H. F., Yusupov, M. M., Yusupova, G. Z., Baucom, A., & Cate, J. H. D. (2002). Translocation of tRNA during protein synthesis. *FEBS Letters*, *514*(1), 11–16.

Ofengand, J., & Bakin, A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts1. *Journal of Molecular Biology*, *266*(2), 246–268.

Olejniczak, M., Dale, T., Fahlman, R. P., & Uhlenbeck, O. C. (2005). Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding. *Nature Structural and Molecular Biology*, *12*(9), 788.

Panek, J., Kolář, M., Vohradský, J., & Shivaya Valášek, L. (2013). An evolutionary conserved pattern of 18S rRNA sequence complementarity to mRNA 5′ UTRs and its implications for eukaryotic gene translation regulation. *Nucleic Acids Research*, *41*(16), 7625–7634.

Paul, C. P., Good, P. D., Winer, I., & Engelke, D. R. (2002). Effective expression of small interfering RNA in human cells. *Nature Biotechnology*, *20*(5), 505.

Polikanov, Y. S., Steitz, T. A., & Innis, C. A. (2014). A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. *Nature Structural & Molecular Biology*, *21*(9), 787.

Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., … Adamiak, R. W. (2012). Automated 3D structure composition for large RNAs. *Nucleic Acids Research*, *40*(14), e112–e112.

Querido, J. B., Mancera-Martínez, E., Vicens, Q., Bochler, A., Chicher, J., Simonetti, A., & Hashem, Y. (2017). The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein. *Structure*, *25*(12), 1785–1794.

Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A., & Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science*, *331*(6018), 730–736.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell*, *108*(4), 557–572.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., & Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature*, *441*(7092), 469.

Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, *11*(1), 129.

Robert, F., & Brakier-Gingras, L. (2003). A functional interaction between ribosomal proteins S7 and S11 within the bacterial ribosome. *Journal of Biological Chemistry*.

Rodnina, M. V., Beringer, M., & Wintermeyer, W. (2007). How ribosomes make peptide bonds. *Trends in Biochemical Sciences*, *32*(1), 20–26.

Roeder, R. G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Letters*, *579*(4), 909–915.

Romero, E. L., & Morilla, M. J. (2008). Drug delivery systems against leishmaniasis? Still an open question. *Expert Opinion on Drug Delivery*, *5*(7), 805–823.

Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M., & Yusupova, G. (2015). Structural insights into the translational infidelity mechanism. *Nature Communications*, *6*, 7251.

Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M., & Yusupova, G. (2016). New structural insights into translational miscoding. *Trends in Biochemical Sciences*, *41*(9), 798–814.

Scaiola, A., Peña, C., Weisser, M., Böhringer, D., Leibundgut, M., Klingauf-Nerurkar, P., … Ban, N. (2018). Structure of a eukaryotic cytoplasmic pre-40S ribosomal subunit. *The EMBO Journal*, e98499.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., … Franceschi, F. (2000). Structure of functionally activated small ribosomal subunit at 3.3 \AA resolution. *Cell*, *102*(5), 615–623.

Schmeing, T. M., Huang, K. S., Strobel, S. A., & Steitz, T. A. (2005). An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature*, *438*(7067), 520.

Schmeing, T. M., & Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature*, *461*(7268), 1234.

Schmeing, T. M., Seila, A. C., Hansen, J. L., Freeborn, B., Soukup, J. K., Scaringe, S. A., … Steitz, T. A. (2002). A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nature Structural and Molecular Biology*, *9*(3), 225.

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., & Cate, J. H. D. (2005). Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science*, *310*(5749), 827–834. https://doi.org/10.1126/science.1117230

Selmer, M., Dunham, C. M., Murphy, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., … Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, *313*(5795), 1935–1942.

Shi, Xinying, Chiu, K., Ghosh, S., & Joseph, S. (2009). Bases in 16S rRNA important for subunit association, tRNA binding, and translocation. *Biochemistry*, *48*(29), 6772–6782.

Shi, Xuesong, Huang, L., Lilley, D. M., Harbury, P. B., & Herschlag, D. (2016). The solution structural ensembles of RNA kink-turn motifs and their protein complexes. *Nature Chemical Biology*, *12*(3), 146.

Sievers, A., Beringer, M., Rodnina, M. V., & Wolfenden, R. (2004). The ribosome as an entropy trap. *Proceedings of the National Academy of Sciences*, *101*(21), 7897–7901.

Sijenyi, F. (2008). *Structure and dynamics of Escherichia coli 16S rRNA helix 23 and human 18S rRNA helix 24 by NMR spectroscopy*. Wayne State University.

Sijenyi, F., Saro, P., Ouyang, Z., Damm-Ganamet, K., Wood, M., Jiang, J., & SantaLucia, J. (2012). The RNA folding problems: different levels of srna structure prediction. In *RNA 3D structure analysis and prediction* (pp. 91–117). Springer.

Spahn, C. M., Beckmann, R., Eswar, N., Penczek, P. A., Sali, A., Blobel, G., & Frank, J. (2001). Structure of the 80S ribosome from Saccharomyces cerevisiae—tRNA-ribosome and subunit-subunit interactions. *Cell*, *107*(3), 373–386.

Spahn, C. M., Gomez-Lorenzo, M. G., Grassucci, R. A., Jørgensen, R., Andersen, G. R., Beckmann, R., … Frank, J. (2004). Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *The EMBO Journal*, *23*(5), 1008–1019.

Stark, H., Mueller, F., Orlova, E. V., Schatz, M., Dube, P., Erdemir, T., … van Heel, M. (1995). The 70S Escherichia coli ribosome at 23 \AA resolution: fitting the ribosomal RNA. *Structure*, *3*(8), 815–821.

Steger, G., Riesner, D., Maurel, M.-C., & Perreault, J.-P. (2017). Viroid Structure. In *Viroids and Satellites* (pp. 63–70). Elsevier.

Steitz, T. A. (2008). A structural understanding of the dynamic ribosome machine. *Nature Reviews Molecular Cell Biology*, *9*(3), 242.

Steverding, D. (2017). Sleeping sickness and nagana disease caused by Trypanosoma brucei. In *Arthropod Borne Diseases* (pp. 277–297). Springer.

Stuart, K., Brun, R., Croft, S., Fairlamb, A., Gürtler, R. E., McKerrow, J., … Tarleton, R. (2008). Kinetoplastids: related protozoan pathogens, different diseases. *The Journal of Clinical Investigation*, *118*(4), 1301–1310.

Sudhandiran, G., & Shaha, C. (2003). Antimonial-induced increase in intracellular Ca2+ through non-selective cation channels in the host and the parasite is responsible for apoptosis of intracellular Leishmania donovani amastigotes. *Journal of Biological Chemistry*, *278*(27), 25120–25132.

Sundar, S., & Chakravarty, J. (2008). Paromomycin in the treatment of leishmaniasis. *Expert Opinion on Investigational Drugs*, *17*(5), 787–794.

Sundar, S., & Goyal, N. (2007). Molecular mechanisms of antimony resistance in Leishmania. *Journal of Medical Microbiology*, *56*(2), 143–153.

Teague, S. J. (2003). Implications of protein flexibility for drug discovery. *Nature Reviews Drug Discovery*, *2*(7), 527.

Tyler, K. M., & Engman, D. M. (2001). The life cycle of Trypanosoma cruzi revisited. *International Journal for Parasitology*, *31*(5–6), 472–481.

