SCREENING FOR PUTATIVE VACCINE CANDIDATES OF *Onchocerca volvulus* USING GENOMIC AND PROTEOMIC ANALYSES

BY

Abraham SHEYIN, BSc (ABU) 2011, PGDE (NTI) 2013

P13SCBC8002

A THESIS SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES, AHMADU BELLO UNIVERSITY, ZARIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE DEGREE IN BIOTECHNOLOGY

DEPARTMENT OF BIOCHEMISTRY,

FACULTY OF SCIENCE,

AHMADU BELLO UNIVERSITY,

ZARIA, NIGERIA

JULY, 2018

Abstract

Onchocerca volvulus, a parasitic worm, is the causative agent of onchocerciasis (river blindness) which is transmitted by blackflies of the genus *Simulium* to humans. It is a neglected tropical disease affecting more than 37 million people worldwide. Nigeria accounts for about 40 percent of the global Onchocerciasis burden and is the most endemic country in the world. There is no vaccine yet for the prevention of the disease. Treatment is done by administering Ivermectin that kills only the microfilaria, but does not kill the adult worm. The danger of resistance to this drug has been reported in Ghana. The development of a vaccine would help solve these challenges. In this study, reverse vaccinology, was used to screen the genome and proteome of *O. volvulus* using bioinformatics tools so as to find proteins of the pathogen that can serve as vaccine candidates. From 6259 proteins of *O. volvulus* in the Uniprot database 4551 uncharacterized protein sequences were collected. Using Subtractive Genomics Approach, NCBI tblastn program was used to screen out sequences similar to humans, to avoid the occurrence of autoimmunity. Hence, 793 uncharacterized protein sequences that are non-orthologues to humans were identified. From these sequences, signal peptides were predicted using the Linux programs: TargetP, SignalP and Phobius. Transmembrane helices were also predicted using the Linux programs: TMHMM and Phobius. As a result, 22 putative proteins containing both signal peptides and a single transmembrane helice (literature shows that the more the helices in a protein the less the correctness of prediction) were subjected to Major Histocompatibility Complex (MHC) I and II peptide binding prediction and B-cell epitopes prediction. This was done using the Linux programs: Immune Epitope Database MHC I and II binding predictors and BepiPred respectively. Consequently, 133 epitopes were identified. Upon interpreting NCBI tblastn result, we suggest that four of the sequences function as components of the impervious cuticle of the pathogen. Furthermore, using comparative genomics via sequence similarity search approach, previously reported potential vaccine candidates of *Dirofilaria immitis* and *Brugia malayi* were compared with the genome of *O. volvulus* using the NCBI tblastn program. As a result, 1 more putative vaccine candidate - a characterized protein, was identified. This brings the number of putative vaccine sequences identified to 23. Experimental validation of these candidates is hereby recommended.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background

The common practice in vaccine production is based on the method used by Louis Pasteur in 1881 to develop a rabies vaccine. It entails the isolation, inactivation and injection of the disease-causing organism (Cafardi *et al.,* 2013) into an appropriate animal model. Josefsberg and Buckland (2012) have also pointed out that from way back, vaccines have been developed based on methods that follow the pattern of isolating, sometimes inactivating and injecting the disease-causing pathogen or pathogen component into the animal model. There are instances where this approach has yielded positive results, but it usually takes a lot of time before a vaccine is eventually developed. Also, the disease-causing organism must be cultivated in a laboratory environment. And more often than not, only proteins that are in large amount are identified(Bertholet *et al.,* 2015). In this study, a faster approach, known as reverse vaccinology, to vaccine development was used. Seib *et al.,* (2009) defined reverse vaccinology as the *in silico* screening of the total genome of a disease-causing organism in order to find DNA sequences encoding proteins which can serve as good vaccine candidates, by working on either the genome of the same pathogenic isolate or the pan-genome (genomic information from different isolates) of a species of pathogen. The reverse vaccinology approach to vaccine production is faster. For instance, in only 18 months, the *Neisseria meningitides* serogroup B (Men B) reverse vaccinology project identified more novel vaccine targets than had been discovered in 40 years of conventional vaccinology (Seib *et al.,* 2009). Also, applying the concept of reverse vaccinology to the heartworm, *Dirofilaria immitis*, revealed drug and vaccine targets (Godel *et*

al., 2012). Here, the idea of reverse vaccinology was applied in analyzing the proteins and genome of *Onchocerca volvulus*, the causative agent of onchocerciasis. Screening of proteins was done using signal peptide and transmembrane helice predictors as well as other bioinformatics programs, while the genome was screened using the NCBI tblastn program.

Onchocerciasis, also known as river blindness is a neglected tropical disease that is prevalent in many sub-Saharan African countries, with minor cases in Yemen, Central and South America. It affects more than 37 million people worldwide (Globisch *et al.,* 2013). The symptoms of the disease include severe itching, disfiguring skin conditions, visual impairment as well as permanent blindness (WHO, 2015b). Since 1989, the Nigerian National Onchocerciasis Control Programme (NNOCP) has been fighting onchocerciasis (Carter Centre, 2015). The Carter Centre, a nongovernmental organization, has also been involved in the fight against onchocerciasis in Nigeria, and in 2014 distributed Mectizan to about 7 million people across nearly 16, 000 villages (Carter Centre, 2015).

1.2 Distribution of onchocerciasis in Nigeria

The World Health Organization has described onchocerciasis as a country-wide public health problem in Nigeria and has shown that in about 30 states, community directed treatment with Ivermectin (CDTI) is needed (WHO, 2015b). Although the Northern part of the country is the worst hit by the disease, some parts of Ebonyi state (Ebonyi Central and North Senatorial zones) in Nigeria have been shown to be onchocerciasis hypo-endemic (low incidence) zones in the state (Okoro *et al.,* 2012). An examination of the map of Nigeria showing the web of onchocerciasis shows that the disease is uncommon in Lagos and Bayelsa states.

1.3*Onchocerca volvulus*

O. volvulus is a parasitic nematode that is transmitted by infected blackflies of the *Simulium* spp. These blackflies live near and breed in fast flowing streams and rivers that are close to fertile land suitable for agriculture (WHO, 2015b). The third-instar larva that infects is transmitted by the blackfly during a blood meal (Mackenzie *et al.*, 2011). These larvae develop into adults in the human host, ultimately accumulating in subcutaneous or deep nodules called onchocercomas (Globisch *et al.*, 2013). According to Shapiro (2012), the female worm is about 33 to 50 cm in length and 270 to 400 μm in diameter, while male worms measure about 19 to 42mm in length and 130 to 210μm in diameter. *O. volvulus* is treated with Ivermectin which although cannot kill the adult worms, paralyses and kills microfilaria, relieving intense skin itching and halting the progression towards blindness, also preventing adult worms from producing more microfilariae for a few months following treatment, hence reducing transmission (WHO, 2015b).

1.4 Life Cycle of *Onchocerca volvulus*

O. volvulus causes infection when a third-instar larva of its life cycle is transmitted to humans via the bite of the blackfly, consequently resulting into onchocerciasis (CDC, 2016).

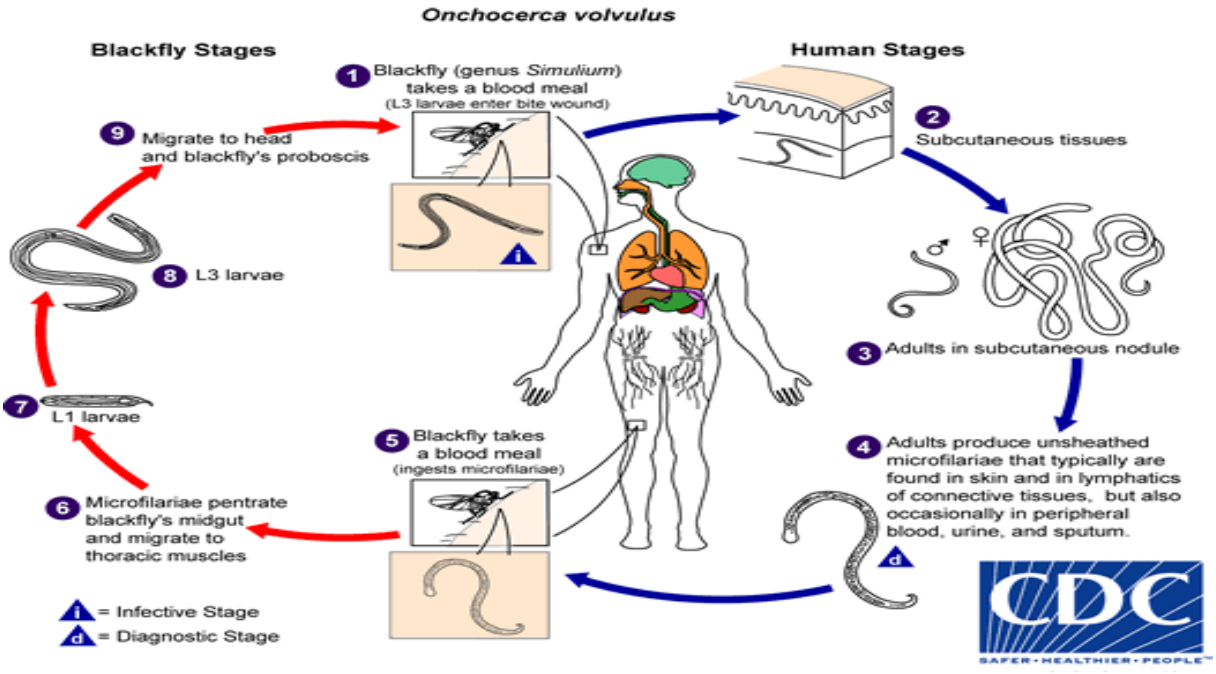


Figure 1. Life cycle of *Onchocerca volvulus*. Adapted from www.cdc.gov/dpdx/onchocerciasis/.

An infected blackfly injects a third stage filarial larvae into the human host during a blood meal. The microfilariae develop into adult worms in the subcutaneous tissues. The adult female worms produce many microfilariae which occasionally are found in peripheral blood, urine, and sputum but are typically found in the skin and in the lymphatics of connective tissues. During a blood meal a blackfly ingests the microfilariae which then moves from the blackfly's midgut through the hemocoel to the thoracic muscles. In the thorax, it develops into first-stage larvae and then into third-stage infective larvae which then migrates to the blackfly's proboscis, ready to infect another human during the fly's blood meal (CDC, 2013).

1.5 Statement of Research Problem

Africa harbors 96% out of some 120 million people throughout the world who are at risk of Onchocerciasis (WHO, 2015b). Also, 99% of the 18 million people infected with the disease and have dermal microfilaria are in Africa and of the 18 million people infected, more than 6.5 million suffer severe itching or dermatitis and 270,000 are blind (WHO, 2015b). The disease is implicated in musculoskeletal pain, reduced body mass index and decreased work productivity (Basanez *et al.,* 2006). More so, Nigeria accounts for about 40 percent of the global Onchocerciasis burden and is the most endemic country in the world (SEOF, 2014). So far, a vaccine to prevent the infection is currently not available. The available drugs give temporary relief, not cure (Cho-Ngwa *et al.,* 2010). Ivermectin is a popular drug used to treat the disease and it kills the microfilaria worms without killing the adult worm (Albers *et al.,* 2012). Furthermore, Makepeace *et al.,* (2009) and Gloeckner *et al.,* (2010) have reported the possibility of Ivermectin resistance emerging.

1.6 Justification

This study is significant because a vaccine against the disease does not exist. Prevention is by avoiding being bitten by flies (Cho-Ngwa *et al.,* 2010). Also, vaccine production will help solve the problem about the possible emergence of Ivermectin resistance (Makepeace *et al.,* 2009). In addition, the conventional approach to vaccine development is more time-consuming (Bertholet *et al.,* 2015, Kanampalliwar *et al.,* 2013). Lastly, a vaccine against the disease will enhance disease prevention which is the most effective approach for health (Kanampalliwar *et al.,* 2013).

1.7 Aim and Objectives

1.7.1 Aim

The aim of this work was to screen the genome and uncharacterized protein sequences of *Onchocerca volvulus* using bioinformatics tools in order to identify putative vaccine targets.

1.7.2 Specificobjectives

The specific objectives of this study were to:

1. Collect uncharacterized protein sequences of *Onchocerca volvulus* not similar to human proteins.
2. Predict secretory signal peptides, transmembrane helices, T cell and B cell epitopes *in silico*.
3. Compare orthologues using *Brugia malayi* and *Dirofilaria immitis in silico.*

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Onchocerciasis (river blindness)

2.1.1 History of onchocerciasis

Kluxen (2011) provides a detailed history of onchocerciasis. He points out that the discovery of the microfilaria in 1874 by John O'Neill, a British naval surgeon attached to His Majesty's Ship (HMS) Decoy at Cape Coast Castle in the Gold Coast (Ghana) was an important event that expanded the frontiers of knowledge of onchocerciasis. The British naval surgeon became interested in studying an irritating and intractable skin disease, known by the locals as 'craw-craw' that resembled scabies which inflicted many people living in parts of the West Coast of Africa. His determination to find the cause of the disease made him study the characteristics of the condition which include papules, vesicles and pustules in patients in Addah Fort Hospital under the care of Dr Thompson of the Glover Expedition. Examination of the contents of pustules and vesicles yielded nothing except leucocytes. However, examination of papules led to the discovery of an organism which he believed was the causative agent of the disease and he reported his observations in 1875. The discovery by John O'Neill was reaffirmed 50 years later by Montpellier and Lacroix who constantly found microfilaria in the dermal layer of the skin of Africans suffering from a form of itch or 'craw-craw'. Hence, they postulated that the 'craw-craw' described by O'Neill was a dermal manifestation of onchocerciasis termed 'gâle filarienne.'

Patrick Manson, a Scottish physician, discovered the adult worm in 1890. From two local inhabitants in Ghana, West Africa, he removed two tumors, one from the scalp and the other

from the chest. Each of the tumors was about the size of a pigeon's egg. He examined the specimens and discovered they contained worms. He sent the worms to Radolf Leuekart in Germany for identification. Several female worms of about 6-70 mm in length and male worms of about half the size of the females were discovered in the tumors. They were confirmed to be adult worms of *O. volvulus*.

2.1.2 Distribution of onchocerciasis

Black flies, the vectors of onchocerciasis, are naturally found close to fast-flowing streams and rivers in the inter-tropical zones (region close to the equator where the northeast and southeast trade winds converge) and their location determines the distribution of onchocerciasis (WHO, 2015b). Since sub-Saharan Africa has a tropical climate, countries in the region bear the burden of the disease with about 30 countries affected (CDC, 2013a). In the Americas, onchocerciasis is still found in limited parts of Venezuela and Brazil (Mectizan Donation Program, 2015). Onchocerciasis is also found in Yemen in the Arabian Peninsula (WHO, 2015b).

2.1.3 Attempts at onchocerciasis control

Etya'ale (2001) observed that the strategies employed globally to combat onchocerciasis are based on *Similium* vector control and large-scale chemotheraphy with Ivermectin (donated free by Merck and Co) and that each could be done separately or in combination. In an effort to eliminate onchocerciasis, the World Health Organization (WHO) and the World Bank in 1974 launched the Onchocerciasis Control Programme (OCP). The strategy used by the programme was weekly aerial larviciding of the breeding sites of the disease vector (*S. damnosium*) with Temephos® (AgrEvo, France) and other environmentally friendly insecticides in rotation and

supplementation by ground larviciding where feasible (Richards *et al.*, 2001). Another programme, the African Programme for Onchocerciasis Control (APOC) was instituted in 1995 with the aim of controlling the disease in some African countries. The APOC and OCP worked together and by the year 2005 distributed Mectizan drugs, which prevents and lessens the conditions of the disease, to more than 45 million people in Sub-Saharan Africa (C.G.D., 2007). In Nigeria the government has been partnering with non-governmental organizations such as Sightsavers to distribute Ivermectin in endemic communities. In 1986, the NOCP was active in the fight against onchocerciasis (Edungbola, 1991). Meredith *et al.*, (2012) reported a decline in prevalence to zero level after about 15-17 years of Ivermectin treatment in two communities in Kaduna State, Nigeria. Similarly, Boatin and Richards (2006) reported that a tremendous success has been achieved in the control of onchocerciasis in the last 30 years both in Africa and South America. In spite of the successes recorded, the WHO (2015b) has described onchocerciasis as 'a country-wide public health problem in Nigeria'.

2.1.4 Mechanism of action of Ivermectin

Ivermectin is a macrocyclic lactone medicine used for the treatment of parasitic worms and ectoparasites, which can be taken orally or through a percutaneous endoscopic gastronomy, that is, (PEG) tube (Smith, 2004). It acts by binding to the microfilaria glutamate-gated chloride channels (GluCls) thereby disrupting the processes of neurotransmission controlled by the activity of the GluCls (Moreno *et al.*, 2010).

2.2 Vaccines

2.2.1 Definition of vaccine

A vaccine is a preparation that has been made non-pathogenic, that is given to elicit immune response and for disease prevention. A vaccine is not supposed to induce autoimmunity (Sridhar, 2006).

2.2.2 Types of vaccines

1) Killed vaccines:these vaccines which are also called inactivated vaccines are made when it will be unsafe to employ live pathogens to prepare vaccines. Hence, they are made non-pathogenic (unable to replicate) by treatment with formaldehyde, heat or gamma radiation. These kind of vaccines are safe and can be administered to immunodeficient people. An example includes rabies virus vaccine (Sridhar, 2006).

2) Live attenuated vaccines:these vaccines are made of live microorganisms that have been weakened by growing them in foreign hosts like animals, embryonated eggs or tissue culture using conditions that render them less virulent. Hence, they cause infection that is just enough to elicit the immune response, but not enough to cause disease. The polio vaccine is an example (Sridhar, 2006).

3) Conjugate vaccines:these vaccines are basically made to fight encapsulated bacteria (that are surrounded by polysaccharide capsule). The polysaccharides when coupled to proteins enable infants to elicit protective immune response (Sridhar, 2006). An example is the *Haemophilus influenzae* type b vaccine (Plotkin, 2014).

4) Anti-idiotypic vaccine: this is a vaccine made by using an antibody's paratope, a part of the antibody that binds to the epitope of the antigen and which reflects the three-dimensional structure of the epitope (idiotype). When the antibody is injected into another animal, antibodies are generated against the idiotype (anti-idiotype antibodies). Injecting the anti-idiotype antibody into a vaccine recipient generates antibodies (anti-anti-idiotype antibodies) (Sridhar, 2006).