Valencia-Sanchez, M. A., Liu, J., Hannon, G. J., & Parker, R. (2006). Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Development*, *20*(5), 515–524.

Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M. ans, & Frank, J. (2003). Locking and unlocking of ribosomal motions. *Cell*, *114*(1), 123–134.

Wasan, K. M., Morton, R. E., Rosenblum, M. G., & Lopez-Berestein, G. (1994). Decreased toxicity of liposomal amphotericin B due to association of amphotericin B with high-density

lipoproteins: Role of lipid transfer protein. *Journal of Pharmaceutical Sciences*, *83*(7), 1006–1010.

Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids. *Nature*, *171*(4356), 737–738.

Weinger, J. S., Parnell, K. M., Dorner, S., Green, R., & Strobel, S. A. (2004). Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nature Structural and Molecular Biology*, *11*(11), 1101.

Weisser, M. (2017). *Structural Insights into Eukaryotic Translation Initiation and Re-Initiation* (PhD Thesis). ETH Zurich.

Weissman, K. J., & Müller, R. (2010). Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Natural Product Reports*, *27*(9), 1276–1295.

Weixlbaumer, A., Jin, H., Neubauer, C., Voorhees, R. M., Petry, S., Kelley, A. C., & Ramakrishnan, V. (2008). Insights into translational termination from the structure of RF2 bound to the ribosome. *Science*, *322*(5903), 953–956.

Weng, J., Aphasizheva, I., Etheridge, R. D., Huang, L., Wang, X., Falick, A. M., & Aphasizhev, R. (2008). Guide RNA-binding complex from mitochondria of trypanosomatids. *Molecular Cell*, *32*(2), 198–209.

Wenzel, S. C., & Mueller, R. (2009). The biosynthetic potential of myxobacteria and their impact in drug discovery. *Current Opinion in Drug Discovery & Development*, *12*(2), 220–230.

Whitworth, D. E. (2008). *Myxobacteria. Multicellularity and differentiation.* ASM press.

Wimberly, B. T., Brodersen, D. E., Clemons Jr, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., … Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature*, *407*(6802), 327.

Wittrup, A., & Lieberman, J. (2015). Knocking down disease: a progress report on siRNA therapeutics. *Nature Reviews Genetics*, *16*(9), 543.

Wlodawer, A., Minor, W., Dauter, Z., & Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *The FEBS Journal*, *275*(1), 1–21.

Wyllie, S., & Fairlamb, A. H. (2006). Differential toxicity of antimonial compounds and their effects on glutathione homeostasis in a human leukaemia monocyte cell line. *Biochemical Pharmacology*, *71*(3), 257–267.

Wyllie, S., Vickers, T. J., & Fairlamb, A. H. (2008). Roles of trypanothione S-transferase and tryparedoxin peroxidase in resistance to antimonials. *Antimicrobial Agents and Chemotherapy*, *52*(4), 1359–1365.

Xu, X., & Chen, S.-J. (2015). A method to predict the 3D structure of an RNA scaffold. In *RNA Scaffolds* (pp. 1–11). Springer.

Yoshizawa, S., Fourmy, D., & Puglisi, J. D. (1999). Recognition of the codon-anticodon helix by ribosomal RNA. *Science*, *285*(5434), 1722–1725.

Young, J. Y., Westbrook, J. D., Feng, Z., Sala, R., Peisach, E., Oldfield, T. J., … Berrisford, J. M. (2017). OneDep: unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive. *Structure*, *25*(3), 536–545.

Youngman, E. M., Brunelle, J. L., Kochaniak, A. B., & Green, R. (2004). The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell*, *117*(5), 589–599.

Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D., & Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 \AA resolution. *Science*, *292*(5518), 883–896.

121

Yusupova, G., & Yusupov, M. (2017). Crystal structure of eukaryotic ribosome and its complexes with inhibitors. *Phil. Trans. R. Soc. B*, *372*(1716), 20160184.

Zok, T., Antczak, M., Zurkowski, M., Popenda, M., Blazewicz, J., Adamiak, R. W., & Szachniuk, M. (2018). RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic Acids Research*, *46*(W1), W30–W35.

9   APPENDICES

9.1   Computer-Aided Drug Design

Molecular modeling is one of the most rapid developing scientific fields, as it comprises of a wide range of theoretical and computational tools used to model and simulate small chemical and biological systems with the purpose of understanding their behavior at an atomistic level.

While experimental techniques can significantly demonstrate the mechanism of action of a biological system, the extensive labor, time reservations and financial shortfalls have led research communities toward enhanced computational alternatives. The discipline of computational chemistry forms part of the nucleus of molecular modeling, allowing for significant medical breakthroughs due to immense improvements in computer hardware and software over recent decades. Starting in the 1960s and progressing rapidly since the late 1980s, these computational techniques have provided a robust platform for biomolecular structure analysis and drug discovery.

Rational drug design is based on the fundamental knowledge that the activity of a drug is obtained from the binding of the compound to a molecular pocket of the biological target. The drug's chemical and geometric stability at the molecular pocket is complementary to successful activity. The computational methods used in rational drug design and structure analysis includes protein modeling (homology modeling), virtual screening, molecular docking and predictive pharmacokinetic analysis.

By combining the above molecular modeling principles with Molecular Dynamic Simulations, the target's flexibility and inhibitor binding landscape, as well as a compound's structural mechanism of action may be analyzed.

A summarized version of the computational drug design process is depicted below. Due to the complexity of molecular dynamic simulations, we will only be delving into the first 4 stages of the molecular modeling techniques.
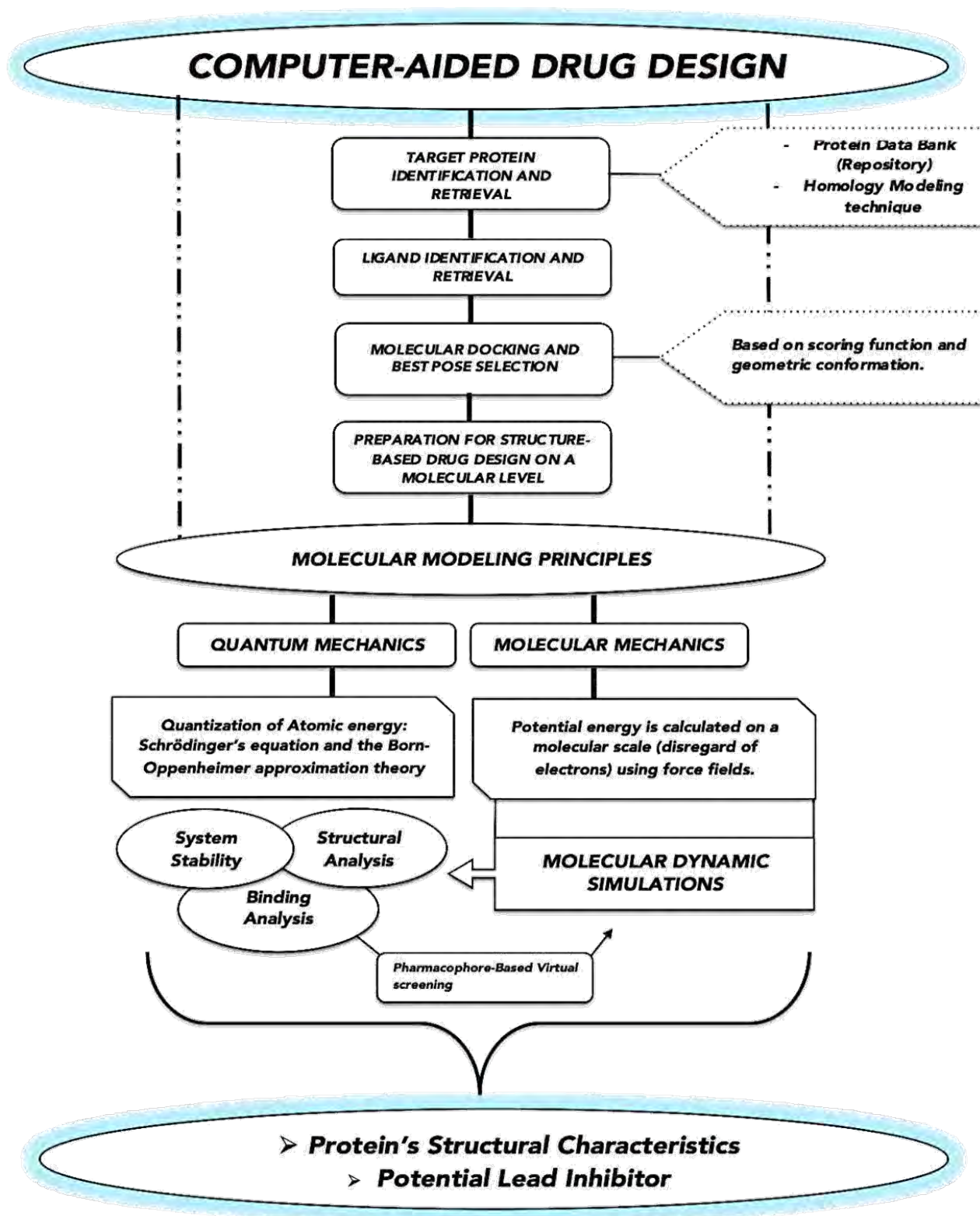
Figure 40: A wholistic image showing a complete process of Computer-Aided Drug Design