5) DNA vaccines:in these kind of vaccines, genes for the target antigens are identified and cloned using genetic engineering. The DNA vaccine produced can be introduced into the vaccine recipient using a gene gun, DNA-coated gold particles and nose drops. Once inside the desired tissue, the DNA is expressed and both humoral and cellular immunity are induced (Sridhar, 2006).

6) Recombinant DNA vaccine: recombinant DNA technology or genetic engineering is employed to make these vaccines. Here, the gene of the antigen is inserted into a vector and the combination when injected into the vaccine recipient generates immune response. An example of this vaccine is the Hepatitis B virus (HBV) vaccine (Sridhar, 2006).

7) Subunit vaccines:this kind of vaccines are composed of purified antigens and not live or whole organisms. The active ingredients of these vaccines include toxoids, sub-cellular fragments or surface antigens. The effectiveness of these vaccines is increased when they are administered in adjuvants. When presented to the major histocompatibility complex (MHC) molecules on antigen-presenting cells, immunity is induced. The tetanus vaccine is an example of this kind of vaccine (Sridhar, 2006).

2.3 Reverse Vaccinology (RV)

2.3.1 Advantages of reverse vaccinology

Kanampalliwar *et al.,* (2013), define reverse vaccinology as the use of computer and genomic information to prepare vaccines without culturing microorganisms. According to Moriel *et al.,* (2013), reverse vaccinology is an *in silico* technique that has to do with screening DNA sequences using comparative analysis and prediction tools with the aim of finding suitable target antigens. Mining the genome sequence of a pathogen in order to identify genes that encode proteins with good vaccine attributes (section 2.4) is critical to reverse vaccinology (Seib *et al.,* 2009). According to Moriel *et al.,* (2013) good vaccine candidates have been identified to be cell surface localized proteins and secreted proteins. This is because they are more accessible to antibodies (Monterrubio-Lopez *et al.,* 2015). Bruno *et al.,* (2015), opined that the technique is the fastest and most comprehensive tool for antigen discovery. Other advantages enumerated by Bruno *et al.,* (2015) include: faster access to virtually every single antigen coded for by a genome responsible for disease, a "short cut" to vaccine discovery unlike the conventional approach which is time consuming, safer and more effective and suitable for all pathogens even non-cultivable organisms. Reverse vaccinology is not the same as vaccinomics. While the former starts with the genome of the pathogen, the later begins with the genome of the host of the pathogen, in order to develop a vaccine that best fits the host (Poland *et al.,* 2013).

2.3.2 Reported cases of reverse vaccinology

2.3.2.1 Anthrax

Anthrax is a disease that attacks animals and humans and it is caused by *Bacillus anthracis* (CDC, 2015). Kashikar and Dipke (2012) used the bioinformatics tool "EMBOSS" to screen the genome of the causative agent for antigenic determinants. Simulated docking experiments indicated the suitability of the epitope as excellent antigen. The kind of anthrax infection usually determines the symptoms noticed. Generally, symptoms include: body aches, fever, etc. (CDC, 2015).

2.3.2.2 Endocarditis

The causative agent of this disease is the gram positive *Streptococus mutans* (Streptococcus, 2012). It is characterized by inflammation or infection of the inner surface of the heart due to bacterial infection (Cabell *et al.,* 2003). Development of a vaccine for endocarditis using reverse vaccinology is in the pipeline (Kanampalliwar *et al.,* 2013). Any of the following symptoms may be exhibited: fatigue and weakness; loss of weight; fever and chills; night sweats; heart murmur; aches and pains; painful nodes in the pads of fingers and toes; red spots on skin of palms and soles; nail abnormalities; swelling of feet, legs, and abdomen; shortness of breath with activity; and blood in the urine (Millar and Moore, 2004).

2.3.2.3 Listerosis

The causative agent of this infectious food borne disease in animals and humans is the gram positive bacterium *Listeria monocytogenes*. Symptoms include fever, muscle aches, sometimes

preceded by diarrhea or other gastrointestinal symptoms (CDC, 2013b). Researchers are currently using the reverse vaccinology approach to develop a vaccine against listerosis (Kanampalliwar *et al.,* 2013).

2.3.2.4 Malaria

Malaria is transmitted by the bites of mosquitoes which inject the parasite *Plasmodium* into the blood stream and the symptoms are fever, headache, and vomiting, and usually appear between 10 and 15 days after the mosquito bite (WHO, 2015a). Malaria has persisted because of the resistance of the causative agents against various drugs (Kanampalliwar *et al.,* 2013). Shuaibu *et al.,* (2010) and other researchers have used the reverse vaccinology approach to search for a vaccine against malaria and WHO (2015a) has mentioned that currently more than 20 subunit vaccine preparations are being evaluated in clinical trials and/or are in advanced preclinical development.

2.3.2.5 Meningitis

The pathogen that causes meningitis is the *Neisseria meningitidis*. The disease is characterized by swelling of the membranes around the spinal cord and brain leading to fever, stiff neck and back, confusion and coma (CDC, 2015). Rappuoli and Sette (2010) first used reverse vaccinology to find a vaccine against meningitidis. They and others at The Institute for Genomic Research (TIGR) sequenced the total DNA of *Neisseria meningitidis* serogroup B, analyzed the sequences using bioinformatics tools and identified more than 600 potential antigens capable of eliciting immune response. In the past, the conventional approach to vaccine development was used, but with no success due to homology of the pathogen protein to humans and also as a result

of the hyper variable nature of the pathogen (Kanampalliwar *et al.*, 2013). Reverse vaccinology overcomes this homology problem by using BLAST to identify and trim off sequences of the pathogen that are homologous to humans.

2.3.2.6 Tuberculosis

Tuberculosis is a contagious and chronic infectious disease caused by *Mycobacterium tuberculosis* (US-NLM, 2012). The causative agent primarily attacks the lung, but can also attack the kidney, spine and brain. The symptoms of the disease include: weight loss, appetite loss, weakness/fatigue, chills, fever, sweating at night, coughing up blood or sputum, a bad cough that lasts three weeks or longer and pain in the chest (CDC, 2012). Reverse vaccinology is safer and has been used by Monterrubio-Lopez *et al.*, (2015) to identify six novel vaccine candidates: EsxL, PE26, PPE65, PE_PGRS49, PBP1 and Erp which can be used to design new TB vaccines.

2.3.2.7 Heartworm, *Dirofilaria immitis*

Godel *et al.*, (2012) employed *in silico* tools to screen the genome of *Dirofilaria immitis* and its endosymbiont, Wolbachia, thereby predicting potential drug and vaccine targets. The program In Paranoid was used to predict *Dirofilaria immitis* orthologues from *Brugia malayi, Caenorhabditis elegans, Homo sapiens* (humans) and *Canis lupus* (dogs) to identify drug targets. Proteins that had an orthologue in the dog or human proteome or had multiple paralogues in *Dirofilaria immitis* were excluded. Proteins that had a *C. elegans* orthologue essential for survival or development and had predicted function as an enzyme or receptor were included. Godel *et al.*, (2012) also surveyed the *D. immitis* genome for molecules currently proposed as vaccine candidates in other onchocercids and identified homologues.

From the foregoing it can be observed that more information is gained from the application of reverse vaccinology to vaccine development to prokaryotes than to Eukaryotes. This is because there are more works of literature on Prokaryotes than Eukaryotes with regards to reverse vaccinology. Goodswen *et al.,* (2013) confirm this in the following statement: 'Applying machine learning algorithms to classify eukaryotic proteins for vaccine discovery is novel and this is reflected by the presence of only a few publications on the topic.' This challenge may be as a result of what Vivona *et al.,* (2006) described as complexities of cellular and tissue organization in eukaryotic systems which tend to make reverse vaccinology less effective with eukaryotes.

2.4 Vaccine Prediction Parameters

2.4.1 Eliminating the risk of homology between pathogen and host proteins

The risk of autoimmunity must be avoided in order to ensure the safety of the recipients of a vaccine. Autoimmunity can be defined as an immune response targeted against a self-antigen (Agarwal, 2015). Lydyard *et al.,* (2004) explains that when an epitope of a pathogen is identical or nearly identical to the host tissue, the same mechanisms triggered to fight the pathogen will be generated against the host tissue. In developing vaccines using reverse vaccinology, sequences (DNA or protein) of pathogen homologous to host sequences, in this case human, are usually identified and subtracted to prevent the occurrence of autoimmunity (Vivona *et al.,* 2006, Seib *et al.,* 2009). This subtractionof undesired sequences in the computational identification of therapeutic targets is referred to as Subtractive Genomics Approach (Sarangi *et al.,* 2009).

2.4.2 Presence of secretory proteins

Any protein whether endocrine or exocrine secreted by a cell is a secretory protein (Wikipedia, 2016). Secretory proteins are accessible to antibodies. Thus are vaccine candidates (Pizza *et al.*, 2000). These proteins are usually directed to the secretory pathway (Freeman, 2000) and contain a signal sequence at their N-terminal (Barlowe and Miller, 2013). A typical signal sequence has a positively charged n-region, followed by a hydrophobic h-region and a neutral but polar c-region (*signal peptides*, 1997) with the cleavage site for signal peptidase (Kapp *et al.*, 2009). The programs SignalP (version 4.0) and TargetP, used in this study, can computationally predict the presence as well as the location of secretory signal peptides cleavage site at the C-terminal end of the presequence (Ellis, 2013) and the presence as well as length of secretory pathway signal peptides in the N-terminal presequence respectively (Emmanuelsson, 2000).

2.4.3 Presence of Transmembrane helices

Transmembrane proteins span across the cell membrane and are permanently attached to it. They have either alpha-helical regions or beta-barrel regions (membrane proteins, 2011) which are found only in some bacteria, mitochondria and chloroplasts (Randall and Baldi, 2010). The major category of transmembrane proteins are the alpha helices, they make up 20 – 30% of the total protein of any organism (Reeb *et al.*, 2014). Since transmembrane proteins of pathogens are found on the membrane, they are accessible to the immune system of their host and are thus antigenic (Ellis, 2013). Transmembrane helices in protein sequences can be predicted using the Program TMHMM which has a high level of accuracy (Krogh*et al.*, 2001; Moller *et al.*, 2001).

2.4.4 Presence of epitopes

The Encyclopedia Britannica (2014) defines epitope (antigenic determinant) as a part of an antigen that can stimulate an immune response. An antigen is a molecule foreign to the host cell that can be recognized by the immune system (Rao, 2001). The immune system's chief function is to eliminate infectious agents (Macpherson and Austyn, 2012). A network of lymphatic organs, tissues and cells make up the immune system (Anonymous, 2013). The Lymp Nodes, the Spleen, Thymus, Tonsils and other Lymphoid tissues spread across the entire body constitute the Lymphatic organs and tissues (*The Lymphatic System*, 2011). The key players of the immune system include B and T Lymphocytes (Kanampalliwar *et al.*, 2013) which are derived in the Bone marrow and Thymus respectively (LeBien and Tedder, 2008). The B cells recognize antigens and produce antibodies to fight them. The T cells also recognize antigens and in some cases generate second signals to activate B cells (Zouali, 2001). The B cells recognize antigens directly with the help of B cell receptors located on their surfaces, whereas antigens must be bound to a Major Histocompatibility Complex (MHC) before they are recognized by T cell receptors (Macpherson and Austyn, 2012). Major Histocompatibility Complex molecules are divided into MHC class I and MHC class II. The class I molecules are found on nucleated cells of the body and present antigens that emanate from within the cell to Cytotoxic T Lymphocytes (CTLs). The class II molecules are present mainly on dendritic cells, macrophages and B-cells – antigen presenting cells (APCs) of the Immune System and showcase exogenous antigens to Helper T Lymphocytes (Penn, 2001). MHC molecules in humans are referred to as Human Leukocyte Antigens (HLAs) and have different forms. They include, HLA class I which is divided into HLA-A, HLA-B and HLA-C and HLA class II also divided into HLA-DP, HLA-DQ

and HLA-DR (Lafuente and Reche, 2009). For T-cell epitopes to be predicted, investigation of the ability of antigens to bind to MHCs must be done (Lafuente and Reche, 2009, Patronov and Doytchinova, 2013). A leading resource for the prediction of antigen binding to MHC I and MHC II is the Immune Epitope Database Analysis Resource (IEDB) (Ellis, 2013). The BepiPred program can also be used to predict linear B-cell epitopes (Larsen *et al.,* 2006).

2.4.5 Homology of pathogen sequence(s) to vaccine candidate sequence(s) previously reported by other researchers.

Two or more sequences are said to share homology if they are descended from a common ancestral origin. Homology is a deduction made from sequence similarity alignment results. For nucleotide sequences, sequence similarity is the same as sequence identity (Xiong, 2006). However, for a protein sequence alignment, the two concepts are not the same: sequence identity means the percentage of the same amino acid residue matches between the aligned sequences. Whereas sequence similarity refers to the percentage of aligned residues that possess similar physicochemical properties like size, charge and hydrophobicity (Xiong, 2006).

Sequences of a pathogen that share significant similarity to identified toxins previously reported by other researchers may also be included in the vaccine candidate list (Gay *et al.,* 2007). Thesignificant similarity between sequencescan be determined using the Basic Local Alignment Search Tool (BLAST) which is very effective (Pearson, 2013). The tool makes use of the expectation value (E-value) as an indicator to ascertain if a sequence alignment is very significant or just as a result of mere chance (Xiong, 2006). The lower the E-value, the less the likelihood that the sequence match is a product of mere chance. Hence, the more significant the

match or alignment. For example, an E-value of <1e-50 (1 x $10^{\wedge -50}$) suggests very strongly that there is homology between the aligned sequences. Also, an E-value between 0.01 and 1e-50 suggests less strongly that there is homology. An alignment is considered not significant if the E-value is between 0.01 and 10, but may suggest a possible distant evolutionary relationship (Xiong, 2006).

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1 Materials

3.1.1 Equipment

The equipment used for this study include: a laptop with Linux operating system, 500Gb hard disk, a universal modem with Airtel network and a pair of dark glasses, worn to avoid eye strain.

3.1.2 Programs

The following were the programs used: SignalP (v4.1), TargetP (v1.1), TMHMM (v2.0), Phobius (v1.01), BepiPred (v1.0), MHC I binding predictor (v2.9.1) and MHC II binding predictor (v2.5.1).

3.1.3 Videos (materials consulted)

In the course of doing this work, the following YouTube videos were downloaded and watched: NCBI blast tutorial video, Immune Epitope Database (IEDB) User Workshop 2015-MHC class I binding prediction and Immune Epitope Database (IEDB) User Workshop 2015-MHC class II and Tepitool.

3.1.4 Websites

The following websites were visited:

www.ncbi.nlm.nih.gov, www.uniprot.org, www.expasy.org,
http://www.cbs.dtu.dk/services/SignalP, http://www.cbs.dtu.dk/services/TargetP,
http://www.cbs.dtu.dk/services/TMHMM, http://phobius.sbc.su.se,
http://tools.immuneepitope.org/mhci, http://tools.immuneepitope.org/mhcii
http://www.cbs.dtu.dk/services/Bepipred/

3.3 Methods

3.3.1 Retrieval of protein sequences

Protein sequences (both characterized and uncharacterized) of *O. volvulus* totaling 6259 were screened, and uncharacterized proteins in fasta format were retrieved by:

(a) logging on to www.uniprot.org

(b) opening 'proteomes' on another tab on the window.

(c) Typing '*Onchocerca volvulus'* in the search box and clicking search/enter.

(d) The number '12225' under protein count was clicked.

(e) Clicking 'Keywords.'

(f) Clicking, one at a time, the numbers assigned to 'PTM' (Post Translational Modification), 'Molecular function', 'Ligand', 'Cellular component' and 'Biological processes.'

(g) Selecting uncharacterized proteins and clicking 'Add to basket' to keep them in Uniprot basket.

(h) Clicking 'Basket' and then clicking 'Download.'

(i) A total of 4551 uncharacterized proteins were downloaded.

3.3.2 Screening of uncharacterized *O. volvulus* proteins.

In order to obtain protein sequences without significant similarity (non-orthologues) to human proteins, the following steps were followed:

(a) logging on to www.ncbi.nlm.nih.gov

(b) opening 'BLAST' in a new tab.

(c) opening 'protein blast' in a new tab.

(d) Clicking 'Browse' and uploading files, one at a time, of uncharacterized proteins from the computer download folder.

(e) Selecting 'Reference proteins (ref seq_protein)' as database.

(f) Typing 'human' in the organism field, 'human (taxid:9606)' immediately appears and is selected.

(g) Clicking 'BLAST.'

(h) Selecting sequences not similar to human proteins (from the download file obtained in 3.3.1 above) and saving them as a separate file.

3.3.3 Prediction of signal peptide cleavage site at C-terminal of protein sequences.

To accomplish this, the 'SignalP' program by Petersen *et al.,* (2011) was used as follows:

(a) logging on to www.cbs.dtu.dk/services/signalp/

(b) Clicking 'Browse' and selecting files, one at a time, of uncharacterized proteins not similar to humans and clicking 'open.'

(c) Selecting 'Short (no graphics)' as output format and 'Eukaryotes' as organism group.

(d) Clicking 'Submit.'

3.3.4 prediction of signal peptides at N-terminal of protein sequences.

To accomplish this, the 'TargetP' program by Emmauelsson *et al.,* (2000) was used as follows:

(a) logging on to www.cbs.dtu.dk/services/TargetP

(b) Clicking 'Browse', selecting files, one at a time, of uncharacterized proteins not similar to humans and clicking 'open' to upload.

(c) Selecting 'Non-plant' as organism group and selecting the 'perform cleavage site predictions box.'

(d) clicking 'Submit'

3.3.5 Prediction of transmembrane helices

The program 'TMHMM' by Krogh *et al.*, (2001) was used as follows:

(a) logging on to www.cbs.dtu.dk/services/TMHMM

(b) Clicking 'Browse', selecting a file (one at a time) of uncharacterized proteins of *Onchocerca volvulus* not similar to humans and clicking 'open' to upload.

(c) Selecting 'one line per protein.' as output format.

(d) clicking 'Submit Query'

The program 'Phobius' by Kall *et al.*, (2004) which discriminates between Transmembrane helices and Secretory peptides (Ellis, 2013) was then used to add endorsement to the predictions as follows:

(a) logging on to http://phobius.sbc.su.se/

(b) pasting the protein sequence in the field provided in FASTA format

(c) selecting 'Long with Graphics' output format

(d) clicking 'Submit Query' and waiting for the results to be displayed.