9.2   Examples Of Chemical Libraries

Virtual chemical libraries form the basis of the rational drug design process. There are numerous chemical libraries available from both commercial and non-commercial sources on the internet.

These libraries contain billions of small molecules, which may be FDA-approved drugs, chemical analogs, patented drugs, lead compounds and compounds synthesized with unknown activity.

A study by Walters (2018) further elaborates on the development, algorithms, demands and utilization of these libraries in the drug discovery domain (publication available in folder provided).

For the purpose of this workshop, we will look at a few of the libraries and elucidate on the potential applications within pharmaceutical science research.

Drugbank

URL: https://www.drugbank.ca/

The latest release of Drug Bank contains 13,443 drug entries including 2,621 approved small molecule drugs, 1,346 approved protein/peptide drugs, 130 nutraceuticals and over 6,335 experimental drugs. Additionally, 5,157 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each Drug Card entry contains more than 200 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

Drugbank offers key information, including:

-   A chemical/ physical description of the compound

-   Prescription products

-   Generic products (if drug is approved)

-   Pharmacology of the drug

-   Pharmacokinetic/dynamic properties

-   Pharmacoeconomics (if on the market)

- Mechanism of action (if available)

- Biological Targets (if available)

- Drug interactions (if available)

Drugbank also offers the option to search for potential analogs by adding in the structure of the known molecule.

2. **Pubchem**

URL: https://pubchem.ncbi.nlm.nih.gov/

PubChem is the world's largest collection of freely accessible chemical information. It is a free tool that was made available by the NIH and permits research data from any source. The database allows you to search chemicals by name, molecular formula, structure, and other identifiers. PubChem contains over 97 million compounds, 237 million chemical substances and 268 million bioactives. PubChem mostly contains small molecules, but also larger molecules such as nucleotides, carbohydrates, lipids, peptides, and chemically modified macromolecules.

PubChem offers key information, including:

- A chemical/physical description of the compound

- Spectral information

- Chemical Vendors

- Pharmacology/biochemistry of the drug

- Toxicity of the drug

- Use and manufacturing

- Biomolecular interactions

- Biological testing studies

PubChem also offers the option to search for potential analogs by adding in the structure of the known molecule.

ZINCdatabase

URL: https://zinc.docking.org/

ZINC was specifically designed for virtual screening. The database contains over 230 million purchasable compounds in ready-to-dock, 3D formats. ZINC also contains over 750 million purchasable compounds you can search for analogs in under a minute. Screening can be done using the molecular name, structure or SMILES. There is also an option to allow for the filtering of compounds for various criteria such as FDA-approved, experimentally approved, currently in trials etc.

Once the information required for screening is added, molecular analogs will be identified and an option to download all the structures is available. There is also an option to identify the characteristic features of each of the screened compounds.

ZINC offers key information, including:

- A chemical/physical description of the compound

- Chemical Vendors

- Biological targets

- Analogs and molecular comparisons

- Clinical studies

ZINC is very useful when virtual screening is required in a study to identify a more favorable compound with structural characteristics.

Unfortunately, with the new release of ZINC15, features that were previously available are now still under optimization, making the site difficult to maneuver.

There are many more chemical libraries available, including ChEMBL (https://www.ebi.ac.uk/chembl/), Chemspider (http://www.chemspider.com/) and SwissSimilarity (http://www.swisssimilarity.ch/), to name a few.

9.3 List of computer tools & Software used in this work

This section list the various software tools that were used in each of the levels of the work that was done and goes further to elaborate in detail how they function, and the hard ware that was used to make this work a success.

- The **Avogadro** software is freely available software that utilizes molecular mechanics methods to draw and optimize molecules.

- **RNA123 Version 2.0.1.3**®- RNA123 as the name suggest is a tool that combines a number of algorithms with the abilities to predict structure, analyze, visualize, structure-based sequence alignment, 3D homology modeling, and de novo modeling of the RNA. Tools for analysis in the RNA 123 suite automatically identify and correct errors in bond length, missing atoms, steric clashes among other errors. Template structures from experimental methods are analyzed automatically identifying hydrogen bonded pairs in it and classify then according to the rules of Leontis-Westhof. The RNA123 suite has a molecular mechanics unique force field that is specifically optimized for RNA allowing it to distinguish native structures from decoys. As shown in Figure 8 RNA 123 has a homology modeling algorithm validated by various sets of RNA targets which include the ribosome, riboswitches and ribozymes. This Validation includes modeling of a known ribosome structure negating all the information generated from its crystal structure, and by using a homologous template a good model was obtained and then compared with the crystal structure from the experimental method. This resulted to a low energy structure model near to the crystal structure, with correct forecast base paring, stacking and tertiary interactions. To add to this RNA 123 has a sophisticated de novo algorithm for modeling of the RNA 3D structure with fewer than~100nucleotides, which was validated using 25 RNAs and showed an average 0f 3.9 Å RMSD. As per now RNA 123 is available commercially to

both academic and industrial licensing runs on windows OS(Fredrick Sijenyi, Pirro Saro et al. 2012).

- *PyMol v0.99®* is a visualization and validation tool which calculates RMSD between atoms if molecules. PyMol is a powerful tool that provides uses ability to view 3D biomolecular images with more than 600settings and 20 representations and can interpreate over 30 different file formats form basic pdb files, multi sdf files to volumetric electron density maps. Its easy to use due to its graphical user interface allowing any type of user to create stunning 3d images from favorite file formats. Saving Images and movies can be done in a cross-platform session file, and viewed exactly as intended, with assurance that object position, colour of atom, representation of molecule,state of molecule, frame and movie viewingis precise. The 20 different ways that data ca be represented include Spheres which provides a CPK-like view,Mesh and surface which provide views that are more volumetric, Sticks and lines which put bond connectivity emphasis, cartoon and ribbon which offer popular ans easy representation for the identification of the secondary structure topology (Grell, Parkin et al. 2006; Ordog 2008; Lua 2012).

- *Discovery studio Accelrys 2.50®* is a suite for small molecule and macromolecule simulation software developed and distributed by the Accelrys Company specializing in scientific software production. Discovery studio is a client server software suite, that is built on a pipeline visual programming, entailing computational chemistry, computational biology, chemiinformatics, molecular simulation and quantum mechanics. Discovery studio major uses involves the development of novel therapeutics medicines which include small molecule drug, therapeutic antibodies, vaccines, synthetic enzymes and others such as consumer products. This property makes it be used over a number of academic and commercial entities such as pharmaceutical, biotech and consumer goods industries. It can

129

run in both windows Red Hat and suse Linux client servers. Discovery studio has a strong academic collaborative program, supportive scientific research to its suite, making a number of software algorithms originally developed in the scientific community which includes CHARMM, MODELLER, DELPHI, ZDOCK, and DMol3 among others (Studio 2009).

- *MolProbity*- Is a server that offers general purpose web service on quality validation of 3D structures of proteins, nucleic acids and other complex molecules. It was developed by the Richardson group and provides all atom detailed contact analysis of any steric problem that may be found within the molecule that is being evaluated, it calculates and displays the H-bond and van der walls contacts in the interfaces between various components. An important step in this process of validation is the ability to add and fully optimize all hydrogen atoms, both polar and non-polar. MolProbity is a very valuable structure validation tool which after the above steps relays results in multiple forms as overall numeric scores, lists, downloadable PDB and graphical files. The results are as informative, easily manipulated 3D kinemage graphics shown online in kiNG tool viewer. This server is available for all and is available freely to all users at http://kinemage.biochem.duke.edu (Davis, Murray et al. 2004; Davis, Leaver-Fay et al. 2007; Chen, Arendall et al. 2010).

- *PROCHECK version 3.6.2 ®* Is a tool developed by European Molecular Biology Laboratory that provides an idea of the stereochemistry quality of the protein chains in a given input PDB structure. It analyses the structure highlighting the protein regions appearing to have unusual geometry and give an overall detailed analysis of the structure as a whole. PROCHECK is easy to use where it allows one to input the protein PDB file and then one can navigate through to find analysis required for validating protein. It

provides four major analytical ways of providing results either as Plot description under which falls; main ramachandran plot, all residue ramachandran plots, All residue chi1-chi2 plots, main chain parameters, side chain parameters, residue properties plot, main chain bond lengths, Main chain bond angles, RMS distances from planarity, and distorted geometry. Print out; residue by residue listing under 3D view; Torsion angle G factor in 3D and finally Miscellaneous; G factor. Result can be evaluated to tell one of finding that is provided (Laskowski, Rullmannn et al. 1996).

- *Swiss-Model server®-* Is a bioinformatics structural tool that is set up as a web server that is dedicated to 3D protein homology modeling. Its fully automated and freely accesible via the ExPASy web server, or from swiss Pdb viewer Deepview. Swiss Model comprises three components integrated tightly. This are SWISS MODEL pipeline a combination of software tool and databases that allow for automated modeling of protein structures. Next is SWISS MODEL workspace a graphical web based user workbench and finally SWISS MODEL Repository which is a database that is continously updated of homology models for set of model organism proteomes of high biomedical intreast (Peitsch 1996; Nicholas Guex 1997; Arnold, Bordoli et al. 2006; Kiefer, Arnold et al. 2009).

- **SCOP** means structure classification of proteins and is used to classify proteins as the name sugests into five levels; class, fold, super family, family. Division of classes is done as alpha, all beta, alpha or beta alpha and beta domains. Super family for example does classify by convergent evolutin where there is unrelated function but same structure as the evolutionaly disctint lineage. Proteins that are orthologous with same functions but different organism; related descent from duplicate common ancestors but different function or paralogous proteins . Family comes finally where diferentiation of these paralogous proteins is done

- *YASARA®* is an acronym of Yet Another Scientific Artificial Reality Application which is a molecular graphics, simulation and modeling software available for windows, linux and Mac OS X users. It has a simple graphical user interface thatallows users to focus on their goal . Its powered by a new developmental frame work providing perfomance better than traditional software the portable vector language. This allows one to visualize much larger proteins and interactive real time simulations with highly accurate forcefields on a standard pc. Is based on academic world hence all methods are reviewed by journals (Krieger, Koraimann et al. 2002).

- **Notepad++** Is a source code editor that supports several languages, runs in a windows platform governed by GPL licence use and available for free. Its based on scintilla a powerful editing component and is written in c++ using pure Win32 API and STL ensuring higher execution speed and smaller program size. Notepad++ is user friendly although optimised in many routenes (Ho 2009).

9.4 Molecular Docking

9.4.1 Molecular Docking With Patchdock

The URL for PatchDock: http://bioinfo3d.cs.tau.ac.il/PatchDock/

NOTE- PatchDock web server may take minutes to several hours to calculate results depending on the submission queue.

Procedure:

1. First, download dimeric state HIV protease (PDB code: 3T11) in .pdb format, save it in a folder called patchdock on your desktop and rename it to HIVPR.

2. Now copy and paste the lopinavir.pdb in the same folder.

3. Click on the above link to open PatchDock web server.

4. Upload the structure of your protein (HIVPR.pdb) as a target and inhibitor (lopinavir.pdb) as a ligand.

   You need to click on "choose file" to be able to do this

5. Provide your email address since the docking results will be sent to it

6. Then click on "submit form"

7. Check your email and click on the link provided to be able to access the results – this should open a new windows like the one below

8. Docking orientation are ranked according to the binding affinities (the best docking score is the first one "result_1.pdb). Click on the best docked complex (result_1.pdb) to download the ligand-enzyme complex to your local hard drive (be aware where have you saved it)

9. Open in visualization software to view docked structure (Chimera software).

9.4.2   Molecular Docking With Chimera

Chimera is a visualization and modelling software. It can handle a wide variety of computational techniques and is especially good for generating images. The user manual can be found at this link should you want to practice more with it: https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/

Once chimera is installed, create a folder on your desktop called "Chimera docking1"

In this folder, copy and paste the "Vina folder" (in software folder). This Vina folder should always be in the folder where you are running docking. This is because, when prompted, Chimera will look for the Vina folder to access the source files. If the Vina folder is not present, docking will not run.

In the "Chimera docking1" folder, you should have the ffg files:

- The HIV protease enzyme (HIVPR.pdb)

- The lopinavir drug (lopinavir.mol2…..recognized by chimera)

134

- Vina folder

Procedure:

1. Open Chimera a file open HIVPR.pdb

2. All unnecessary ligand and substances should be removed, as to minimize and hinderance that may occur when the drug is binding.

   Go to **Select** > All Residues > All nonstandard (you will see that the small molecules are highlight is green, if nothing is highlighted, then there were no small molecules and if you delete, you will delete the entire enzyme and will have to start afresh) à **Actions** > atoms and bonds delete.