3.3.6 Prediction of T cell epitopes

This was done in two phases using the 'IEDB Analysis Resource' as follows:

I. Predicting MHC I binding, which was done by:

a) logging on to http://tools.immuneepitope.org/mhci

b) copying and pasting each of the twenty-two uncharacterized *Onchocerca volvulus* protein sequences found to be dissimilar to the human proteome and which contain a signal peptide and a single transmembrane helice in the field provided.

c) Selecting 'IEDB recommended' as the prediction method.

d) Selecting 'human' as the MHC source species.

e) Selecting 'HLA-A*01:01' and 'All lengths' as MHC allele and length respectively.

f) Clicking 'submit,' waiting for the results to be displayed and selecting peptides with the lowest percentile rank.

II. Predicting MHC II binding, which was done by:

a) logging on to http://tools.immuneepitope.org/mhcii

b) copying and pasting each of the twenty-two uncharacterized *Onchocerca volvulus* protein sequences in the field provided.

c) Selecting 'IEDB recommended' as prediction method.

d) Selecting 'Human, HLA-DR' as 'Species/locus'

e) selecting 'DRB1*01:01' as allele

f) clicking the 'submit' button, waiting for the results to be displayed and selecting only a peptide(s) with the lowest percentile rank.

3.3.7 Prediction of linear B cell epitopes

To accomplish this the 'BepiPred' programme was used as follows:

a) logging on to http://www.cbs.dtu.dk/services/bepipred

b) clicking 'Browse,' selecting file on the computer containing all the twenty-two uncharacterized protein sequences of *Onchocerca volvulus* found to be dissimilar to human and which contain a signal peptide and a single transmembrane helice and clicking 'open' to upload it.

c) Clicking 'submit,' waiting for the results to be displayed, saving results in pdf and selecting amino acid predicted to be part of an epitope (residues with an 'E' annotation).

3.3.8 Determining homologous sequences between *O. volvulus* and established potential vaccine candidate of *Dirofilaria immitis.*

Fructosebisphosphate aldolase (FBA) of *Dirofilaria immitis* is an established potential vaccine candidate (Godel *et al.*, 2012). Blast search (tblastn) was carried out to find out if this established vaccine candidate is homologous to any protein(s) in the *Onchocerca volvulus* nucleotide database (Gay *et al.*, 2007). This was done via the following steps:

(a) logging on to www.ncbi.nlm.nih.gov

(b) opening 'protein' in a new tab.

(c) Typing 'fructose bisphosphate aldolase *Dirofilaria immitis'* in the search box and clicking 'search'.

(d) Clicking 'Run BLAST' and then 'tblastn'.

(e) Typing 'Onchocerca volvulus' in the organism field. Before typing all characters, 'Onchocerca volvulus (taxid:6282)' is automatically suggested and is selected.

(f) Clicking 'BLAST' and saving results in pdf format.

3.3.9 Determining homologous sequences between *O. volvulus* and established potential vaccine candidate of *Brugia malayi.*

Chitinase is an established potential vaccine candidate of *Brugia malayi* (Godel *et al.*, 2012). To find homology, blast search (tblastn) was carried out (Gay *et al.*, 2007) by:

(a) logging on to www.ncbi.nlm.nih.gov

(b) opening 'protein' in a new tab.

(c) Typing 'Chitinase *Brugia malayi'* in the search box and clicking 'search'.

(d) Clicking 'Run BLAST' and then 'tblastn'

(e) Typing '*Onchocerca volvulus'* in the organism field. Before typing all characters in the name, 'Onchocerca volvulus (taxid:6282)' is automatically suggested and is selected.

(f) Clicking 'BLAST' and saving results in pdf format.

3.3.10 Prediction of possible function of some uncharacterized proteins (putative vaccine candidates).

To accomplish this, the following steps were followed:

a) logging on to 'www.ncbi.nih.nlm.gov

b) opening 'Blast' in a new tab.

c) Opening 'tblastn' in a new tab.

d) Clicking 'Browse' and uploading the file of the 22 uncharacterized putative vaccine candidate sequences from the computer.

e) Selecting 'Nucleotide collection (nr/nt)' as database.

f) Typing 'Dirofilaria immitis' in the organism field.

g) Clicking 'BLAST.'

h) repeating the above steps for *Brugia malayi*.

CHAPTER FOUR

4.0 RESULTS

4.1 Uncharacterized Protein sequences of *O. volvulus* non-orthologues to human proteins

From the 4551 uncharacterized protein sequences of *O. volvulus* screened using the NCBI BLAST, 802 were found to be non-orthologues to human proteins and their sequence entriesare in Appendix I. They are not stated here for convenience since they would occupy a lot of space and make the reading of this work boring. These 802 sequences were further screened by the programs: SignalP, TargetP, TMHMM and Phobius. Details are stated below.

4.2 Predicted secretory signal peptides

From the 802 uncharacterized protein sequences of *O. volvulus* found to have no homology with humans, 413 sequences were predicted by TargetP to have signal peptides. Both TargetP and SignalP were used to predict cleavage sites of signal peptidase. The signal peptides predicted agree with the standard nature of signal peptides that has been described by (*signal peptides*, 1997, Kapp*et al.*, 2009). There was a slight variation in the cleavage sites predicted by TargetP and SignalP; the cleavage sites predicted for each protein sequence by SignalP was one amino acid higher than that of TargetP. For example, SignalP classified sequence A0A044RS55 (S/N 1) to have a cleavage site at position 26, whereas TargetP predicted 25 as the cleavage site. An exception is seen in S/N 5 and a few others (S/N 3, 15, 16, etc.) shown in Appendix II. SignalP and TargetP predicted 22 and 24 respectively as cleavage site for S/N 5 (Appendix II). Some signal peptides predicted by TargetP and SignalP were not endorsed by Phobius. On the prediction column n7-18c25/26 means that the signal peptide is at position 7-18 with the

cleavage site between amino acid 25 and 26. The length of the signal peptide core was calculated to be 8-14 residues long. These observations can be seen in Appendix II.

4.3 Predicted transmembrane helices in protein sequences

A hundred and fifty-one (151) protein sequences were predicted by both Phobius and TMHMM to have a single transmembrane helice. The protein topology, that is, the in/out orientation with respect to the membrane, of each of the sequences was also given. For example, sequence entry (S/N 1) in table 4.1 below was predicted to have a topology of o18-40i by TMHMM. This means that the loop of the helice starts on the outside (o) of the membrane away from the cytoplasm and ends on the inside (i) towards the cytoplasm with the helice at position 18-40 of the amino acid sequence. All the 151 transmembrane helices predicted by the TMHMM program are 18-23 residues long. Phobius shows that 17-34 is the residues length of transmembrane helices predicted.

Table 4.1. Topology of single transmembrane helice present in uncharacterized proteins of *O. volvulus* non-orthologues to humans

S/N	Sequence entry	TMHMM		Phobius	
		Transmembrane helice	Topology (AAs in TMH)	Transmembrane helice	Topology (AAs in TMH)
1	A0A044U6S7	Y	o18-40i (23)	Y	o20-43i (24)
2	A0A044U8H1	Y	o15-37i (23)	Y	o18-35i (18)
3	A0A044UL22	Y	o10-29i (20)	Y	o6-30i (25)
4	A0A044SLQ4	Y	i51-70o (20)	Y	i52-70o (19)
5	A0A044UJI2	Y	i78-100o (23)	Y	i78-95o (18)
6	A0A044UXB1	Y	o345-367i (23)	Y	o340-367i (28)
7	A0A044TMU6	Y	o81-103i (23)	Y	o82-103i (22)
8	A0A044TM82	Y	o182-204i (23)	Y	o182-204i (23)
9	A0A044SA24	Y	o24-46i (23)	Y	o24-46i (23)
10	A0A044V148	Y	o10-32i (23)	Y	o6-29i (24)
11	A0A044U0N1	Y	o10-29i (20)	Y	o6-29i (24)
12	A0A044REF6	Y	i7-29o (23)	Y	i7-33o (27)
13	A0A044TD61	Y	o34-56i (23)	Y	o30-56i (27)
14	A0A044UXB1	Y	o345-367i (23)	Y	o340-367i (28)
15	A0A044RP47	Y	i62-81o (20)	Y	o63-82i (20)
16	A0A044SI04	Y	o10-32i (23)	Y	o6-30i (25)
17	A0A044RWM1	Y	o20-42i (23)	Y	o12-36i (25)
18	A0A044V6E0	Y	o10-29i (20)	Y	o12-30i (19)
19	A0A044UAE9	Y	i13-35o (23)	Y	i12-34o (23)
20	A0A044U293	Y	o268-290i (23)	Y	o266-284i (19)
21	A0A044URQ3	Y	i7-24o (18)	Y	i7-23o (17)
22	A0A044V9Z2	Y	o165-187i (23)	Y	o165-186i (22)
23	A0A044QK85	Y	i5-27o (23)	Y	o6-24i (19)
24	A0A044RIX6	Y	o227-249i (23)	Y	o227-249i (23)

25	A0A044SF87	Y	o77-99i (23)	Y	o76-99i (24)
26	A0A044RK54	Y	i13-35o (23)	Y	i7-36o (30)
27	A0A044V6Q4	Y	i7-29o (23)	Y	o6-30i (25)
28	A0A044SVE5	Y	i61-83o (23)	Y	i59-79o (21)
29	A0A044UIH1	Y	i12-34o (23)	Y	o12-34i (23)
30	A0A044SG87	Y	o192-211i (20)	Y	i192-211o (20)
31	A0A044U4G7	Y	o20-42i (23)	Y	o22-42i (21)
32	A0A044RYR5	Y	o10-27i (18)	Y	o12-29i (18)
33	A0A044U8Q1	Y	i13-35o (23)	Y	i12-33o (22)
34	A0A044SJU9	Y	i7-29o (23)	Y	o6-28i (23)
35	A0A044TS24	Y	i12-34o (23)	Y	i12-31o (20)
36	A0A044SQ33	Y	o10-32i (23)	Y	o12-32i (21)
37	A0A044V7H5	Y	o221-243i (23)	Y	o220-243i (24)
38	A0A044TV73	Y	o10-32i (23)	Y	i9-33o (25)
39	A0A044UZN4	Y	o6-28i (23)	Y	o6-27i (22)
40	A0A044TUF9	Y	10-32i (23)	Y	o12-33i (22)
41	A0A044UBT4	Y	o5-27i (23)	Y	o6-25i (20)
42	A0A044UHX1	Y	i19-38o (20)	Y	i18-38o (21)
43	A0A044RXC4	Y	o15-37i (23)	Y	o12-35i (24)
44	A0A044TXV3	Y	i13-35o (23)	Y	o15-35i (21)
45	A0A044R0T9	Y	i7-26o (20)	Y	o6-24i (19)
46	A0A044SFP8	Y	i13-35o (23)	Y	o13-39i (17)
47	A0A044U3W2	Y	o10-32i (23)	Y	o12-34i (23)
48	A0A044UE02	Y	i47-69o (23)	Y	o47-76i (30)
49	A0A044V3R2	Y	i56-78o (23)	Y	i56-80o (25)
50	A0A044TUY7	Y	i41-63o (23)	Y	i35-63o (29)
51	A0A044VEP2	Y	i28-45o (18)	Y	o29-49i (21)
52	A0A044UGZ1	Y	o230-252i (23)	Y	o231-252i (22)

53	A0A044SAS9	Y	i41-63o (23)	Y	o37-70i (34)
54	A0A044UBG0	Y	i13-32o (20)	Y	i12-30o (19)
55	A0A044SCX7	Y	o192-214i (23)	Y	o195-215i (21)
56	A0A044UWG2	Y	o72-94i (23)	Y	o73-98i (26)
57	A0A044T6M2	Y	i7-29o (23)	Y	o6-25i (20)
58	A0A044R5Q7	Y	o513-535i (23)	Y	o513-535i (23)
59	A0A044T2I2	Y	o652-674i (23)	Y	o650-674i (25)
60	A0A044SRL3	Y	o5-27i (23)	Y	o6-30i (25)
61	A0A044TDH5	Y	i27-49o (23)	Y	i27-47o (21)
62	A0A044V9S5	Y	o10-29i (20)	Y	o12-29i (18)
63	A0A044S8R4	Y	i48-70o (23)	Y	o47-70i (24)
64	A0A044S9L1	Y	i7-29o (23)	Y	o12-36i (25)
65	A0A044V7J3	Y	o15-37i (23)	Y	o23-40i (18)
66	A0A044RH86	Y	o5-27i (23)	Y	o6-27i (22)
67	A0A044UED5	Y	i2-24o (23)	Y	o6-25i (20)
68	A0A044URH4	Y	i19-41o (23)	Y	i20-41o (22)
69	A0A044RPP5	Y	o4-26i (23)	Y	o6-28i (23)
70	A0A044TW79	Y	o15-37i (23)	Y	o15-40i (26)
71	A0A044RGI8	Y	o13-35i (23)	Y	o14-33i (20)
72	A0A044RWL2	Y	i12-29o (18)	Y	i12-29o (18)
73	A0A044UZI4	Y	o23-42i (20)	Y	o23-42i (20)
74	A0A044RXY6	Y	i7-24o (18)	Y	o6-24i (19)
75	A0A044V4F5	Y	o38-60i (23)	Y	o36-60i (25)
76	A0A044VEF0	Y	o10-29i (20)	Y	o12-29i (18)
77	A0A044TH05	Y	i17-39o (23)	Y	o17-37i (21)
78	A0A044V368	Y	o10-32i (23)	Y	o12-31i (20)
79	A0A044QX16	Y	o5-27i (23)	Y	o6-27i (22)
80	A0A044U6E4	Y	i62-84o (23)	Y	o60-84i (25)

81	A0A044UG76	Y	o156-178i (23)	Y	o158-178i (21)
82	A0A044UF95	Y	o15-37i (23)	Y	o15-37i (23)
83	A0A044VAP0	Y	o224-246i (23)	Y	o223-246i (24)
84	A0A044VHR0	Y	o457-479i (23)	Y	o457-479i (23)
85	A0A044TJ65	Y	o942-961i (20)	Y	o942-961i (20)
86	A0A044TBY3	Y	i100-122o (23)	Y	o103-122i (20)
87	A0A044UNM4	Y	o112-134i (23)	Y	o113-134i (22)
88	A0A044RG47	Y	o144-166i (23)	Y	o141-166i (26)
89	A0A044SL30	Y	o15-37i (23)	Y	o20-40i (21)
90	A0A044RCT5	Y	o10-32i (23)	Y	o12-32i (21)
91	A0A044UFD5	Y	o10-29i (20)	Y	o12-29i (18)
92	A0A044SF54	Y	i7-29o (23)	Y	o6-29i (24)
93	A0A044V9K7	Y	o15-37i (23)	Y	o16-41i (26)
94	A0A044RAF2	Y	i5-27o (23)	Y	i5-27o (23)
95	A0A044UFR5	Y	o15-37i (23)	Y	o14-38i (25)
96	A0A044TR52	Y	o10-29i (20)	Y	o12-30i (19)
97	A0A044RA41	Y	o15-37i (23)	Y	o20-44i (25)
98	A0A044RAX8	Y	o10-32i (23)	Y	o14-33i (20)
99	A0A044TIW5	Y	o5-27i (23)	Y	o6-29i (24)
100	A0A044V134	Y	i33-52o (20)	Y	i21-46o (26)
101	A0A044UAX4	Y	o4-26i (23)	Y	o6-25i (20)
102	A0A044SJT9	Y	i7-25o (19)	Y	o6-25i (20)
103	A0A044ST73	Y	o15-37i (23)	Y	o12-36i (25)
104	A0A044RSX7	Y	o5-27i (23)	Y	o6-27i (22)
105	A0A044V598	Y	o20-38i (19)	Y	o20-38i (19)
106	A0A044U434	Y	o37-59i (23)	Y	o40-59i (20)
107	A0A044QU07	Y	o5-24i (20)	Y	o6-24i (19)
108	A0A044V4V1	Y	i7-29o (23)	Y	i7-30o (24)

109	A0A044V8Z0	Y	o10-27i (18)	Y	i7-25o (19)
110	A0A044RMF0	Y	o517-539i (23)	Y	o520-539i (20)
111	A0A044S787	Y	o169-191i (23)	Y	i169-191o (23)
112	A0A044SUI8	Y	o5-24i (20)	Y	o6-24i (19)
113	A0A044T088	Y	o4-21i (18)	Y	o6-25i (20)
114	A0A044TDN4	Y	o96-118i (23)	Y	o96-118i (23)
115	A0A044V9Y1	Y	o10-32i (23)	Y	o6-30i (25)
116	A0A044S7B9	Y	o24-46i (23)	Y	o28-47i (20)
117	A0A044U7F2	Y	o20-42i (23)	Y	o27-43i (17)
118	A0A044V4A7	Y	o15-35i (21)	Y	o18-35i (18)
119	A0A044S696	Y	i30-52o (23)	Y	i30-53o (24)
120	A0A044V7P6	Y	o15-37i (23)	Y	o18-37i (20)
121	A0A044RSM7	Y	o36-58i (23)	Y	o31-55i (25)
122	A0A044TSZ1	Y	o15-34i (20)	Y	o12-33i (22)
123	A0A044U1F8	Y	o55-77i (23)	Y	o51-77i (27)
124	A0A044VAN4	Y	i169-191o (23)	Y	o169-197i (29)
125	A0A044R301	Y	o20-42i (23)	Y	o20-42i (23)
126	A0A044R021	Y	o4-21i (18)	Y	o6-23i (18)
127	A0A044V6D0	Y	o10-29i (20)	Y	o6-30i (25)
128	A0A044VHZ6	Y	o10-32i (23)	Y	o12-30i (19)
129	A0A044UXT5	Y	o13-30i (18)	Y	o12-30i (19)
130	A0A044QW19	Y	o5-27i (23)	Y	o6-27i (22)
131	A0A044V136	Y	i7-29o (23)	Y	o6-30i (25)
132	A0A044V162	Y	o208-230i (23)	Y	o206-229i (24)
133	A0A044S1I6	Y	o15-37i (23)	Y	o6-31i (26)
134	A0A044RMM8	Y	o352-374i (23)	Y	o352-374i (23)
135	A0A044T009	Y	o371-393i (23)	Y	o369-394i (26)
136	A0A044QSB0	Y	o367-389i (23)	Y	o369-389i (21)

137	A0A044UB98	Y	i284-303o (20	Y	o284-310i (27)
138	A0A044SEB3	Y	o612-634i (23)	Y	o611-634i (24)
139	A0A044R521	Y	o363-385i (23)	Y	o363-385i (23)
140	A0A044TJF8	Y	o184-206i (23)	Y	o183-206i (24)
141	A0A044QSP0	Y	o217-239i (23)	Y	o217-240i (24)
142	A0A044RW31	Y	o5-27i (23)	Y	o6-27i (22)
143	A0A044QVL5	Y	i13-32o (20)	Y	i12-32o (21)
144	A0A044TUS5	Y	o10-29i (20)	Y	o6-29i (24)
145	A0A044U4D4	Y	o219-241i (23)	Y	o216-241i (26)
146	A0A044V9B1	Y	o35-57i (23)	Y	o39-57i (23)
147	A0A044T7U3	Y	o10-32i (23)	Y	o6-34i (29)
148	A0A044SRJ6	Y	o10-32i (23)	Y	i12-31o (20)
149	A0A044REF1	Y	i7-29o (23)	Y	i7-33o (27)
150	A0A044V915	Y	i23-41o (19)	Y	i23-41o (19)
151	A0A044VE43	Y	i177-199o (23)	Y	i174-199o (26)

'Y' = single transmembrane helice predicted,TMHMM= transmembrane hidden markov model, TMH = transmembrane helice and AAs = amino acids.