3. Now open lopinavir (File à Open àlopinavir.mol2)

4. Based on my reading, I know that the binding site of the HIV protease contains a calytic triad of ALA, TYR and ASP, so I can search for these by two options: 1- move the cursor along the enzyme ribbon, 2- Favorites- command line à type "Sel: Asp, Tyr, Ala"**.**

5. To start the docking process, go to tools à surface/ Binding Analysisà Autodock Vina

6. When the pop-up opens a click Browse and go to chimera docking folder a save with the word "complex"   (this   means   that   data   will   be   saved   as   complex.
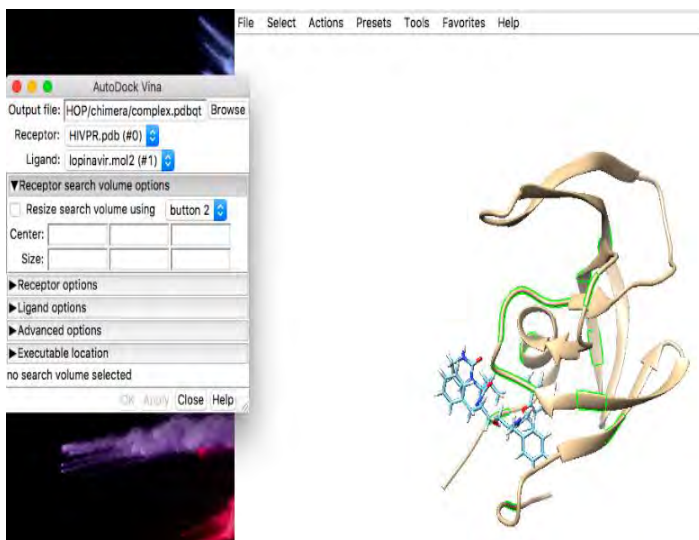
7. The receiver has to be HIVPR and the ligand can be changes to lopinavir.mol2 (use drop-down arrow).

9. Now click on resize using volume button and change all blank squares to 10. You will notice that a green box popped up on the screen. That is the gridbox for docking. It can be moved in the x, y and z direction by the mouse or by changing values in the grid.

10. Maneuver box to fit over catalytic triad.
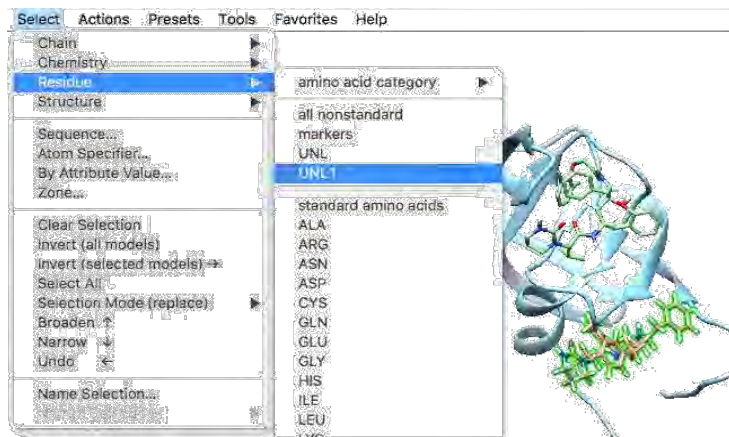
11. All receptor and ligand options should be changed to the optimized protocol (seen below)

12. The Executable location is the Opal Web service, or through the vina file, in the Vina folder located in your docking folder.

13. And then hit ok.
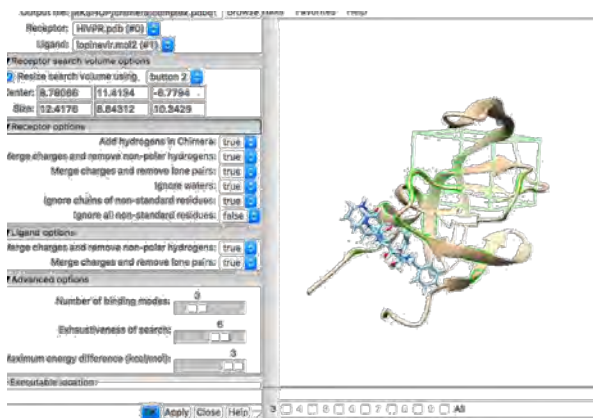
14. Once the docking has been completed, a replicate drug will be placed in proximity to the gridbox at different geometries and a score will be allocated to each conformation. The best

score is the most negative and is usually chosen as the final conformation for the docked structure.
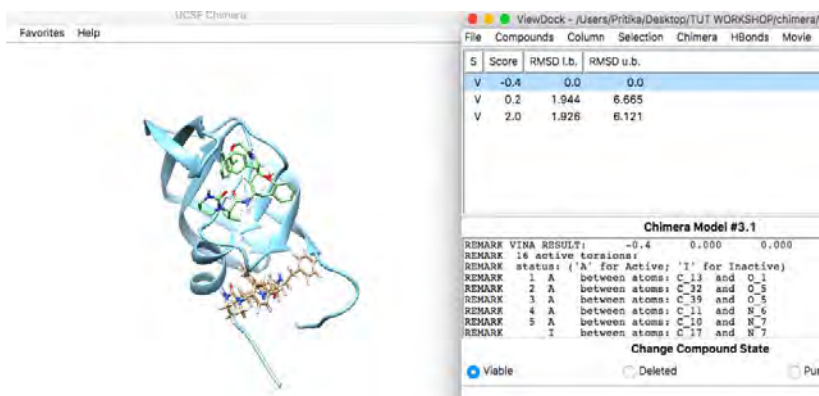
15. The original compound can be deleted by clicking SelectàResidues à clicking the correct compound (will light up green) à Actions atoms and bonds à delete.

16. Click File à save pdb à next to "save models" select HIVPR and model 1(hold shift and click both) à make sure it saves as a single file à call it complex à click save (see below).
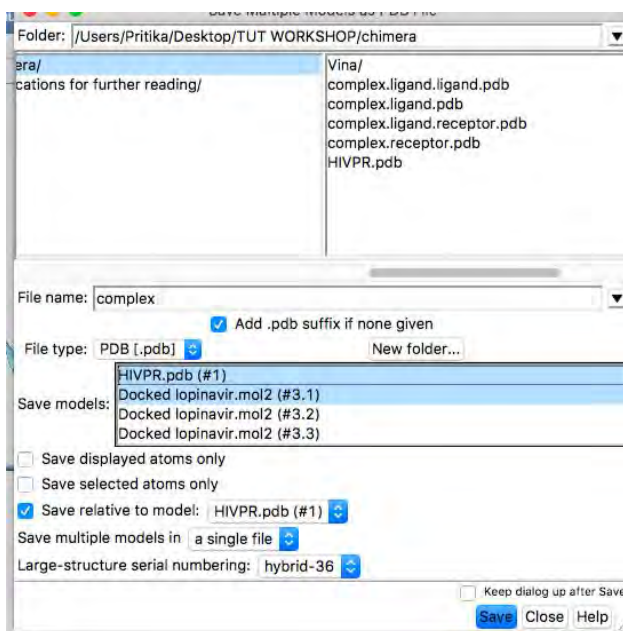
17. Close chimera and re-open it. à file open à complex.pdb à you should have a single file with the compound and enzyme. You can now change colors and formats.