4.4 Predicted T cell epitopes

There was consensus among the programs: SignalP, TargetP, TMHMM and Phobius that only 22 sequences have both signal peptides and a single transmembrane helice. The IEDB MHC I binding prediction tool was used to screen these 22 sequences. Consequently, a total of 29 peptides with the lowest percentile rank that can bind to the allele HLA-A*01:01 were selected and are shown in table 4.2 below. Sequence entries having serial numbers 22, 15 and 11 have three peptides each that would theoretically bind to the allele. S/N 1 has two peptides that would bind to the allele and the remaining 18 sequence entries have one peptide each that would bind to

the allele. The lengths of the peptides are 9-14 residues long. The lowest and highest percentile rank are 0.1 and 0.7 respectively as shown in table 4.4.1. Many of the peptides have Tyrosine and other aromatic amino acid residues.

Table 4.2 Peptides of *O. volvulus* predicted to bind to MHC IHLA-A*01:01 allele

S/N	Sequence Entry	Peptide	Length	Start-End	Percentile rank
1	A0A044TMU6	MMCPQNSLTYYY	12	54-65	0.4
		FVSLFLFGY	9	8-16	0.4
2	A0A044TM82	ITDQRGYFF	9	64-72	0.4
3	A0A044U293	ETDTEEINSSHMY	13	58-70	0.1
4	A0A044V9Z2	NTTATPEKEY	10	147-156	0.5
5	A0A044RIX6	CSAINPHGY	9	91-99	0.25
6	A0A044SF87	ETEHLNGLSLPNEY	14	122-135	0.2
7	A0A044SCX7	STNDHRLYY	9	122-130	0.2
8	A0A044UWG2	PASSIFHYY	9	48-56	0.45
9	A0A044R5Q7	STDDFILPQLTY	12	213-224	0.1
10	A0A044VAP0	RTSTPSLTPIAY	12	324-335	0.2
11	A0A044VHR0	VIDENGCTLDSY	12	343-354	0.2
		LSSKHADFNANHEY	14	158-171	0.2
		NSDATMHDY	9	140-148	0.2
12	A0A044TBY3	CSKESDSCY	9	43-51	0.7
13	A0A044UNM4	FSSSSPFQCY	10	15-24	0.2
14	A0A044RG47	NTSAIHEGIY	10	104-113	0.3
15	A0A044RMF0	VSSDFSLFLY	10	428-437	0.2
		VMDDVSSDFSLFLY	14	424-437	0.2
		SSDFSLFLY	9	429-437	0.2

16	A0A044TDN4	LTDQIAEKAKQTVV	14	38-51	0.4
17	A0A044V162	LTTMCILQWIFNRY	14	219-232	0.2
18	A0A044RMM8	PTDLMAGQEAHVY	13	223-235	0.1
19	A0A044QSB0	PTDLMAGQEAHVY	13	230-242	0.1
20	A0A044SEB3	SSSPAASTVIEMLY	14	396-409	0.2
21	A0A044R521	LSDLMAGQEAHVY	13	233-245	0.1
22	A0A044TJF8	CTSWVQPQIGIY	12	21-32	0.2
		TLDIRMTKTDRY	12	58-69	0.2
		KTDRYDYLLQFCTY	14	65-78	0.2

The IEDB MHC II binding prediction tool was also used to screen the 22 sequences predicted by the programs SignalP, TargetP, TMHMM and Phobius to have signal peptides and a single transmembrane helice. Consequently, a total of 40 peptides with the lowest percentile rank that can bind to the allele HLA-DRB1*01:01 were selected and are shown in table 4.3 below. S/Ns 12 and 17 have five peptides each that can bind to the allele. S/N 18 has four peptides, S/Ns 16 and 10 have three each. S/Ns 1, 5 and 20 have two peptides each and the remaining 14 sequences have one peptide each that can bind to the allele. The length of each of the peptides was found to be 15. The lowest and highest percentile rank observed for the peptides predicted are 0.03 and 5.88 respectively. Table 4.3 also shows that many of the peptides have Tyrosine and other aromatic amino acid residues.

Table 4.3 Peptides of *O. volvulus* predicted to bind to MHC II HLA-DRB1*01:01 allele

S/N	Sequence Entry	Peptide	Length	Start-End	Percentile rank
1	A0A044TMU6	LFLFGYLNVNSAQAM	15	11-25	3.95
		FLFGYLNVNSAQAMN	15	12-26	3.95
2	A0A044TM82	LWWFLPNAANALFAL	15	8-22	0.09
3	A0A044U293	LLFILLFCNMLLARL	15	7-21	0.77
4	A0A044V9Z2	TILIIIVASAIVLYT	15	173-187	0.60
5	A0A044RIX6	VLTSFVTLVTLIIMR	15	236-250	0.96
		LTSFVTLVTLIIMRF	15	237-251	0.96
6	A0A044SF87	SAILFTALIVTFTLL	15	75-89	3.24
7	A0A044SCX7	LAYFMIILLIVLHLL	15	200-214	2.51
8	A0A044UWG2	TWVIVVLAVIAVMML	15	72-86	2.05
9	A0A044R5Q7	LLSMWILSGLSLVSI	15	512-526	2.91
10	A0A044VAP0	MKILISLAIIIAARL	15	1-15	0.77
		KILISLAIIIAARLH	15	2-16	0.77
		LISLAIIIAARLHSI	15	4-18	0.77
11	A0A044VHR0	QSPFKGIHAMLVRNC	15	317-331	2.18
12	A0A044TBY3	MIILLTFFLLHASLL	15	1-15	0.04
		IILLTFFLLHASLLW	15	2-16	0.04
		ILLTFFLLHASLLWN	15	3-17	0.04
		LLTFFLLHASLLWNV	15	4-18	0.04
		LTFFLLHASLLWNVV	15	5-19	0.04
13	A0A044UNM4	ITNIHLFISNLFITF	15	114-128	5.88
14	A0A044RG47	QFFASILSSLFIAVL	15	140-154	0.77

15	A0A044RMF0	KARLIVIASARDILV	15	252-266	0.62
16	A0A044TDN4	ILILILLKMLWSGLV	15	101-115	2.05
		LILILLKMLWSGLVW	15	102-116	2.05
		LILLKMLWSGLVWCI	15	104-118	2.05
17	A0A044V162	TIVSLLFVIWSPAIK	15	4-18	5.46
		IVSLLFVIWSPAIKT	15	5-19	5.46
		VSLLFVIWSPAIKTA	15	6-20	5.46
		SLLFVIWSPAIKTAQ	15	7-21	5.46
		LLFVIWSPAIKTAQV	15	8-22	5.46
18	A0A044RMM8	FGFSMFMGLSIALIA	15	352-366	0.42
		GFSMFMGLSIALIAA	15	353-367	0.42
		FSMFMGLSIALIAAV	15	354-368	0.42
		SMFMGLSIALIAAVI	15	355-369	0.42
19	A0A044QSB0	CFSNGTFAILIAMFG	15	363-377	2.51
20	A0A044SEB3	VLGFFMLLLGGAIGV	15	613-627	0.03
		LGFFMLLLGGAIGVG	15	614-628	0.03
21	A0A044R521	ILLIILFGAMLAIGS	15	363-377	0.28
22	A0A044TJF8	IWFLIFLAFILLLCL	15	185-199	1.15

4.5 Predicted linear B-cell epitopes

The total of 69 linear B-cell epitopes predicted from the 22 sequences earlier screened for MHC I and MHC II peptide binding are presented in Table 4.4. Their lengths are from 1 – 13 amino acid residues long, but epitopes with length less than two were not considered. The epitopes contain a few Tyrosine (Y) residues and many Arginine (R), Glycine (G) and Serine (S) residues.

Table 4.4 Linear B-cell epitopes of *Onchocerca volvulus*

S/N	Sequence entry	Epitope (score 1.30)
1	A0A044TMU6	-
2	A0A044TM82	NEGD, SSPDAR, GSGY, DNGPNPDD, TD
3	A0A044U293	QPSSWPNDSTTDH, DSSESETNPE, SDQTN
4	A0A044V9Z2	PG, GNESEKK, NTTATPEKEYS
5	A0A044RIX6	-
6	A0A044SF87	DPL
7	A0A044SCX7	NP, RI
8	A0A044UWG2	-
9	A0A044R5Q7	DSPNG, QD, NIP, EFQTSPTTSTS, PDV, SGE, SPKTSET, STR, TV
10	A0A044VAP0	EYESSGD, KNAN, NFD
11	A0A044VHR0	FTP, QSM, RK, RR
12	A0A044TBY3	-
13	A0A044UNM4	TCKG
14	A0A044RG47	SEVPEST
15	A0A044RMF0	-
16	A0A044TDN4	-
17	A0A044V162	EP

18	A0A044RMM8	GVEGEPEIE, RNDEG, GGPTGQ, NG, SE, CSEP, GAATR
19	A0A044QSB0	GVEGDPEI, CRSDEG, GPTGEPV, NG, GECARPECPEPQ, QPVQQPVE, DDNH
20	A0A044SEB3	TT, ESGK, VK, NR, ETS
21	A0A044R521	GVEGEPEIE, CRSDSG, GPTGT, DGKG, CPRPQCPE, PQ, AQ
22	A0A044TJF8	NGA, IRPPP, SDSPSESAH, PRT, EKA

4.6 Orthologue of *Dirofilaria immitis* potential vaccine candidate

The BLAST alignment result showed that Fructose-1,6-bisphosphate aldolase (fba) mRNA of *O. volvulus* shares a sequence identity of 97% and sequence similarity of 98% with the query, established potential vaccine candidate- Fructose bisphosphate aldolase of *Dirofilaria immitis*.

The result also gave an E value of 0.0 as can be seen in figure 2 below. However, it was dropped as a vaccine candidate because of its homology to humans.

Onchocerca volvulus fructose 1,6 bisphosphate aldolase (Fba) mRNA, complete cds

Sequence ID: **AF155220.1** Length: 1172 Number of Matches: 1

Range 1: 14 to 1102

Score	Expect	Method	Identities	Positives	Gaps	Frame
735 bits(1898)	0.0()	Compositional matrix adjust.	351/363(97%)	357/363(98%)	0/363(0%)	+2

Features:

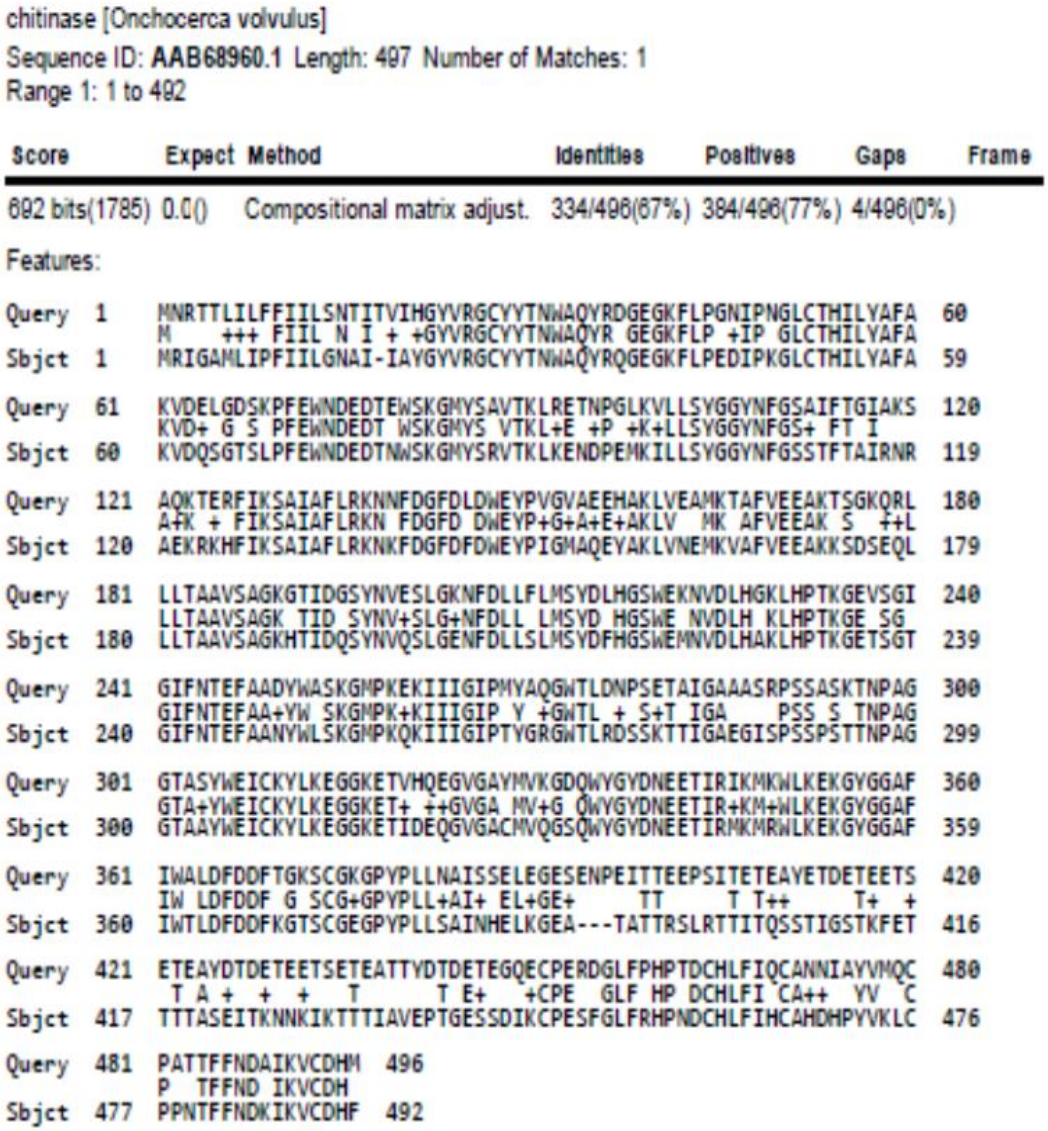
```
Query  1     MTSYSQFLTEAQKDELRQIANQIVAPGKGILAADESVGSMDKKLKPIGLENVEENRRLYR
             MTSYSQFLT+AQKDELR+IANQIV PGKGILAADES GSMDKKLKPIGLENVEENRRLYR
Sbjct  14    MTSYSQFLTDAQKDELRKIANQIVTPGKGILAADESTGSMDKKLKPIGLENVEENRRLYR

Query  61    QLLFTAGDEMSKYISGVIMFHESFYHKADDGTPFVQILQKKGILPGIKVDKGVVPMAGTV
             QLLFTAGDEMSKYISGVIMFHE+FY K DDGTPFVQILQKKGILPGIKVDKGV+PMAGTV
Sbjct  194   QLLFTAGDEMSKYISGVIMFHETFYQKGDDGTPFVQILQKKGILPGIKVDKGVIPMAGTV

Query  121   GEGTTQGLDDLNSRCAQYKKGGAQFAKWRCVHKIGATTPSHMALVEIAEVLARYASICQQ
             GEGTTQGLDDLNSRCAQYKK GAQFAKWRCVHKIGATTPSHMALVEIAEVLARYASICQQ
Sbjct  374   GEGTTQGLDDLNSRCAQYKKDGAQFAKWRCVHKIGATTPSHMALVEIAEVLARYASICQQ

Query  181   HGLVPIVEPETLPDGEHDVHRCQKVTELVLSYTYKALIDHHVYLEGTLLKPNMCMPGMQF
             HGLVPIVEPETLPDGEHDVHRCQKVTELVLSYTYKALIDHHVYLEGTLLKPNMCMPGMQF
Sbjct  554   HGLVPIVEPETLPDGEHDVHRCQKVTELVLSYTYKALIDHHVYLEGTLLKPNMCMPGMQF

Query  241   KGQCSHEEIARATVTALRRTVPVAVPGIVFLSGGQSEEDATLNLNAINQFSGKKPWALTF
             KGQCSHEEIARATVTAL+RTVPVAVPGIVFLSGGQSEEDATLNLNAINQF GKKPWALTF
Sbjct  734   KGQCSHEEIARATVTALQRTVPVAVPGIVFLSGGQSEEDATLNLNAINQFPGKKPWALTF

Query  301   SYGRALQASALAAWGGKPENIQAAKEAFLKRAEANSLAQLGKYTGGAGSGAAGENLYIAN
             SYGRALQASALAAWGGKPENIQAAKEAFLKRA+ANSLAQLGKYTGGAGSGAAGENLYIAN
Sbjct  914   SYGRALQASALAAWGGKPENIQAAKEAFLKRAQANSLAQLGKYTGGAGSGAAGENLYIAN

Query  361   HAY  363
             HAY
Sbjct  1094  HAY  1102
```

Figure 2. National Centre for Biotechnology Information (NCBI) BLAST output showing alignment of query (established potential vaccine candidate- Fructose bisphosphate aldolase of *Dirofilaria immitis*) with subject (Fructose -1,6- bisphosphate aldolase of *Onchocerca volvulus*) and their alignment details.

4.7 Orthologue of *Brugia malayi* potential vaccine candidate

The Blast alignment result showed that chitinase of *O. volvulus* shared a sequence identity of 87% and a sequence similarity of 77% with the *Brugia malayi* query.

Also, the alignment had an E value of 0.0 as can be seen in figure 3 below.



chitinase [Onchocerca volvulus]

Sequence ID: **AAB68960.1** Length: 497 Number of Matches: 1

Range 1: 1 to 492

Score	Expect	Method	Identities	Positives	Gaps	Frame
692 bits(1785)	0.0()	Compositional matrix adjust.	334/496(67%)	384/496(77%)	4/496(0%)	

Features:

```
Query  1    MNRTTLILFFIILSNTITVIHGYVRGCYYTNWAQYRDGEGKFLPGNIPNGLCTHILYAFA  60
            M    +++ FIIL N I + +GYVRGCYYTNWAQYR GEGKFLP +IP GLCTHILYAFA
Sbjct  1    MRIGAMLIPFIILGNAI-IAYGYVRGCYYTNWAQYRQGEGKFLPEDIPKGLCTHILYAFA  59

Query  61   KVDELGDSKPFEWNDEDTEWSKGMYSAVTKLRETNPGLKVLLSYGGYNFGSAIFTGIAKS  120
            KVD+ G S PFEWNDEDT WSKGMYS VTKL+E +P +K+LLSYGGYNFGS+ FT I
Sbjct  60   KVDQSGTSLPFEWNDEDTNWSKGMYSRVTKLKENDPEMKILLSYGGYNFGSSTFTAIRNR  119

Query  121  AQKTERFIKSAIAFLRKNNFDGFDLDWEYPVGVAEEHAKLVEAMKTAFVEEAKTSGKQRL  180
            A+K + FIKSAIAFLRKN FDGFD DWEYP+G+A+E+AKLV  MK AFVEEAK S  ++L
Sbjct  120  AEKRKHFIKSAIAFLRKNKFDGFDFDWEYPIGMAQEYAKLVNEMKVAFVEEAKKSDSEQL  179

Query  181  LLTAAVSAGKGTIDGSYNVESLGKNFDLLFLMSYDLHGSWEKNVDLHGKLHPTKGEVSGI  240
            LLTAAVSAGK TID SYNV+SLG+NFDLL LMSYD HGSWE NVDLH KLHPTKGE SG
Sbjct  180  LLTAAVSAGKHTIDQSYNVQSLGENFDLLSLMSYDFHGSWEMNVDLHAKLHPTKGETSGT  239

Query  241  GIFNTEFAADYWASKGMPKEKIIIGIPMYAQGWTLDNPSETAIGAAASRPSSASKTNPAG  300
            GIFNTEFAA+YW SKGMPK+KIIIGIP Y +GWTL + S+T IGA    PSS S TNPAG
Sbjct  240  GIFNTEFAANYWLSKGMPKQKIIIGIPTYGRGWTLRDSSKTTIGAEGISPSSPSTTNPAG  299

Query  301  GTASYWEICKYLKEGGKETVHQEGVGAYMVKGDQWYGYDNEETIRIKMKWLKEKGYGGAF  360
            GTA+YWEICKYLKEGGKET+ ++GVGA MV+G QWYGYDNEETIR+KM+WLKEKGYGGAF
Sbjct  300  GTAAYWEICKYLKEGGKETIDEQGVGACMVQGSQWYGYDNEETIRMKMRWLKEKGYGGAF  359

Query  361  IWALDFDDFTGKSCGKGPYPLLNAISSELEGESENPEITTEEPSITETEAYETDETEETS  420
            IW LDFDDF G SCG+GPYPLL+AI+ EL+GE+     TT     T T++     T+  +
Sbjct  360  IWTLDFDDFKGTSCGEGPYPLLSAINHELKGEA---TATTRSLRTTITQSSTIGSTKFET  416

Query  421  ETEAYDTDETEETSETEATTYDTDETEGQECPERDGLFPHPTDCHLFIQCANNIAYVMQC  480
             T A +  +  +   T      T E+   +CPE  GLF HP DCHLFI CA++  YV  C
Sbjct  417  TTTASEITKNNKIKTTTIAVEPTGESSDIKCPESFGLFRHPNDCHLFIHCAHDHPYVKLC  476

Query  481  PATTFFNDAIKVCDHM  496
            P  TFFND IKVCDH
Sbjct  477  PPNTFFNDKIKVCDHF  492
```

Figure 3. NCBI BLAST output showing alignment of query (established potential vaccine candidate- Chitinase of *Brugia malayi*) with subject (Chitinase of *O. volvulus*) and their alignment details.

4.8 Suggested function of some putative vaccine candidates

The total of eight (8) of the 22 predicted vaccine candidates (uncharacterized proteins) which are orthologues to *D. immitis* and/or *B. malayi* and are likely to have the following functions are shown in table 4.5 below.

Table 4.5 Possible function of some of the 22 putative vaccine candidates.

***Onchocerca volvulus* protein**	**Predicted function/description**	***Dirofilaria immitis* orthologue**			***Brugia malayi* orthologue**		
		Acession no	E value	% identity	Acession number	E value	% identity
A0A044RMM8	Impervious cuticlin protein	AF453385.1	0.0	62	XM 001898524.1	0.0	90
A0A044R5Q7	Impervious cuticlin protein	-	-	-	XM 001902142.1	1.E-65	88
A0A044QSB0	Impervious cuticlin protein	-	-	-	XM 001895330.1	0.0	73
A0A044R521	Impervious cuticlin protein	-	-	-	XM 001902757.1	0.0	77
A0A044VHR0	Zona pellucida-like domain containing protein	-	-	-	XM 001900794.1	0.0	62
A0A044RMF0	Hypothetical protein	-	-	-	XM 001896012.1	0.0	60
A0A044SEB3	Hypothetical protein	-	-	-	XM 001897289.1	0.0	68
A0A044TJF8	Hypothetical protein	-	-	-	XM 001899259.1	1E-138	55

Only tblastn E -value of 0.0 or E < 1e-50 was considered. '-' = no orthologue. *O. volvulus* proteins are referred to by their Uniprot sequence entry. *Dirofilaria immitis* and *Brugia malayi* orthologues are referred to by their accession number in National center for Biotechnology information (NCBI). Expectation value (E value) and % identity are NCBI blast output parameters.

CHAPTER FIVE

5.0 DISCUSSION

5.1 Protein sequences of *O. volvulus* non orthologues to human proteins.

The lack of similarity observed between protein sequences of *O. volvulus* and protein sequences of human can be attributed to lack of aligned amino acid residues. Thus, it can be inferred that these 802 sequences of *O. volvulus* (Appendix I) do not have a homologous relationship or share homology with humans. This agrees with the position of Xiong (2006). There would not likely be risk of autoimmunity if vaccine candidates are drawn from them (Vivona *et al.,* 2006, Seib *et al.,* 2009).

5.2 Predicted secretory signal peptides

The slight variation that was observed in the cleavage sites predicted by the programs TargetP and SignalP may be attributed to what Ellis (2013) pointed out, that TargetP detects signal peptides at the N-terminal presequence. Whereas, SignalP detects cleavage sites of signal peptides at the C-terminal end of the presequence.

Some signal peptides predicted by TargetP and SignalP were not endorsed by Phobius, a combined signal peptide and transmembrane helice predictor, probably, because of the reason already stated by Krogh *et al.,* (2001), Ellis (2013) and Reeb *et al.,* (2014): signal peptides and transmembrane helices both contain a hydrophobic region that can be easily misclassified by a prediction program. Only non-orthologues to human *O. volvulus* sequencesunanimously

predicted by theprograms: TargetP, SignalP, TMHMM and Phobius to contain signal peptides and a single transmembrane helicewere used for epitope prediction.

The pattern of the signal peptides observed agrees with that reported by Signal peptides (1997) and Kapp *et al.,* (2009). The length of the signal peptide core calculated does not conflict with the known length of signal peptides already stated by Kall *et al.*, (2004).

The validity of these signal peptides predicted has been attested to by their pattern and length which are typical of signal peptides. This indicates that they are likely to be immunogenic (Ellis, 2013).

5.3 Predicted transmembrane helices in protein sequences

The topology of proteins containing transmembrane helice observed are correct since they agree with the description of transmembrane helice given by Ellis (2013). There is therefore a high likelihood that they can trigger immune response (Krogh *et al.*, 2001).

Some transmembrane helice predicted by TMHMM were not endorsed by Phobius, probably, because of the reason already stated by Krogh *et al.,* (2001), Ellis (2013) and Reeb *et al.*, (2014): that the hydrophobic region present in both signal peptides and transmembrane helice can confuse a program, making prediction difficult.

The position of Krogh *et al.,* (2001) supports the transmembrane helice residues length predicted in this study by both TMHMM and Phobius.

5.4 Predicted T-cell epitopes

The length (9 – 14) of the 29 peptides that can bind to MHC I predicted did not agree with that (8 – 11) stated by Patronov and Doytchinova (2012). However, the length (15) of the 40 peptides predicted, which can bind to MHC II agreed with the length (12 – 25) noted by the afore-mentioned authors.

The very low percentile ranks of these predicted peptides indicate that they have a high likelihood of triggering immune response and might therefore be potential vaccine candidates (Pandey *et al.,* 2018). Also, the likelihood that the epitopes predicted would elicit immune response is high because they have many Tyrosine and other aromatic acid residues that have been shown to be immunogenic (Wehland *et al.,* 1984, Steen 2010, Peng *et al.,* 2014).

These peptides can trigger immune response from T lymphocytes when they are displayed by MHC I and MHC II (Macpherson and Austyn, 2012) and would enhance humoral and cell-mediated immunity (Pandey *et al.,* 2018).

5.5 Predicted linear B-cell epitopes

The 69 linear B-cell epitopes predicted do not need MHCs; they would be recognized directly with the aid of B-cell antibody receptors located on their surfaces (Macpherson and Austyn, 2012). The prospects for these predicted linear B-cell epitopes are high. Su *et al.,* (2012) have observed that in wet-lab experiments contiguous peptide sequences (from which linear B-cell epitopes are constructed) are synthesized without difficulties.

The predicted epitopes contain Tyrosine, Arginine, Glycine and Serine which have been shown to be very immunogenic portions of antigens (Wehland *et al.,* 1984, Steen 2010, Peng *et al.,* 2014).

In addition, they would enhance humoral immunity in recipients of the vaccine produced based on them (Pandey *et al.,* 2018).

5.6 Orthologue of *Dirofilaria immitis* potential vaccine candidate

Sequence identity of 97% shows that there is a 97% match of the same amino acid residues (Xiong *et al.,* 2006) between Fructose 1,6-bisphosphate aldolase of *D. immitis* and that of *O. volvulus*. This indicates homology between the two sequences.

Sequence similarity (positives) of 98% shows that the percentage of aligned residues that have similar physicochemical properties and can be readily substituted for each other is 98% (Xiong *et al.,* 2006). This is also an indication of homology.

An E value of 0.0, i.e., <1e-179 (NCBI, 2011) observed indicates that the likelihood that the match is as a result of chance is less than 1e-179 (very low); the match is very significant and hence, homology is inferred (Xiong *et al.,* 2006).

However, using the NCBI Blastp programme and selecting 'Reference proteins' (refseq_protein) as database name, the Blast alignment output of this potential candidate was found to have very low E values and shared an identity of more than 60% with at least four human proteins (fructose-1,6-bisphosphate aldolase A1, A2, B and C). Therefore, it was dropped in order to avoid the occurrence of autoimmunity. The findings of this research contradict the position of

McCarthy *et al.,* (2002) that fructose-1,6-bisphosphate aldolase of *O. volvulus* can serve as a potential vaccine candidate.

5.7 Orthologue of *Brugia malayi* potential vaccine candidate

Sequence identity of 87% shows that there is an 87% match of the same amino acid residues (Xiong *et al.,* 2006) between Chitinase of *Brugia malayi* and that of *O. volvulus*. This indicates homology between the two sequences.

Sequence similarity (positives) of 77% shows that the percentage of aligned residues that have similar physicochemical properties and can be readily substituted for each other is 77% (Xiong *et al.,* 2006). This is also an indication of homology.

An E value of 0.0, i.e., <1e-179 (NCBI, 2011) observed shows that the likelihood that the match is as a result of chance is less than 1e-179 (very low); the match is very significant and hence, homology is inferred (Xiong *et al.,* 2006).

The position of Gay *et al.,* (2007) that if an antigen has a similar structure to a known toxin it can be included in the vaccine candidate list justifies our consideration of chitinase as a putative vaccine candidate.

Interestingly, our inclusion of chitinase in the vaccine candidate list coincides with an earlier position of Harrison *et al.,* (1999) that chitinase of *O. volvulus* has some vaccine potential.

5.8 Suggested function of some putative vaccine candidates

The eight (8) uncharacterized protein sequences, which are part of the putative vaccine candidates predicted and whose possible functions have been predicted in table 4.5 above are

mostly orthologues of *B. malayi*. The sequence, A0A044RMM8, was found to be an orthologue of both *B. malayi* and *D. immitis*. They are considered orthologues because they are homologous protein sequences found in different organisms: *B. malayi*, *D. immitis* and *O. volvulus* and probably have the same function in these different organisms (Xiong 2006).

As there are no animal models for *O. volvulus*, chitinase and the other uncharacterized sequences, which have been found to be orthologues with *B. malayi* and *D. immitis*potential vaccine candidate sequences, seem to have provided an escape route. These potential vaccine candidates of *O. volvulus*would be used and tested on animal models for *B. malayi* and/or *D. immitis*.

CHAPTER SIX

6.0 CONCLUSION AND RECOMMENDATION

6.1 Conclusion

Using Subtractive Genomic Approach, 802 uncharacterized protein sequences of *O. volvulus* found to be non-orthologues to humans were screened successfully for signal peptides and transmembrane helice via Bioinformatics programs. Consequently, 22 putative vaccine candidates were revealed from which many T-cell and B-cell epitopes were predicted. Using comparative genomics (sequence similarity search approach), one more putative vaccine candidate similar to already established potential vaccine candidate of *B. malayi* was also identified. The findings of this study showed that fructose-1,6-bisphosphate aldolase of *O. volvulus* reported to be a potential vaccine candidate by other researchers cannot be considered so, as it is very similar to the fructose-1,6-bisphosphate aldolase of humans, posing a risk of autoimmunity should a vaccine be made based on it. Furthermore, the problem of lack of animal models for *O. volvulus* seems to have been circumvented; the putative vaccine candidates of *O. volvulus* found to be orthologues with *B. malayi* and *D. immitis* would be tested on animal models for the duo. Also, the possible function of some of the uncharacterized proteins were suggested. All these might be promising candidates for vaccine design against onchocerciasis.

6.2 Recommendation

Experimental validation of the suggested function of some of the 22 uncharacterized proteins proposed in this study as well as the use of the putative vaccine candidates predicted is required.

REFERENCES

Agarwal, S. K. (2015). Autoimmunity and Autoimmune Diseases [PowerPoint slides]. Retrieved from www.uth.tmc.edu/pathology/medic/immunology/immuno/Autoimmunity Agarwal2013.pdf

Albers, A., Esum, M. E., Tendongfor, N., Enyong, P., Klarmann, U., Wanji, S., Hoerauf, A. and Pfarr, K. (2012) Retarded *Onchocerca volvulus* L1 to L3 larval development in the *Simulium damnosum* vector after anti-wolbachial treatment of the human host. *Parasites and Vectors*, 5 (12). doi:10.1186/1756-3305-5-12.

Anonymous (2013). Human Physiology/ The Immune System. Retrieved from www.saylor.org/site/wp-content/uploads/2010/11/The-Immune-System.pdf

Barlowe, C. K. and Miller, E. A. (2013). Secretory protein biogenesis and traffic in the early secretory pathway. *Genetics*, 193, 383 – 4110. doi: 10.1534/genetics.112.142810

Basanez, M., Pion, S. D. S., Churcher, T. S., Breitling, L. P., Little, M. P. and Boussinesq, M. (2006). River blindness: A success story under threat? *PLoS Med*, 3(9), 1 – 7. http://dx.doi.org/10.1371/journal.pmed.0030371

Bertholet, S., Reed, S. G. and Rappuoli, R. (2015). Reverse Vaccinology, the Art and Science of Tuberculosis Vaccine Development, 2nd edition. Retrieved from: http://tbvaccines.usm.my/finlay/sites/default/files/Chapter%203.1.pdf

Boatin, B.A., and Richards, F.O. (2006). Control of Onchocerciasis. *Advances in Parasitology*, 61, 349-394. doi: 10.1016/S0065-308X(05)61009-3

Bruno, L., Cortese, M., Rappuoli, R. and Merola M. (2015). Lessons from Reverse Vaccinology for Viral Vaccine Design. *Current Opinion in Virology*, 11, 89-97. doi:10.1016/j.coviro.2015.03.001

Cabell, C.H., Abrutyn, E. and Karchmer, A.W. (2003). Bacterial endocarditis: the disease, treatment and prevention. *Circulation*, 1-4. doi: 10.1161/01.CIR.0000071082.36561.F1

Cafardi, V., John, L. T. and Serruto, D. (2013). Bacterial Genomes and Vaccine Design. *Immunomics Reviews*, 5, 13 – 37. DOI 10.1007/978-1-4614-5070-2_2.

Carter centre (2015). Eliminating riverblindness. Retrieved October 8, 2015, from https://www.cartercenter.org/countries/nigeria.html

Center for Disease, Control and Prevention (2016). Onchocerciasis. Retrieved December 30, 2017, from: https://www.cdc.gov/dpdx/onchocerciasis/index.html

Center for Global Development (2007). Case 7: Controlling Onchocerciasis in Sub-Saharan Africa. Retrieved October 8, 2015, from http://www.cgdev.org/page/case-7-controlling-onchocerciasis-sub-saharan-africa

Centers for Disease Control and Development (2012). Tuberculosis (TB). Retrieved October 16, 2015, from http://www.cdc.gov/tb/topic/basics/default.htm

Centers for Disease Control and Prevention (2013a). Epidemiology and risks factors. Retrieved October 8, 2015, from http://www.cdc.gov/parasites/onchocerciasis/epi.html

Centers for Disease Control and Prevention (2013b). Listeria (Listeriosis). Retrieved October 15, 2015, from http://www.cdc.gov/listeria/definition.html

Centers for Disease Control and Prevention (2015a). Anthrax. Retrieved October 14, 2015, from http://www.cdc.gov/anthrax/index.html

Centers for Disease Control and prevention (2015b). Meningococcal Disease. Retrieved October 15, 2015, from http://www.cdc.gov/meningococcal/about/causes-transmission.html

Centers for Disease Control and Prevention-CDC (2013c). Onchocerciasis. Retrieved June 26, 2015, from: http://www.cdc.gov/parasites/onchocerciasis

Cho-Ngwa, F., Liu, J. and Lustigman S. (2010). The *Onchocerca volvulus* cysteine proteinase inhibitor, Ov-CPI-2, is a target of protective antibody response that increases with age. *PLoS Neglected Tropical Disease*, 4(8), doi: 10.1371/journal.pntd.0000800.