18. Docking score should be noted as x kcal/mol.

The Chimera / Autodock plugin is a more widely used tool and has proven to be more accurate when docking compounds to a known active site

9.5 Work plan

9.6 Work Plan objective and tasks description

Objective 1: Target Selection. One RNA subdomain will be selected as a prioritized target from the four potential targets chosen from the RNA "regions of interest" identified in Phase 1 project.	
Activities	1. Construct a high resolution functional rRNA homology model crystal structure, rRNA motifs drug targets database for design drugs using homology modeling. 2. Homology and *de novo* modeling of the three dimensional structure of the 18S rRNA from any of the mentioned pathogens. Target will be chosen based on structural similarity with currently available crystal structures.
Milestone:	Atomic level structure information on Trypanosome and Leishmania

Objective 2: Target selection & refinement:	
Activities	1. We will visually and *in silico* inspect generated models for structural consistency and composition of active sites. This will shed light into possible binding sites for candidate ligands. 2. Use homology and *De novo* modeling to determine the essential structural components of the target. RNA motifs will be selected as a prioritized target from the potential targets chosen from the RNA "regions of interest" identified. 3. We will use software tools to probe the known functionally important regions of the rRNA
Milestone:	Prioritized compound against selected target, with backup lower target

Objective 3: Target screening: *In silico* screening of targets against the virtual pathogen box compound libraries using Schrödinger suite & rDOCK. RNA specific docking tool

Activities	1. Screening of potential binding ligands will be performed on the targets, together with the solved *Homo sapiens* 18S rRNA structure as a control to identify specific binders.
	2. Validation using targets and compounds: We will perform Comparative docking study using 3D structure of compounds and coupled rRNA homology modeling for application to the discovery of rRNA Ligands. Structural studies of target/hit complexes, optimization of hit compounds, and validation of the target using *in vitro* and *in vivo* assays.
Milestone	Identification of hits that have specificity for the ribosome target

Objective 4: Validation of targets and compounds: We will perform validation of the target using *in vitro* and *in vivo* assays.	
Activities	1. We will perform validation of the target using in vitro and in vivo assays.
	2. Determine the nature of the interaction of the hits with their targets using compound arrays and by capillary electrophoresis gel retardation binding analysis of lead-target complexes
Milestone	One or more compounds that are potential leads, and target validated, 15-24 months.

9.7 Project milestones that define key decision points and/or the success of the project

Milestone 1 - Atomic level structure information on Trypanosome and Leishmania

Planned month (from project start): October 2016

Milestone 2 - Prioritized compound against selected target, with backup lower target

Planned month (from project start): April 2017

Milestone 3 - Identification of hits that have specificity for the ribosome target

Planned month (from project start): November 2017

Milestone 4 - One or more compounds that are potential leads, and target validated, 15-24 months.

Planned month (from project start): April 2018- August 2019

Work Package No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	29	30	31	32	33	34	35	36	Site
WP1	■	■	■	■	■	■	■	■																												
WP2						■	■	■	■	■	■	■	■	■	■	■																				
WP3														■	■	■	■	■	■	■	■	■	■	■												
WP4																								■	■	■	■	■	■	■	■	■	■	■	■	

* Work Package numbers – matching with the Work Plan described above

Project Funding over stated work plan

NAME																																						

9.9 RESEARCH GOALS

RESEARCH GOALS		SON 2016	D 2016	JFM 2017	AMJ 2017	JAS 2017	OND 2017	JFM 2018	AMJ 2018	JAS 2018	OND 2018	JFM 2019	AMJ 2019
Phase I: Project Development, Data Collection & Atomic level structure information acquisition on Trypanosome and Leishmania													
Review computer modeling of biological systems literature		X											
Review of Trypanosome and Leishmania ribosome and their importance to drug discovery		X											
Reconnaissance work - *Experimental methods + theoretical* methods of obtaining biological molecules structure		X	X										
Attend training on experimental, theoretical methods of obtaining structures, and Drug docking				X									
Begin using Homology and de novo modeling software to familiarize myself with chosen pathogens				X									
Explore existing ribosome models to be used as template				X	X								
Literature review of the experimental methods and theoretical histories and their mergers in research				X	X								
Compile preliminary information on the processes of entire project				X	X	X							
Plan and execute modeling work on Trypanosome and Leishmania ribosomal 18S rRNA Atomic level structure				X	X	X							
Present at a departmental seminar				X	X	X							
Submit research Thesis/Dissertation				X	X	X							
Phase 2: Prioritized compound against selected target, with backup lower target & Identification of hits that have specificity for the ribosome target													
Begin visual and in silico inspection of the models for structural consistency and compound active sites							X						
Begin target selection and refinement analyses of Trypanosome and Leishmania ribosomal 18S rRNA Atomic level structure								X					
Present talk at the Biochemistry department, University of Nairobi								X	X				
Prepare manuscript on homology and de novo model results								X					
Complete follow up on rRNA motif selection as a prioritized potential target								X					
Prepare manuscripts on important regions of the Trypanosome and Leishmania ribosomal 18S rRNA in context to drug screening								X					
Target Compound Modeling Course and begin to model the selected compounds to be screened									X				
Adjust course material to improve this project									X				
Present at chemistry department Meeting, University of Nairobi									X				
Begin Target screening against compound libraries using docking software on the Trypanosome and Leishmania ribosomal 18S rRNA									X	X			
Perform docking validation structure studies and optimization of hit compounds on the target										X	X		
Prepare manuscript on in silico screening deviations and success										X			
Compare findings to similar studies on a global scale, discuss the finding with various stakeholders in field of drug discovery										X	X		
Present talk at the departmental meeting on the success and progress of the work										X	X		
Phase 3: Completion One or more compounds that are potential leads, and target validated													
Begin *in vitro* assay validation of target compounds of Trypanosome and Leishmania ribosomal 18S rRNA											X	X	
Determine the nature of interaction using compound array and by capillary electrophoresis gel retardation binding analysis of lead-target complexes											X	X	
Publish papers, assemble thesis, compile questions and issues to explore in future research													X
Thesis write-up and editing													X

141

9.10 Budget

	Budget Item	Amount	Sub-Total	Justification
a	**Equipment**			Most of this category of equipment is available in the laboratory the cost of each equipment listed hence this could be accounted for as provisional equipment required total requested amount is **$7,499.00** for the HIGH END Mac pro desktop computers and mac pro laptop
	HIGH END Mac pro desktop computers and mac pro laptop	$7,499.00		
	Agilent HPLC 1100 system with diode array detector	$6,499.99		
	Agilent capillary electrophoresis G1601 unit	$5,299.00		
	PCR thermocycler	$14,500.00		
	gel documentation system	$7,725.00		
	HITACHI UV/Visible spectrophotometer	$2,495.00		
	HICEN high speed ultra centrifuge	$3,820.00		
	Aclirimide gel electrophoresis system	$2,980.00		
	-80 degree's freezer	$3,050.00		
			$53,867.00	
b	**Software**			Homology and de novo modeling, potentials for protein-RNA docking, compound screening
	RNA123®	$4,490.00		
	Accelrys® -Discovery Studio®	$6,490.00		
	Schrödinger's suite®	$4,000.00		
	ChemOffice® 14	$4,000.00		
	rDock®	$2,500.00		
	ModeRNA ®	$1,480.00		
	DARS-RNP® and QUASI-RNP®,	$3,900.00		
	Computer data Protection SOFTWARE	$490.00		
			$23,350.00	
c	**Materials and consumables**			**$6,165.00**
	Reagents required in the in vitro process	$4,030.00		
	Reagents Measuring equipment, Viles, plates, glassware apparatus, pipettes, Gloves among others	$2,135.00		
			$6,165.00	
d	**Printing & Publishing**			**$1,690.00**
	Laboratory note books	$20.00		
	Stationary estimates	$100.00		
	Binding (hard and loose)	$480.00		
	Printing and photocopying services	$1070.00		
	Box files	$20.00		
			$1,690.00	
e	**Workshops and Conferences**			**$1,250.00**
	Registration fees and Presentation	$1,250.00		
			$1250.00	
f	**Direct Expenses**:			**$400.00**
	Transcription of training material and information	$400.00		