Edungbola, L.D. (1991). Onchocerciasis Control in Nigeria. *Parasitology Today*, 7(5), 97 – 99. Retrieved from http://www.sciencedirect.com/science/article/pii/016947589190165K/part/first-page-pdf

Ellis, J. (2013). User guide for Vacceed version 1.1: An in silico vaccine discovery pipeline for eukaryotic pathogens. Retrieved from https://github.com/sgoodswe/vacceed/blob/v1.1/Vacceed_User_Guide.pdf

Emanuelsson, O., Nielsen, H., Brunak, S. and Heijne, V. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol*. 300, 1005 – 1016. doi: 10.1006/jmbi.2000.3903.

Epitope (2016). Encyclopaedia Britannica. Retrieved May 29, 2016, from www.britannica.com/science/epitope

Etya'ale, D. (2001). Vision 2020: Update on Onchocerciasis. *Community Eye Health*, 14(38), 19-21. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1705929/pdf/jceh_14_38_019.pdf

Freeman, W. H. (2000). *Overview of the secretory pathway*. Retrieved from www.ncbi.nlm.nih.gov/books/NBK21471/

Gay, C. J., Zuerner, R., Bannantine, J. P., Lillehoj, H. S., Zhu, J. J., Green, R. and Pastoret, P. P. (2007). Genomics and Vaccine Development. *Rev. sci. tech. Off. int. Epiz.*,26(1), 49-67

Globisch, D., Moreno, A. Y., Hixon, M. S., Nunes, A. A. K., Denery, J. R., Specht, S., Hoerauf, A., and Janda, K. D. (2013) *Onchocerca volvulus*-neurotransmitter tyramine is a biomarker for river blindness, Proceedings of the National Academy of Science USA, 110(11), pp 4218 – 4223.

Gloeckner, C., Garner, A. L., Mersha, F., Oksov, Y., Tricoche, N., Eubanks, L. M., … Janda, K. M. (2010)Repositioning of an existing drug for the neglected tropical disease Onchocerciasis, *PNAS*, 107 (8), 3424 – 3429. doi:10.1073/pnas.0915125107

Godel, C., Sujai, K., Georgios, K., Philip, L., Daniel, N., Francesco, C., … Pascal, M. (2012). The Genome of the heartworm, *Dirofilaria immitis*, reveals drug and vaccine targets. The *Journal of the Federation of American Societies for Experimental Biology*, 26(11), 4650 – 4661.

Goodsween, S. J., Kennedy, P. J. and Ellis, J. T. (2013). A novel strategy for classifying the output from an in silico vaccine discovery pipeline for eukaryotic pathogens using machine learning algorithms. *BMC Bioinformatics*, 14(315), 1 – 13. https://doi.org/10.1186/1471-2105-14-315

Jain, L., and Kumar, V., (2014). Recent Approaches in Vaccinology. Research News for U, 13, pp 136 – 142. Retrieved from http://doublehelixresearch.com/files/magazines/rnfu/vol-13/RECENT-APPROACHES-IN-VACCINOLOGY.pdf

Josefsberg, O. J. and Buckland, B. (2012). Vaccine process technology. *Biotechnology and Bioengineering Journal*, 109(6), doi 10.1002/bit.24493.

Kall, L., Krogh, A. and Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 338, 1027-1036. doi: 10.1016/j.jmb.2004.03.016

Kanampalliwar, A. M., Soni, R., Girdhar, A. and Tiwari, A. (2013) Reverse Vaccinology: Basics and Applications, *Journal of Vaccines and Vaccinology*, 4, 194. doi: 10.4172/2157-7560.1000194.

Kapp, K., Schrempf, S., Lemberg, M. K. and Dobberstein, B. (2009). Post-targeting functions of signal peptides. *Madame Curie Bioscience Database [internet]*. Retrieved from https://www. ncbi.nlm.nih.gov/books/NBK6322/

Kashikar, P. and Dipke, C. (2012). Insilico design and development of vaccine by reverse vaccinology approach for anthrax. *Journal of Advanced Bioinformatics Applications and Research*, 3(1), 262-266. Retrieved from http://www.bipublication.com

Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., ... Peters, B. (2012). Immune epitope database analysis resource. *Nucleic Acids Research*, 40, 1 – 6. DOI: 10.1093/nar/gks438

Kluxen, G. (2011). A Short History of Onchocerciasis. Retrieved October 7, 2015, from http://www.jean-hissette.de/Oncho_JH_GB.html

Krogh, A., Larsson, B., Heijne, G. V. and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol*. 305, 567 – 580. doi: 10.1006/jmbi.2000.4315

Lafuente, E. M. and Reche, P. A. (2009). Prediction of MHC-Peptide binding: A Systematic and comprehensive overview. *Current Pharmaceutical Design*. 15(28), 3209 – 3220. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.9790&rep=rep1&type=pdf

Larsen, J. E. P., Lund O. and Nielsen M. (2006). Improved method for predicting Linear B-cell epitopes. *Immunome Research*, 2(2), 1 – 7. doi:10.1186/1745-7580-2-2

LeBien, T. W. and Tedder, T. F. (2008). B Lymphocytes: how they develop and function. *Blood*. 112(5), 1570 – 1580. doi:10.1182/blood-2008-02-078071

Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., and Nielsen, M. (2008). NetMHC-3.0: Accurate web accessible predictions of Human, Mouse, and Monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Research*, 36, 509-512. doi: 10.1093/nar/gkn202

Lydyard, P. M., Whelan, A. and Fanger, M. W. (2004). *Instant Notes in Immunology*. Retrieved from www. dphu.org/uploads/attachments/books/books_5453_0.pdf

Mackenzie, C. D., Homeida, M. M., Hopkins, A. D. and Lawrence, J. C. (2012). Elimination of onchocerciasis from Africa: possible? *Trends in Parasitology*, 28(1), 16 – 22. doi: 10.1016/j.pt.2011.10.003.

Macpherson, G. and Austyn, J. (2012). The Immune System. In Exploring Immunology: Concepts and Evidence. (pp 5 – 47). Retrieved from http://www.wiley-vch.de/books/sample/3527324127_c01.pdf

Makepeace, B. L., Jensen, S. A., Laney, S. J., Nfon, C. K., Njongmeta, L. M., Tanya, V. N., Williams, S. A., Bianco, A. E. and Trees, A. J. (2009) Immunisation with a multivalent, subunit vaccine reduces patent infection in a natural bovine model of onchocerciasis

during intense field exposure. *PLoS Neglected Tropical Diseases*, 3(11), 1 – 7. doi:10.1371/journal.pntd.0000544.

McCarthy, J. S., Wieseman, M., Tropea, J., Kaslow, D., Abraham, D., Lustigman, S., … Nutman, T. B. (2002). Onchocerca volvulus glycolytic enzyme fructose-1,6-bisphoshate aldolase as a target for a protective immune response in humans. *Infection and Immunity*, 70(2), pp 851 – 858. doi: 10.1128/IAI.70.2.851–858.2002

Mectizan Donation Program (2015). Onchocerciasis Endemic Countries. Retrieved October 8, 2015, from http://www.mectizan.org/onchocerciasis-endemic-countries

Membrane proteins (2011). Retrieved May 22, 2016, from cavanaghlab.bch.ncsu.edu/bch701/Membrane_proteins.pdf

Meredith, S.E.O., Cross, C. and Amazigo, U.V. (2012). Empowering communities in combating riverblindness and the role of NGOs: case studies from Cameroon, Mali, Nigeria and Uganda. *Health Research Policy and Systems*, 10(16), 1-16. doi:10.1186/1478-4505-10-16

Millar, B. C., & Moore, J. E. (2004). Emerging Issues in Infective Endocarditis. *Emerging Infectious Diseases, 10*(6), 1110-1116. https://dx.doi.org/10.3201/eid1006.030848.

Moller, S., Croning, M. D. R. and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*. 17 (7), 646 – 653. doi: 10.1093/bioinformatics/17.7.646

Monterrubio-Lopez, G.P., Gonzalez-Y-Merchand, J.A. and Ribas-Aparicio, R.M. (2015). Identification of novel potential vaccine candidates against tuberculosis based on reverse vaccinology. *Biomed Research International*, 2015, 1-16. DOI: http://dx.doi.org/10.1155/2015/483150

Moreno, Y., Nabhan, J. F., Solomon, J., Mackenzie, C. D. and Geary, T. G. (2010). Ivermectin disrupts the function of the excretory-secretory apparatus in microfilariae of *Brugia malayi*. *PNAS*, pp 1 – 6. www.pnas.org/cgi/doi/10.1073/pnas.1011983107

NCBI. (2011, November 2). BLAST results: expect values, part 2[video file]. Video posted tohttp://www.youtube.com/watch?v=Z7ek7UoP7Bg

Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S. L., Lamberth, K., Buus, S., Brunak, S. and Lund, O. (2003). Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci*, 12(5), 1007-1017. DOI: 10.1110/ps.0239403

Okoro, N., Nwali, U. N., Nnamdi, O. A., Innocent, O. C., Somadina, O. C. and Shedrack, E. O. (2014) The Prevalence and Distribution of Human Onchocerciasis in two Senatorial

Districts in Ebonyi state, Nigeria. *American Journal of Infectious Diseases and Microbiology*, 2(2), 39 – 44. DOI:10.12691/ajidm-2-2-3

Pandey, R. K., Bhatt, T. K. and Prajapati, V. K. (2018). Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating Anopheles salivary protein. Scientific Reports. 8. DOI:10.1038/s41598-018-19456-1

Patronov, A. and Doytchinova, I. (2013). T-cell Epitope Vaccine Design by Immunoinformatics. *Open Biology*, 3, 1 – 13. doi: 10.1098/rsob.120139

Pearson, W. R. (2013). An introduction to sequence similarity ('homology') searching, *Current Protocols in Bioinformatics*, 3(10), pp 1-9.doi:10.1002/0471250953.bi0301s42

Peng, H. P., Lee, K. H., Jian, J. W., and Yang, A. S. (2014). Origins of specificity and affinity in antibody–protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), E2656–E2665. http://doi.org/10.1073/pnas.1401131111

Penn, D. J. and Ilmonen, P. (2002). Major Histocompatibility Complex (MHC). *John Wiley and Sons Encyclopaedia of Life Science*, 1 – 8. DOI:10.1038/NPG.ELS.0003986

Peters, B., and Sette, A. (2005). Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 6(132), 1 – 9. doi:10.1186/1471-2105-6-132

Petersen, T. N., Brunak, S., Heijne, G. V. and Nielssen, H. (2011). Nature methods. 8, pp 785 – 786. doi: 10.1038/nmeth.1701

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M. M., Arico, B., Comanducci, M., ... Rappuoli, R. (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Research Gate*, 287, 1816 – 1820. Retrieved from https://www.researchgate.net/profile/Beatrice_Arico/publication/12606081_Identification_of_vaccine_candidates_against_serogroup_B_meningococcus_by_whole-genome_sequencing/links/0fcfd50a267c2afe86000000.pdf/download?version=vs

Plotkin, S. (2014). History of vaccination. Retrieved from http://www.pnas.org/content/111/34/12283.full.pdf

Poland, G. A., Kennedy, R. B., McKinney, B. A., Ovsyannikova, I. G., Lambert, N. D., Jacobson, R. M. and Oberg, A. L. (2013).Vaccinomics, adversomics, and the immune response network theory: individualized vaccinology in the 21st century. *Seminars in immunology*, 25(2), 89 – 103. doi:10.1016/j.smim.2013.04.007

Randall, A. and Baldi, P. (2010). Transmembrane beta-barrel protein structure prediction. *Structural bioinformatics of membrane proteins*, 2010, 83 – 102. doi: 10.1007/978-3-7091-0045-5_5

Rao, K. V. S. (2001). Antigens. Encyclopedia of Life Sciences. Retrieved May 29, 2016, from http://mi.med.stu.edu.cn/pdf/%E6%8A%97%E5%8E%9F.pdf

Rappuoli, R. and Sette, A. (2010). Reverse vaccinology: Developing vaccines in the era of genomics. *Immunity*, 33(4): 530 – 534. doi:10.1016/j.immuni.2010.09.017

Reeb, J., Kloppmann, E., Bernhofer, M. and Rost, B. (2014). Evaluation of transmembrane helix predictions in 2014. *Proteins*. 83, 473 – 484. doi: 10.1002/prot.24749

Richards, F.O., Boatin, B., Sauerbrey, M. and Seketeli, A. (2001). Control of Onchocerciasis Today: Status and Challenges. *TRENDS in Parasitology*, 17(12), 558-563. doi: https://doi.org/10.1016/S1471-4922(01)02112-2

Rouchka, E. C. and Cha, I. E. (2009) Current Trends in Pseudogene Detection and Characterisation, *Journal of Current Bioinformatics*, 4(2), pp 112 – 119.

Sarangi, A. N., Aggarwal, R., Rahman, Q. and Trivedi, N. (2009). Subtractive Genomics Approach for *in Silico* Identification and Characterization of Novel Drug Targets in *Neisseria Meningitidis Serogroup B. Journal of Computer Science and Systems Biology* 2: 255-258. doi:10.4172/jcsb.1000038

Secretory Protein. (2016). Wikipedia, the free encyclopedia. Retrieved 16 May, 2016, from http://en.wikipedia.org/wiki/secretory_protein

Seib, K.L., Dougan, G. and Rappuoli, R. (2009). The key role of genomics in modern vaccine and drug design for emerging infectious diseases. *PLoS Genetics* 5(10), 1 – 8. doi:10.1371/journal.pgen.1000612.

Shapiro, L. (2012). *Onchocerca volvulus*. Retrieved June 26, 2015, from: http://www.eol.org/data_objects/17764063

Shuaibu, M. N., Kikuchi, M., Cherif, M. S., Helegbe, G. K., Yanagi, T. and Hirayama, K. (2010). Selection and identification of malaria vaccine target molecule using bioinformatics and DNA vaccination. *Elsevier*, 28(42), pp 6868 – 6875. https://doi.org/10.1016/j.vaccine.2010.08.019

Sidney, J., Assarsson, E., Moore, C., Ngo, S., Pinilla, C., Sette, A. and Peters, B. (2008). Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Research* 4(2), 1 – 14. doi:10.1186/1745-7580-4-2

Signal peptides (1997). Retrieved May 17, 2016, from www.cryst.bbk.ac.uk/pps97/assignments/projects/dulai/signal.html

Sir Emeka Offor Foundation (2014). The river blindness elimination commitment in Nigeria. Retrieved November 26, 2016, from http://www.sireofforfoundation.org/index.php/news-and-events/224-the-river-blindness-elimination-commitment-in-nigeria

Smith J. (2004). Ivermectin. Retrieved February 6, 2018, from: https://www.dermnetnz.org/topics/ivermectin/

Sridhar, R. P. N. (2006). Vaccines. Retrieved from https://www.microrao.com/micronotes/vaccines.pdf

Steen, J. (2010). Characterisation of antigenic properties and high throughput protein purification. Retrieved from https://www.diva-portal.org/smash/get/diva2:360129/FULLTEXT02.pdf

Streptococcus mutans. (2012). In Farlex partner medical dictionary. Retrieved from http://medical-dictionary.thefreedictionary.com/Streptococcus+mutans

Su, C. H., Pal, N. R., Lin, K. L. and Chung, I. F. (2012). Identification of Amino acids propensities that are strong determinants of linear B-cell epitope using Neural networks. PLOS ONE. 7(3). https://doi.org/10.1371/annotation/58d3a7cb-d039-40ca-b960-bd60abfe0ae3

The Lymphatic System and Lymphoid organs and tissues (2011). Retrieved from www.phschool.com/atschool/florida/pdfbooks/Sci_Marieb/pdf/Marieb_ch20.pdf

United States National Library of Medicine (2012). Tuberculosis (TB). Retrieved October 16, 2015, from http://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0024668/

Vivona, S., Bernante, F. and Filippini, F. (2006). NERVE: New Enhanced Reverse Vaccinology Environment. *BMC Biotechnology*, 6(35), 1 – 8. doi:10.1186/1472-6750-6-35

Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A. and Peters, B. (2008). A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Computational Biology*, 4(4), 1 – 10. doi: *10.1371/journal.pcbi.1000048*

Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M. and Peters, B. (2010). Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*, 11(568), 1 – 12. doi: 10.1186/1471-2105-11-568

Wehland, J., Schroder, H. C., and Weber, K. (1984). Amino acid sequence requirements in the epitope recognized by the α-tubulin-specific rat monoclonal antibody YL1/2. The EMBO

Journal, 3(6), 1295 – 1300. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1460-2075.1984.tb01965.x

World Health Organization (2015a). Malaria. Retrieved October 15, 2015, from http://www.who.int/malaria/areas/vaccine/en/

World Health Organization (2015b). Onchocerciasis. Retrieved June 26, 2015, from: http://www.who.int/apoc/onchocerciasis/en/

Xiong, J. (2006). Essential Bioinformatics. Retrieved from http://www.aun.edu.eg/molecular_biology/Procedure%20Bioinformatics22.23-4-2015/Xiong%20-%20Essential%20Bioinformatics%20send%20by%20Amira.pdf

Yan, R., Chen, Z., and Zhang, Z. (2011). Outer Membrane can be simply identified using secondary structure element alignment, *BMC Bioinformatics*, 12(76), 1 – 9. doi: 10.1186/1471-2105-12-76

Zouali, M. (2001). Antibodies. Encyclopedia of Life Science. Retrieved from http://www1.gdou.edu.cn/nxy/dwmyx/resources%5Cpdf%5C%E6%8A%97%E4%BD%93%EF%BC%88PDF%EF%BC%8C%E8%8B%B1%E6%96%87%EF%BC%89.pdf

APPENDICES

Appendix I:

1.0 Sequence entries of uncharacterized proteins of *Onchocerca volvulus* non-orthologues to humans used for screening

A0A044U8Q6, A0A044U5A6, A0A044UIG6, A0A044UMG0, A0A044V4M7, A0A044U1C0, A0A044U425, A0A044UKA7, A0A044RBY2, A0A044SS18, A0A044RS55, A0A044QKI3, A0A044UBU6, A0A044QQ29, A0A044V9E9, A0A044R3V3, A0A044R6T0, A0A044T6E6, A0A044T183, A0A044UMA2, A0A044TW81, A0A044QZA3, A0A044RB35, A0A044RQZ1, A0A044STA6, A0A044SJY0, A0A044V4B5, A0A044UDD9, A0A044TUS0, A0A044UQ33, A0A044U8D5, A0A044RM78, A0A044T5L2, A0A044RB19, A0A044V1B5, A0A044T3E2, A0A044U6S7, A0A044RGT4, A0A044SYW6, A0A044QUD5, A0A044QQ79, A0A044QSP5, A0A044V8J4, A0A044SCV7, A0A044TUX9, A0A044RCD9, A0A044RKA6, A0A044SB89, A0A044R8N6, A0A044R8T6, A0A044TEN5, A0A044RLS7, A0A044U8H1, A0A044UL22, A0A044UEA6, A0A044V110, A0A044RPR7, A0A044SYK0, A0A044U9G9, A0A044SAT5, A0A044QKL9, A0A044UQ24, A0A044R864, A0A044RHS4, A0A044SYA7, A0A044TT84, A0A044QJS4, A0A044URQ7, A0A044T9S9, A0A044QVN5 A0A044S3P5, A0A044V716, A0A044UL49, A0A044QVR1, A0A044U3P3, A0A044SW64, A0A044TJB3, A0A044R276, A0A044QZ68, A0A044RVC7, A0A044SLQ4, A0A044UJI2, A0A044QQ18, A0A044V7D3, A0A044S8K1, A0A044SD38, A0A044S0V5, A0A044SGM6, A0A044UEP2, A0A044T0B2, A0A044VD77, A0A044UT04, A0A044SXP0, A0A044V990, A0A044SA24, A0A044V148, A0A044S852, A0A044U5M2, A0A044T0Y2, A0A044SBA6, A0A044RJ57, A0A044QJT5, A0A044UB72, A0A044RXA1, A0A044V182, A0A044U0N1, A0A044U697, A0A044S450, A0A044S9P3, A0A044UAP7, A0A044QL01, A0A044RA12, A0A044V703, A0A044R4W1, A0A044SKA7, A0A044REF6, A0A044TIY8, A0A044T320, A0A044TK75, A0A044TD61, A0A044UT72, A0A044R7C4, A0A044S9R4, A0A044ULE8, A0A044UXB1, A0A044TKC3, A0A044TMU6, A0A044V9M2, A0A044RPX3, A0A044S4C5, A0A044SS76, A0A044TFH0, A0A044RY15, A0A044UHI2, A0A044SCF4, A0A044UJX8, A0A044V966, A0A044RP47, A0A044UE82, A0A044TZH7, A0A044SI04, A0A044RY69, A0A044SG43, A0A044VG57, A0A044RWM1, A0A044RPS7, A0A044SE40, A0A044TM82, A0A044V6E0, A0A044VCR9, A0A044SZ09, A0A044V3Y4, A0A044S121, A0A044UAE9, A0A044QLB9, A0A044RJP9, A0A044U293, A0A044UXF1, A0A044RSJ7, A0A044RK54, A0A044QSN9, A0A044V6Q4, A0A044SWW2, A0A044TAN2, A0A044V643, A0A044RM40, A0A044SPF1, A0A044RXZ7, A0A044SVE5, A0A044UIH1, A0A044URQ3, A0A044V9Z2, A0A044SG87, A0A044S9U7, A0A044U4G7, A0A044SFK0, A0A044RYR5, A0A044UU42, A0A044V2Y0, A0A044TYZ8, A0A044T3F0, A0A044V7S7, A0A044SPG6, A0A044R308, A0A044TZS1, A0A044V1B0, A0A044QLH0, A0A044T8I3, A0A044U8Q1, A0A044SJU9, A0A044RXR8, A0A044S3I9, A0A044UCC4, A0A044TS24, A0A044R2H0, A0A044TQM6, A0A044SLR1, A0A044SQ33, A0A044R3T1, A0A044R9E4, A0A044RMT2, A0A044V7H5, A0A044VJA5, A0A044RY06, A0A044RPG3, A0A044T377, A0A044R5Y9, A0A044SCH6, A0A044TV73, A0A044TP02, A0A044QK85, A0A044RIX6, A0A044RQP7, A0A044V478, A0A044SMJ4, A0A044UZN4, A0A044TUF9, A0A044V9E6, A0A044UBT4, A0A044UHX1, A0A044VAG2, A0A044R4D8, A0A044RXC4, A0A044TXV3, A0A044V7J6, A0A044R0T9, A0A044SM21, A0A044SAQ9, A0A044R4C5, A0A044R4I2, A0A044UCG6, A0A044RGE6, A0A044RFU5, A0A044S6A4, A0A044TH32, A0A044QSA2, A0A044RQ68, A0A044RTL0, A0A044SFP8, A0A044QXD7, A0A044T0Q3, A0A044UJ48, A0A044S5G7, A0A044T7K9, A0A044RDA7, A0A044QLG4, A0A044T2A6, A0A044TFG4, A0A044U3W2, A0A044R0F9, A0A044VDA4, A0A044R963, A0A044UE02, A0A044VF92, A0A044RBE0, A0A044TS47, A0A044ULG2, A0A044ST32, A0A044RVG8, A0A044T082, A0A044UT67, A0A044SF73, A0A044V3R2, A0A044TYE4, A0A044RG85, A0A044TUY7, A0A044VEP2, A0A044TME1, A0A044U536, A0A044UBK0, A0A044V8G4, A0A044R2H5, A0A044S5A9, A0A044UJM3, A0A044UGZ1, A0A044SF87, A0A044SAS9, A0A044T2I2, A0A044U4A3, A0A044VI80, A0A044QV79, A0A044RW03, A0A044SBX3, A0A044SI40, A0A044TCT8, A0A044SRL3, A0A044TDH5, A0A044R0H2, A0A044RM57, A0A044SQD0, A0A044V9G4, A0A044VAQ4, A0A044UPY3, A0A044QL28, A0A044RY03, A0A044U3Q0, A0A044TLY2, A0A044SS11, A0A044UBG0, A0A044QPL1, A0A044V0S5, A0A044R2V9, A0A044U006, A0A044VIV6, A0A044QKI7, A0A044V124, A0A044UM51, A0A044RER0,

A0A044R930, A0A044SB57, A0A044V9S5, A0A044UPZ0, A0A044U7Z9, A0A044UVJ4, A0A044QUE3, A0A044S8R4, A0A044UT03, A0A044SUT4, A0A044UF12, A0A044S9T2, A0A044TYI5, A0A044VCM0, A0A044V5X7, A0A044UJU0, A0A044RPX9, A0A044SSA4, A0A044TCD5, A0A044RFG0, A0A044QVS5, A0A044SCX7, A0A044QZH7, A0A044V1M0, A0A044S9L1, A0A044V7J3, A0A044QYA0, A0A044RTH0, A0A044RH86, A0A044RI52, A0A044TXX4, A0A044U6R2, A0A044UWG2, A0A044TTM3, A0A044UED5, A0A044URH4, A0A044TDE7, A0A044V954, A0A044RV21, A0A044RT58, A0A044R514, A0A044RPP5, A0A044VHX7, A0A044RIT5, A0A044U9P5, A0A044VCF0, A0A044RIL5, A0A044RP25, A0A044TW79, A0A044VF51, A0A044SLH9, A0A044V444, A0A044SKM0, A0A044SZ88, A0A044TJP8, A0A044V933, A0A044QLY0, A0A044RGI8, A0A044TSD8, A0A044RUQ4, A0A044VG01, A0A044U836, A0A044RWL2, A0A044SX98, A0A044ST38, A0A044U4F2, A0A044S1M8, A0A044T6M2, A0A044RFL6, A0A044SBY6, A0A044RS55, A0A044S6X4, A0A044V312, A0A044UXR5, A0A044SUF3, A0A044UZI4, A0A044TPD5, A0A044RDH6, A0A044UKS6, A0A044QTY1, A0A044TTI6, A0A044RU70, A0A044RXY6, A0A044V4F5, A0A044VEF0, A0A044TH05, A0A044TMY2, A0A044V368, A0A044RW45, A0A044QSL0, A0A044TY25, A0A044QX16, A0A044S836, A0A044SD15, A0A044T653, A0A044T0P0, A0A044RP87, A0A044RTE4, A0A044SAQ4, A0A044SJI2, A0A044U6E4, A0A044R5Q7, A0A044RGN5, A0A044QU01, A0A044SL30, A0A044RCT5, A0A044SL96, A0A044VGE2, A0A044V9E9, A0A044RXX0, A0A044UFD5, A0A044R6C9, A0A044VA40, A0A044S3E3, A0A044UCC9, A0A044T4L8, A0A044U2C2, A0A044UG76, A0A044RQS2, A0A044UF95, A0A044UJ83, A0A044UUJ2, A0A044VJG1, A0A044SYM5, A0A044U131, A0A044RAV7, A0A044RH48, A0A044V9F1, A0A044RW00, A0A044SF54, A0A044TQV4, A0A044VAP0, A0A044RL99, A0A044RPI9, A0A044TDR8, A0A044TCI1, A0A044RTU4, A0A044T9V8, A0A044V562, A0A044SJG6, A0A044T4F4, A0A044V9K7, A0A044TBG6, A0A044V6Z6, A0A044U8B5, A0A044RET9, A0A044SEG2, A0A044RAF2, A0A044UFR5, A0A044V8X7, A0A044QUR6, A0A044U2X6, A0A044U8E5, A0A044T513, A0A044RX04, A0A044TNW9, A0A044VCM8, A0A044RVB1, A0A044T6P1, A0A044UQ20, A0A044TR52, A0A044RA41, A0A044SHW3, A0A044TFJ2, A0A044UF77, A0A044S1E7, A0A044RAX8, A0A044VDK0, A0A044R983, A0A044S1P8, A0A044R8U7, A0A044VG24, A0A044VH58, A0A044S742, A0A044QXC6, A0A044VE78, A0A044RJU5, A0A044TEN0, A0A044QXM7, A0A044VGV8, A0A044RU62, A0A044UNI7, A0A044SI27, A0A044VHR0, A0A044R2V4, A0A044RVW1, A0A044VDB0, A0A044RGT0, A0A044TKF5, A0A044T546, A0A044UBX4, A0A044SVZ2, A0A044TIW5, A0A044U8M5, A0A044S299, A0A044SWJ6, A0A044S6G9, A0A044U019, A0A044QRK6, A0A044TJ65, A0A044SNK6, A0A044V134, A0A044TBY3, A0A044SVZ7, A0A044RWL0, A0A044UAX4, A0A044U109, A0A044RJ42, A0A044U1I6, A0A044V489, A0A044RC25, A0A044T7H0, A0A044UA29, A0A044TZL9, A0A044RP43, A0A044UNM4, A0A044V9J1, A0A044RG47, A0A044SJT9, A0A044RY83, A0A044TPK0, A0A044QLJ3, A0A044SEQ6, A0A044V2B6, A0A044RAZ9, A0A044RZI1, A0A044ST73, A0A044SCI7, A0A044RQ39, A0A044RSX7, A0A044RPG4, A0A044TJH6, A0A044QPE5, A0A044S0W6, A0A044UVK0, A0A044RX39, A0A044S9S6, A0A044RQ45, A0A044S9J1, A0A044V598, A0A044UM18, A0A044U434, A0A044QU07, A0A044U015, A0A044U9S0, A0A044UCF9, A0A044V4V1, A0A044VGD4, A0A044V9N8, A0A044UHW5, A0A044SQX9, A0A044RKR8, A0A044SLV5, A0A044UZ60, A0A044T0M0, A0A044VF63, A0A044T039, A0A044VCR1, A0A044S138, A0A044U5P1, A0A044TQ34, A0A044SUS3, A0A044U3Q1, A0A044SNZ2, A0A044V8Z0, A0A044TC69, A0A044QWG2, A0A044THL2, A0A044T3G3, A0A044V043, A0A044VA30, A0A044RWT9, A0A044S650, A0A044QXU9, A0A044SYV3, A0A044UI66, A0A044V445, A0A044UN34, A0A044RNV9, A0A044QK32, A0A044UA49, A0A044RZJ5, A0A044RIZ3, A0A044TZH9, A0A044R6S7, A0A044S7F4, A0A044RTY5, A0A044RWY4, A0A044TVC2, A0A044UMU2, A0A044QUK6, A0A044THG2, A0A044VIS7, A0A044V9Y1, A0A044R4S0, A0A044VG14, A0A044TBK5, A0A044UYY9, A0A044S7B9, A0A044RMF0, A0A044SWB1, A0A044TPU2, A0A044V770, A0A044QLH7, A0A044SNG8, A0A044U7F2, A0A044US99, A0A044VAC7, A0A044U243, A0A044VGU6, A0A044TB37, A0A044V4A7, A0A044RC50, A0A044S696, A0A044V7P6, A0A044U5Y5, A0A044UYJ7, A0A044RJZ0, A0A044RSM7, A0A044SR48, A0A044S787, A0A044SIP6, A0A044TSZ1, A0A044RMJ0, A0A044U1F8, A0A044R222, A0A044UM25, A0A044SUI8, A0A044RBE7, A0A044URL5, A0A044S9W3, A0A044V728, A0A044RM52, A0A044RJC4, A0A044RD51, A0A044S8N0, A0A044TP81, A0A044RJ67, A0A044TA93, A0A044TS84, A0A044RT52, A0A044UZZ0, A0A044VAN4, A0A044RQX9, A0A044SD63, A0A044R301, A0A044R021, A0A044SL10, A0A044T088, A0A044RNY2, A0A044QMC5, A0A044SLX7, A0A044U104, A0A044V6D0, A0A044VBE5, A0A044VF57, A0A044QSQ5, A0A044RW46, A0A044TW88, A0A044TCE6, A0A044QPP3, A0A044QVD2, A0A044SH33, A0A044SWS4, A0A044VHC2, A0A044UPQ4, A0A044SGV4, A0A044QKX5, A0A044V9X2, A0A044VHZ6, A0A044S0E3, A0A044RXY3, A0A044UL06, A0A044TW32, A0A044SCL2, A0A044TDN4, A0A044UXT5, A0A044S609,

A0A044T655, A0A044RTR8, A0A044RXR2, A0A044S669, A0A044QRJ8, A0A044U4D5, A0A044QW19, A0A044RIG3, A0A044V136, A0A044STL4, A0A044UBD0, A0A044S9V7, A0A044SID1, A0A044V8Y1, A0A044V162, A0A044V5K8, A0A044RTY8, A0A044TNX5, A0A044T3L2, A0A044UTQ0, A0A044QPW3, A0A044VHS6, A0A044RJ90, A0A044TJW6, A0A044R4Z0, A0A044RW31, A0A044S1I6, A0A044TAF1, A0A044U323, A0A044U3T2, A0A044RNY8, A0A044S5Z6, A0A044S8W7, A0A044TUH8, A0A044QVL5, A0A044TUS5, A0A044VG75, A0A044RP61, A0A044RMM8, A0A044TC63, A0A044S9K0, A0A044S426, A0A044U8A2, A0A044UPJ3, A0A044V9Y6, A0A044T009, A0A044UFM5, A0A044V4C2, A0A044UKR1, A0A044TT39, A0A044RUX0, A0A044T7K1, A0A044T4L7, A0A044R7M0, A0A044UMK0, A0A044SYN7 A0A044TZJ6, A0A044V5S3, A0A044R4D0, A0A044R5D8, A0A044TYY0, A0A044QQQ4, A0A044VHK0, A0A044TF01, A0A044SHG1, A0A044UCE0, A0A044QSB0, A0A044UB98, A0A044SEB3, A0A044R521, A0A044U4D4, A0A044VCB0, A0A044UVD8, A0A044VF19, A0A044TJF8, A0A044V9N9, A0A044TRG3, A0A044S8I7, A0A044SQ43, A0A044SKR1, A0A044VEJ7, A0A044R0R1, A0A044T193, A0A044V6Y3, A0A044URP2, A0A044R6Z6, A0A044RVU1, A0A044V4Q9, A0A044RA05, A0A044V9B1, A0A044QZD1, A0A044T7U3, A0A044RVD3, A0A044RM49, A0A044QQD0, A0A044TKN4, A0A044V132, A0A044SBT5, A0A044T0J7, A0A044SQV9, A0A044S1T8, A0A044U337, A0A044SMS5, A0A044SRJ6, A0A044QSP0, A0A044U4V5, A0A044REF1, A0A044V915, A0A044UGA0, A0A044RGR7, A0A044VDG5, A0A044VE43, A0A044SUU0, A0A044U5U6, A0A044U9I9 and A0A044SNM2.