	Total Direct Expenses			
g	**Travel**			$960.00
	Domestic Travel (Transportation, Accommodation, Meals) Domestic Trips for conferences and workshops	$960.00		
	Total Travel Expenses		$960.00	
	TOTAL EXPENSES		$1,360.00	$1,360.00
h	**Overhead:** @ 10% overseen cost		$4,131.00	$4,131.00
	TOTAL REQUEST:			$45,445.00

$1 = 100ksh (As per time of preparation)

Kindly note some of the equipment to be used in the project are available in the institution hence the only item required is the HIGH END Mac pro desktop computers and mac pro laptop, therefore the total request as by the start of project in the institution is **$45,445.00**

9.11 >M12676.1 *T.brucei* small subunit rRNA gene

GATCTGGTTGATTCTGCCAGTAGTCATATGCTTGTTTCAAGGACTTAGCCATGCATGCCTCAGA
ATCACTGCATTGCAGGAATCTGCGCATGGCTCATTACATCAGACGTAATCTGCCGCCAAAATTT
TGCGGTCTCCGCATTACTGGATAACTTGGCGAAACGCCAAGCTAATACATGAACCAATCGGAC
GCTCTCTTTTTCTATGTCGCGGCTTGTGTTTACGCACTTGTCGTGGCGATGAGACGTCCAGCGA
ATGAATGAAATTAGAACCAACGCCTCCACCCGGGGGCAGTAACACTCAGACGTGTTGACTCA
ATTCATTCCGTGCGAAAGCCGAGCCTTTTTGCTCGGCGTCTACTGACGAACAACTGCCCTATC
AGCCAGTGATGGCCGTGTAGTGGACTGCCATGGCGTTGACGGGAGCGGGGGATTAGGGTTCG
ATTCCGGAGAGGGAGCCTGAGAAATAGCTACCACTTCTACGGAGGGCAGCAGGCGCGCAAAT
TGCCCAATGTCGAAAAAATACGATGAGGCAGCGAAAGAAATAGAGCCGACCGTGCCCTAGT
GCATGGTTGTTTTCAATGGGGGATACTCAAACCCATCCAATATCGAGTAACAATTGGAGGACA
AGTCTGGTGCCAGCACCCGCGGTAATTCCAGCTCCAAAAGCGTATATTAATGCTGTTGCTGTTA
AAGGGTTCGTAGTTGAACTGTGGGCCACGTAGTTTTGTGCCGTGCCAGTCCCGTCCACCTCGG
ACGTGTTTTGACCCACGCCCTCGTGGCCCGTGAACACACTCAGATACAAGAAACACGGGAGC
GGTTCCTCCTCACTTTCACGCATGTCATGCATGCGAGGGGGCGTCCGTGAATTTTACTGTGAC
CAAAAAAGTGCGACCAAAGCAGTCTGCCGACTTGAATTACAAAGCATGGGATAACGAAGCAT
CAGCCCTGGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGGCC
GCGTGCCTTGGGTCGGTGTTTCGTGTCTCATTTTTGTGGCGCGCACATTCGGCTCTTCGTGATG
TTTTTTTACATTCATTGCGACGCGCGGCTTCCAGGAATGAAGGAGGGTAGTTCGGGGGAGAAC
GTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTACAGCGAAGGCATT
CTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCAT
TGTAGTCCACACTGCAAACCATGACACCCATGAATTGGGGAACATCATTGGGTGCCCGTGTGG
CGGCCTTTTGTGCCGACCCTCGGCCCCAATTTATTTATCAATTTACGTGCCTATTCTATCACCCC
CGGTTCCCTCTTTTGAGGTTCTTCCGGGGTTTTTTACGGGAATATCCTCAGCACGTTTCTTACT
TCTTCACGCGAAAGCTTGGAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACT
TAAAGAAATTGACGGAATGGCACCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACAC
GGGGAACTTTACCAGATCCGGACAGGGTGAGGATTGACAGATGGAGTGTTCTTTCTCGATCCC
CTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTGATTCCGTCAAC
GGACGAGATCCAAGCTGCCCAGTAGGTGCCGGGATTGTCCACACAGGACAGCAGTCCCTCCG
GCGGGGATTTTTTCCCCAACGGTGGTCGTCATCCTTCTTTTTACAGGCCCCTTCTCTGCGGGAT
TCCTTGCTTTTCGCGCAAGGTGAGATTTTGGGCAACAGCAGGTCTGTGATGCTCCTCAATGTT
CTGGGCGACACGCGCACTACAATGTCAGTGAGAACAAGAGTCCGAGCGGCACTTCACAATGT
CGCTCCCGCTTGATCAAAAGAGCGGGGAAACCACGGAATCACGTAGACCCACTTGGGACCGA
GTATTGCAATTATTGGTCGCGCAACGAGGAATGTCTCGTAGGCGCAGCTCATCAAACTGTGCC
GATTACGTCCCTGCCATTTGTACACACCGCCCGTCGTTGTTTCCGATGATGGTGCAATACAGGT
GATCGGACCGTCGCTCGTCTCGGGCGACCGAAAGTTCACCGATATTGCTTCAATAGAGGAAGC
AAAAGTCGTAACAAGGTAGCTGTAGGTGAACCTGCAGCTGGATCATTT

>AF245382.1 Trypanosoma cruzi strain Dm28c 18S ribosomal RNA gene, complete sequence

AACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTTTCAAGGACTTAGCCATGCATGCCTCAGA
ATCACTGCATTGCAGGAATCTGCGCATGGCTCATTACATCAGACGTAATCTGCCGCAAAAATCT

TGCGGTCTCCGCAAAATTGGATAACTTGGCGAAACGCCAAGCTAATACATGAACCAACCGGAT
GTTCTCTGTTCCGGCGGCAGGGCAACCTGCTGCCATGGGACGTCCAGCGAATGAATGAAAGT
AAAACCAATGCCTTCACCGGGCAGTAACACTCAGAAGTGTTGATTCAATTCATTCCGTGCGAA
AGCCGGGTTTTTTATCCGGCGTCTTTTGACGAACAACTGCCCTATCAGCCAGCGATGGCCGTG
TAGTGGACTGCCATGGCGTTGACGGGAGCGGGGGATTAGGGTTCGATTCCGGAGAGGGAGCC
TGAGAAATAGCTACCACTTCTACGGAGGGCAGCAGGCGCGCAAATTGCCCAATGTCAAAAAA
AAAAGATGAGGCAGCGAAAGAAATAGAGCCGACAGTGCTTTTGCATTGTCGTTTTCAATGG
GGGATATTTAAACCCATCCAAAATCGAGTAACAATTGGAGGACAAGTCTGGTGCCAGCACCCG
CGGTAATTCCAGCTCCAAAAGCGTATATTAATGCTGTTGCTGTTAAAGGGTTCGTAGTTGAATT
GAGGGCCTCTAAGGCGCAATGGTTTAGTCCCATCCACTTCGGATTGGTGACCCATGCCCTTGT
GGTCCGTGAACAGACATATTCAGAAACAAAAAACACGGGAGTGGTACCTTTTCTGATTATCGC
ATGTCATGCATGCCAGAGGGCGCCCGTGATTTTTTACTGTGACTAAAAAAGTGTGACCAAAGC
AGTCATTCGACTTGAATTAGAAAGCATGGGATAACAAAGGAGCAGCCTCTGGGCCACCGTTT
CGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCAGTGTGACAAGCGGCTGGGTG
GTTATTCCACACACGCACACTCCTTTTTTGGATGTTGTGTCTCTGTGTGTGTGGCACTCGTC
GCCTTTGTGGGAAATCCGTGTGGCACTTGTTTGGTGTTGTTGGCAGACTTCGGTCTTGCCTTC
ACACACGTTTCACATGTGTCATGCCTTCCCTCAACTCACGGCATCCAGGAATGAAGGAGGGTA
GTTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACT
ACAGCGAAGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAG
ATGATTAGAGACCATTGTAGTCCACACTGCAAACGATGACACCCATGAATTGGGGAGTTTTTG
GTCGTAGGCGTGGTCGGGCTTGATTATATTATTTTTCATCCCGTTCCTCGTCTCGCCAATGAATA
TATTAAATTTACGTGCATATTCTTTTTGGTCTTCGTTTTTTTACGGCGAGGACCTTTAACGGGAA
TATCCTCAGCACGTTATCTGACTTCTTCACGCGAAAGCTTTGAGGTTACAGTCTCAGGGGGGA
GTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAATGGCACCACAAGACGTGGAGCG
TGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATTGACA
GATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTG
ATTTGTTTGGTTGATTCCGTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCC
CATAGGATAGCAATCCCTTCCGCGGGTTTTACCCAAGGGGGGGCGGTATTCGCTTGTATCCTTC
TCTGCGGGATTCCTTGTTTTGCGCAAGGTGAGATTTTGGGCAACAGCAGGTCTGTGATGCTCC
TCAATGTTCTGGGCGACACGCGCACTACAATGTCAGTGAGAACAAGAAAAACGACTCTTGTC
GGACCTACTTGATCAAAAGAGTGGGAAAACCCCGGAATCACGTAGACCCACTTGGGACCGAG
TATTGCAATTATTGGTCGCGCAACGAGGAATGTCTCGTAGGCGCAGCTCATCAAACTGTGCCG
ATTACGTCCCTGCCATTTGTACACACCGCCCGTCGTTGTTTCCGATGATGGTGCAATACAGGTG
ATCGGACAGTCGAGTGCTTCACTTGACCGAAAGTTCACCGATATTTCTTCAATAGAGGAAGCA
AAAGTCGTAACAAGGTAGCTGTAGGTGAACCTGCAGAAGGATCAAGCTT

>AC005806.3:10707-12910 *Leishmania major* strain Friedlin chromosome 27 clone L7886, complete sequence

GATCTGGTTGATTCTGCCAGTAGTCATATGCTTGTTTCAAGGACTTAGCCATGCATGCCTCAGA
ATCACTGCATTTGCAGGAATCTGCGCATGGCTCATTACATCAGACGTAATCTGCCGCAAAAATC
TTGCGGTTTCCGCAAAATTGGATAACTTGGCGAAACGCCAAGCTAATACATGAACCAACCGGG

TGTTCTCCACTCCAGACGGTGGGCAACCATCGTCGTGAGACGCCCAGCGAATGAATGACAGT
AAAACCAATGCCTTCACTGGCAGTAACACCCAGCAGTGTTGACTCAATTCATTCCGTGCGAAA
GCCGGCTTGTTCCGGCGTCTTTTGACGAACAACTGCCCTATCAGCTGGTGATGGCCGTGTAGT
GGACTGCCATGGCGTTGACGGGAGCGGGGGATTAGGGTTCGATTCCGGAGAGGGAGCCTGAG
AAATAGCTACCACTTCTACGGAGGGCAGCAGGCGCGCAAATTGCCCAATGTCAAAACAAAAC
GATGAGGCAGCGAAAAGAAATAGAGTTGTCAGTCCATTTGGATTGTCATTTCAATGGGGGATA
TTTAAACCCATCCAATATCGAGTAACAATTGGAGGACAAGTCTGGTGCCAGCACCCGCGGTAA
TTCCAGCTCCAAAAGCGTATATTAATGCTGTTGCTGTTAAAGGGGTTCGTAGTTGAACTGTGGGC
TGTGCAGGTTTGTTCCTGGTCGTCCCGTCCATGTCGGATTTGGTGACCCAGGCCCTTGCAGCC
CGTGAACATTCAAAGAAACAAGAAACACGGGAGTGGTTCCTTTCCTGATTTACGCATGTCATG
CATGCCAGGGGGCGTCCGTGATTTTTTACTGTGACTAAAGAAGCGTGACTAAAGCAGTCATTT
GACTTGAATTAGAAAGCATGGGATAACAAAGGAGCAGCCTCTAGGCTACCGTTTCGGCTTTTG
TTGGTTTTAAAGGTCTATTGGAGATTATGGAGCTGTGCGACAAGTGCTTTCCCATCGCAACTTC
GGTTCGGTGTGTGGCGCCTTTGAGGGGTTTAGTGCGTCCGGTGCGAGCTCCGGTTCGTCCGG
CCGTAACGCCTTTTCAACTCACGGCCTCTAGGAATGAAGGAGGGTAGTTCGGGGGAGAACGT
ACTGGGGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTACAGCGAAGGCATTCT
TCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGAGATCGAAGATGATTAGAGACCATTG
TAGTCCACACTGCAAACGATGACACCCATGAATTGGGGATCTTATGGGCCGGCCTGCGGCAGG
GTTTACCCTGTGTCAGCACCGCGCCCGCTTTTACCAACTTACGTATCTTTTCTATTCGGCCTTTA
CCGGCCACCCACGGGAATATCCTCAGCACGTTTTCTGTTTTTTCACGCGAAAGCTTTGAGGTT
ACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAATGGCAC
CACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGAC
AGGATGAGGATTGACAGATTGAGTGTTCTTTCTCGATTCCCTGAATGGTGGTGCATGGCCGCT
TTTGGTCGGTGGAGTGATTTGTTTGGTTGATTCCGTCAACGGACGAGATCCAAGCTGCCCAGT
AGAATTCAGAATTGCCCATAGAATAGCAAACTCATCGGCGGGTTTTACCCAACGGTGGGCCGC
ATTCGGTCGAATTCTTCTCTGCGGGATTCCTTTGTAATTGCACAAGGTGAAATTTTGGGCAACA
GCAGGTCTGTGATGCTCCTCAATGTTCTGGGCGACACGCGCACTACAATGTCAGTGAGAACA
AGAAAAACGACTTTTGTCGAACCTACTTGATCAAAAGAGTGGGGAAACCCCGGAATCACATA
GACCCACTTGGGACCGAGGATTGCAATTATTGGTCGCGCAACGAGGAATGTCTCGTAGGCGC
AGCTCATCAAACTGTGCCGATTACGTCCCTGCCATTTGTACACACCGCCCGTCGTTGTTTCCGA
TGATGGTGCAATACAGGTGATCGGACAGGCGGTGTTTTATCCGCCCGAAAGTTCACCGATATT
TCTTCAATAGAGGAAGCAAAAGTCGTAACAAGGTAGCTGTAGGTGAACCTGCAGCTGGATCA
TTTT