Appendix II

2.0 Signal peptides and cleavage sites predicted in uncharacterized proteins of *O. volvulus*non-orthologues to humans

S/N	Sequence Entry	SignalP program		TargetP program		Phobius program		
		Signal peptide	Cleavage site	Signal peptide	Cleavage site	Signal peptide	prediction	Length of signal peptide core
1	A0A044RS55	Y	26	Y	25	Y	n7-18c25/26	12
2	A0A044SS18	Y	21	Y	20	Y	n11-22c26/27	12
3	A0A044UBU6	Y	**34**	Y	**20**	-	-	
4	A0A044QQ29	Y	22	Y	21	Y	n5-16c21/22	12
5	A0A044V9E9	Y	**22**	Y	**24**	Y	n6-16c21/22	11
6	A0A044R3V3	Y	19	Y	18	Y	n2-13c18/19	12
7	A0A044U1C0	N	-	Y	18	-	-	-
8	A0A044T183	Y	20	Y	19	-	-	-
9	A0A044SJY0	Y	30	Y	29	-	-	-
10	A0A044QUD5	Y	20	Y	19	-	-	-
11	A0A044QSP5	Y	23	Y	22	-	-	-
12	A0A044TUX9	Y	27	Y	26	-	-	-
13	A0A044RCD9	Y	20	Y	19	-	-	-
14	A0A044R864	Y	29	Y	28	-	-	-
15	A0A044RHS4	Y	**22**	Y	**20**	-	-	-
16	A0A044QJS4	Y	**17**	Y	**85**	-	-	-
17	A0A044QVN5	Y	25	Y	24	-	-	-
18	A0A044S3P5	Y	26	Y	25	-	-	-
19	A0A044V716	Y	19	Y	18	-	-	-
20	A0A044U3P3	Y	20	Y	19	-	-	-
21	A0A044SW64	Y	29	Y	28	-	-	-

22	A0A044R276	Y	20	Y	19	-	-	-
23	A0A044R6T0	N	-	Y	38	-	-	-
24	A0A044UMA2	N	-	Y	26	-	-	-
25	A0A044RQZ1	N	-	Y	38	-	-	-
26	A0A044STA6	N	-	Y	27	-	-	-
27	A0A044V4B5	N	-	Y	33	-	-	-
28	A0A044UDD9	N	-	Y	57	-	-	-
29	A0A044TUS0	N	-	Y	24	-	-	-
30	A0A044UQ33	N	-	Y	34	-	-	-
31	A0A044RM78	N	-	Y	168	-	-	-
32	A0A044U6S7	N	-	Y	33	-	-	-
33	A0A044SCV7	N	-	Y	42	-	-	-
34	A0A044R8T6	N	-	Y	69	-	-	-
35	A0A044TEN5	N	-	Y	43	-	-	-
36	A0A044RLS7	N	-	Y	40	-	-	-
37	A0A044U8H1	N	-	Y	33	-	-	-
38	A0A044UL22	N	-	Y	25	-	-	-
39	A0A044UEA6	N	-	Y	28	-	-	-
40	A0A044RPR7	N	-	Y	37	-	-	-
41	A0A044URQ7	N	-	Y	24	Y	n11-19C24/25	9
42	A0A044QVR1	N	-	Y	21	-	-	-
43	A0A044SLQ4	N	-	Y	33	-	-	-
44	A0A044UJI2	N	-	Y	26	-	-	-
45	A0A044SBK1	N	-	Y	36	-	-	-
46	A0A044SGM6	N	-	Y	29	-	-	-
47	A0A044VD77	Y	18	Y	17	Y	n2-12c17/18	11
48	A0A044SXP0	Y	18	Y	17	-	-	-
49	A0A044V990	Y	29	Y	28	-	-	-

50	A0A044U5M2	Y	18	Y	17	-	-	-
51	A0A044T320	Y	**25**	Y	**27**	-	-	-
52	A0A044UXB1	Y	20	Y	19	Y	n4-15c19/20	12
53	A0A044TMU6	Y	25	Y	24	Y	n6-17c24/25	12
54	A0A044TM82	Y	19	Y	18	Y	n4-12c18/19	9
55	A0A044SA24	N	-	Y	37	Y	n2-11c15/16	10
56	A0A044V148	N	-	Y	34	-	-	-
57	A0A044SBA6	N	-	Y	32	-	-	-
58	A0A044UB72	N	-	Y	21	-	-	-
59	A0A044U0N1	N	-	Y	21	-	-	-
60	A0A044U697	N	-	Y	87	-	-	-
61	A0A044S9P3	N	-	Y	18	-	-	-
62	A0A044REF6	N	-	Y	18	-	-	-
63	A0A044TD61	N	-	Y	17	-	-	-
64	A0A044ULE8	N	-	Y	28	-	-	-
65	A0A044UXB1	N	-	Y	19	Y	n4-15c19/20	12
66	A0A044RPX3	N	-	Y	19	-	-	-
67	A0A044TFH0	N	-	Y	31	-	-	-
68	A0A044UHI2	N	-	Y	27	-	-	-
69	A0A044SCF4	N	-	Y	70	-	-	-
70	A0A044V966	N	-	Y	43	-	-	-
71	A0A044RP47	N	-	Y	22	-	-	-
72	A0A044UE82	N	-	Y	23	-	-	-
73	A0A044TZH7	N	-	Y	123	-	-	-
74	A0A044SI04	N	-	Y	19	-	-	-
75	A0A044RY69	N	-	Y	24	-	-	-
76	A0A044RWM1	N	-	Y	44	-	-	-
77	A0A044V6E0	N	-	Y	32	-	-	-

78	A0A044S121	N	-	Y	81	-	-	-
79	A0A044UAE9	N	-	Y	30	-	-	-
80	A0A044RJP9	Y	18	Y	17	Y	n4-12c17/18	9
81	A0A044U293	Y	25	Y	24	Y	n4-14c24/25	11
82	A0A044URQ3	Y	25	Y	24	-	-	-
83	A0A044V9Z2	Y	**19**	Y	**23**	Y	n5-16c23/24	12
84	A0A044T8I3	Y	35	Y	34	-	-	-
85	A0A044R2H0	Y	26	Y	25	Y	n4-15c20/21	12
86	A0A044SLR1	Y	18	Y	17	Y	n2-13c17/18	12
87	A0A044SCH6	Y	20	Y	19	Y	n6-14c19/20	9
88	A0A044QK85	Y	**27**	Y	**28**	-	-	-
89	A0A044RIX6	Y	20	Y	19	Y	n3-14c19/20	12
90	A0A044V9E6	Y	24	Y	23	Y	n7-18c23/24	12
91	A0A044R4D8	Y	26	Y	25	Y	n8-19c29/30	12
92	A0A044R963	Y	23	Y	22	Y	n4-15c19/20	12
93	A0A044SF87	Y	22	Y	21	Y	n4-16c21/22	13
94	A0A044UXF1	N	-	Y	44	-	-	-
95	A0A044RSJ7	N	-	Y	18	Y	n2-13c18/19	12
96	A0A044RK54	N	-	Y	32	-	-	-
97	A0A044QSN9	N	-	Y	29	-	-	-
98	A0A044V6Q4	N	-	Y	32	-	-	-
99	A0A044SWW2	N	-	Y	39	-	-	-
100	A0A044V643	N	-	Y	131	-	-	-
101	A0A044RM40	N	-	Y	115	-	-	-
102	A0A044SVE5	N	-	Y	21	-	-	-
103	A0A044UIH1	N	-	Y	34	-	-	-
104	A0A044SG87	N	-	Y	16	-	-	-
105	A0A044S9U7	N	-	Y	26	-	-	-

106	A0A044U4G7	N	-	Y	41	-	-	-
107	A0A044RYR5	N	-	Y	24	-	-	-
108	A0A044T3F0	N	-	Y	73	-	-	-
109	A0A044U8Q1	N	-	Y	26	-	-	-
110	A0A044SJU9	N	-	Y	22	-	-	-
111	A0A044RXR8	N	-	Y	33	Y	n17-28c33/34	12
112	A0A044TS24	N	-	Y	26	-	-	-
113	A0A044SQ33	N	-	Y	32	-	-	-
114	A0A044R9E4	N	-	Y	35	-	-	-
115	A0A044V7H5	N	-	Y	23	Y	n6-17c29/30	12
116	A0A044TV73	N	-	Y	38	-	-	-
117	A0A044RQP7	N	-	Y	33	-	-	-
118	A0A044V478	N	-	Y	21	-	-	-
119	A0A044UZN4	N	-	Y	31	-	-	-
120	A0A044TUF9	N	-	Y	30	-	-	-
121	A0A044UBT4	N	-	Y	20	-	-	-
122	A0A044UHX1	N	-	Y	42	-	-	-
123	A0A044VAG2	N	-	Y	51	-	-	-
124	A0A044RXC4	N	-	Y	31	-	-	-
125	A0A044TXV3	N	-	Y	39	-	-	-
126	A0A044R0T9	N	-	Y	28	-	-	-
127	A0A044SAQ9	N	-	Y	19	-	-	-
128	A0A044SFP8	N	-	Y	31	-	-	-
129	A0A044QXD7	N	-	Y	30	-	-	-
130	A0A044T7K9	N	-	Y	146	-	-	-
131	A0A044TFG4	N	-	Y	30	-	-	-
132	A0A044U3W2	N	-	Y	24	-	-	-
133	A0A044UE02	N	-	Y	20	-	-	-

134	A0A044RVG8	N	-	Y	30	-	-	-
135	A0A044UT67	N	-	Y	38	-	-	-
136	A0A044SF73	N	-	Y	51	-	-	-
137	A0A044V3R2	N	-	Y	76	-	-	-
138	A0A044TYE4	N	-	Y	14	Y	n3-14c19/20	12
139	A0A044TUY7	N	-	Y	46	-	-	-
140	A0A044VEP2	N	-	Y	46	-	-	-
141	A0A044UBK0	N	-	Y	47	-	-	-
142	A0A044R2H5	N	-	Y	113	-	-	-
143	A0A044UJM3	N	-	Y	35	-	-	-
144	A0A044UGZ1	N	-	Y	22	-	-	-
145	A0A044SAS9	N	-	Y	77	-	-	-
146	A0A044UBG0	N	-	Y	32	-	-	-
147	A0A044R930	N	-	Y	29	-	-	-
148	A0A044SCX7	Y	19	Y	18	Y	n6-13c18/19	8
149	A0A044QYA0	Y	17	Y	16	Y	n4-11c18/19	8
150	A0A044UWG2	Y	18	Y	17	Y	n2-12c17/18	11
151	A0A044VG01	Y	21	Y	20	Y	n8-15c20/21	8
152	A0A044T6M2	Y	25	Y	24	-	-	-
153	A0A044RFL6	Y	**20**	Y	**20**	Y	n7-15c19/20	9
154	A0A044RS55	Y	26	Y	25	Y	n7-18c25/26	12
155	A0A044UXR5	Y	20	Y	19	Y	n4-11c19/20	8
156	A0A044TPD5	Y	**23**	Y	**14**	-	-	-
157	A0A044R5Q7	Y	21	Y	20	Y	n4-15c20/21	12
158	A0A044T2I2	N	-	Y	19	Y	n4-14c19/20	11
159	A0A044U4A3	N	-	Y	25	-	-	-
160	A0A044VI80	N	-	Y	22	-	-	-
161	A0A044RW03	N	-	Y	20	-	-	-

162	A0A044SBX3	N	-	Y	50	-	-	-
163	A0A044SI40	N	-	Y	36	-	-	-
164	A0A044SRL3	N	-	Y	34	-	-	-
165	A0A044TDH5	N	-	Y	48	-	-	-
166	A0A044RM57	N	-	Y	52	-	-	-
167	A0A044SQD0	N	-	Y	109	-	-	-
168	A0A044RY03	N	-	Y	19	Y	n8-18c23/24	11
169	A0A044SS11	N	-	Y	30	-	-	-
170	A0A044V0S5	N	-	Y	28	-	-	-
171	A0A044U006	N	-	Y	40	-	-	-
172	A0A044QKI7	N	-	Y	25	-	-	-
173	A0A044RER0	N	-	Y	30	Y	n14-25c30/31	12
174	A0A044V9S5	N	-	Y	27	-	-	-
175	A0A044S8R4	N	-	Y	23	Y	n11-19c23/24	9
176	A0A044UF12	N	-	Y	32	-	-	-
177	A0A044TYI5	N	-	Y	50	-	-	-
178	A0A044RPX9	N	-	Y	35	-	-	-
179	A0A044TCD5	N	-	Y	26	-	-	-
180	A0A044QVS5	N	-	Y	117	-	-	-
181	A0A044QZH7	N	-	Y	35	-	-	-
182	A0A044V1M0	N	-	Y	31	-	-	-
183	A0A044S9L1	N	-	Y	32	-	-	-
184	A0A044V7J3	N	-	Y	36	-	-	-
185	A0A044RH86	N	-	Y	31	-	-	-
186	A0A044U6R2	N	-	Y	42	-	-	-
187	A0A044UED5	N	-	Y	29	-	-	-
188	A0A044URH4	N	-	Y	38	-	-	-
189	A0A044RV21	N	-	Y	23	-	-	-

190	A0A044RT58	N	-	Y	52	-	-	-
191	A0A044RPP5	N	-	Y	18	-	-	-
192	A0A044VHX7	N	-	Y	35	-	-	-
193	A0A044RIT5	N	-	Y	25	-	-	-
194	A0A044U9P5	N	-	Y	29	-	-	-
195	A0A044RIL5	N	-	Y	20	-	-	-
196	A0A044RP25	N	-	Y	28	-	-	-
197	A0A044TW79	N	-	Y	32	-	-	-
198	A0A044V933	N	-	Y	27	-	-	-
199	A0A044QLY0	N	-	Y	27	-	-	-
200	A0A044RGI8	N	-	Y	33	-	-	-
201	A0A044RWL2	N	-	Y	30	-	-	-
202	A0A044ST38	N	-	Y	16	-	-	-
202	A0A044S1M8	N	-	Y	50	-	-	-
203	A0A044SBY6	N	-	Y	105	-	-	-
204	A0A044SUF3	N	-	Y	115	-	-	-
205	A0A044UZI4	N	-	Y	44	-	-	-
206	A0A044TTI6	N	-	Y	38	-	-	-
207	A0A044RXY6	N	-	Y	30	-	-	-
208	A0A044V4F5	N	-	Y	56	-	-	-
209	A0A044VEF0	N	-	Y	20	-	-	-
210	A0A044TH05	N	-	Y	37	-	-	-
211	A0A044V368	N	-	Y	28	-	-	-
212	A0A044QX16	N	-	Y	25	-	-	-
213	A0A044RTE4	N	-	Y	99	-	-	-
214	A0A044U6E4	N	-	Y	25	Y	n17-28c36/37	12
215	A0A044UG76	Y	21	Y	20	-	-	-
216	A0A044UF95	Y	33	Y	32	-	-	-

217	A0A044RAV7	Y	23	Y	22	-	-	-
218	A0A044TQV4	Y	**31**	Y	**24**	-	-	-
219	A0A044VAP0	Y	21	Y	20	Y	n3-13c17/18	11
220	A0A044SEG2	Y	20	Y	19	Y	n5-14c19/20	10
221	A0A044R983	Y	**17**	Y	**37**	Y	n2-9c14/15	8
222	A0A044VG24	Y	20	Y	19	Y	n7-17c21/22	11
223	A0A044VHR0	Y	23	Y	22	Y	n9-17c22/23	9
224	A0A044TJ65	Y	22	Y	21	-	-	-
225	A0A044TBY3	Y	22	Y	21	Y	n4-15c20/21	12
226	A0A044UA29	Y	25	Y	24	Y	n9-19c24/25	11
227	A0A044UNM4	Y	17	Y	16	Y	n2-11c16/17	10
228	A0A044RG47	Y	18	Y	17	Y	n2-12c17/18	11
229	A0A044QPE5	Y	**28**	Y	**24**	Y	n14-22c27/28	9
230	A0A044RQ45	Y	**17**	Y	**22**	Y	n4-14c22/23	11
231	A0A044UM18	Y	23	Y	22	Y	n6-17c22/23	12
232	A0A044SL30	N	-	Y	43	-	-	-
233	A0A044RCT5	N	-	Y	27	-	-	-
234	A0A044VGE2	N	-	Y	44	-	-	-
235	A0A044UFD5	N	-	Y	28	-	-	-
236	A0A044R6C9	N	-	Y	134	-	-	-
237	A0A044UCC9	N	-	Y	19	-	-	-
238	A0A044U2C2	N	-	Y	20	-	-	-
239	A0A044V9F1	N	-	Y	40	-	-	-
240	A0A044SF54	N	-	Y	22	-	-	-
241	A0A044RL99	N	-	Y	23	-	-	-
242	A0A044RPI9	N	-	Y	30	-	-	-
243	A0A044V562	N	-	Y	28	Y	n13-23c28/29	11
244	A0A044V9K7	N	-	Y	48	-	-	-

245	A0A044RET9	N	-	Y	100	-	-	-
246	A0A044RAF2	N	-	Y	30	-	-	-
247	A0A044UFR5	N	-	Y	31	-	-	-
248	A0A044U8E5	N	-	Y	27	-	-	-
249	A0A044RX04	N	-	Y	25	-	-	-
250	A0A044TNW9	N	-	Y	34	-	-	-
251	A0A044RVB1	N	-	Y	18	-	-	-
252	A0A044UQ20	N	-	Y	20	-	-	-
253	A0A044TR52	N	-	Y	28	-	-	-
254	A0A044RA41	N	-	Y	34	-	-	-
255	A0A044UF77	N	-	Y	27	-	-	-
256	A0A044S1E7	N	-	Y	32	-	-	-
257	A0A044RAX8	N	-	Y	35	-	-	-
258	A0A044VH58	N	-	Y	24	-	-	-
259	A0A044S742	N	-	Y	22	-	-	-
260	A0A044RJU5	N	-	Y	34	-	-	-
261	A0A044VGV8	N	-	Y	23	-	-	-
262	A0A044RU62	N	-	Y	60	-	-	-
263	A0A044SVZ2	N	-	Y	19	-	-	-
264	A0A044TIW5	N	-	Y	21	-	-	-
265	A0A044SWJ6	N	-	Y	42	-	-	-
266	A0A044U019	N	-	Y	41	-	-	-
267	A0A044SNK6	N	-	Y	29	-	-	-
268	A0A044V134	N	-	Y	46	-	-	-
269	A0A044SVZ7	N	-	Y	86	-	-	-
270	A0A044RWL0	N	-	Y	37	-	-	-
271	A0A044UAX4	N	-	Y	21	-	-	-
272	A0A044RJ42	N	-	Y	16	-	-	-

273	A0A044V489	N	-	Y	17	-	-	-
274	A0A044T7H0	N	-	Y	29	Y	n3-14c19/20	12
275	A0A044RP43	N	-	Y	30	-	-	-
276	A0A044SJT9	N	-	Y	25	-	-	-
277	A0A044QLJ3	N	-	Y	30	-	-	-
278	A0A044SEQ6	N	-	Y	34	-	-	-
279	A0A044RAZ9	N	-	Y	69	-	-	-
280	A0A044RZI1	N	-	Y	21	-	-	-
281	A0A044ST73	N	-	Y	29	-	-	-
282	A0A044SCI7	N	-	Y	30	-	-	-
283	A0A044RQ39	N	-	Y	30	-	-	-
284	A0A044RSX7	N	-	Y	22	-	-	-
285	A0A044TJH6	N	-	Y	26	-	-	-
286	A0A044S0W6	N	-	Y	51	-	-	-
287	A0A044RX39	N	-	Y	30	-	-	-
288	A0A044S9J1	N	-	Y	30	-	-	-
289	A0A044V598	N	-	Y	33	-	-	-
290	A0A044U434	N	-	Y	19	-	-	-
291	A0A044QU07	N	-	Y	22	-	-	-
292	A0A044V4V1	N	-	Y	27	-	-	-
293	A0A044VGD4	N	-	Y	29	-	-	-
294	A0A044UHW5	N	-	Y	88	-	-	-
295	A0A044RKR8	N	-	Y	30	-	-	-
296	A0A044U5P1	Y	22	Y	21	Y	n4-16c21/22	13
297	A0A044V8Z0	Y	**33**	Y	**27**	-	-	-
298	A0A044RWT9	Y	22	Y	21	Y	n7-18c23/24	12
299	A0A044R6S7	Y	22	Y	21	Y	n6-17c21/22	12
300	A0A044RMF0	Y	27	Y	26	Y	n10-21c26/27	12

301	A0A044TPU2	Y	30	Y	29	Y	n13-24c29/30	12
302	A0A044RC50	Y	20	Y	19	-	-	-
303	A0A044S787	Y	21	Y	20	-	-	-
304	A0A044SUI8	Y	22	Y	21	-	-	-
305	A0A044RM52	Y	24	Y	23	Y	n9-17c21/22	13
306	A0A044S8N0	Y	18	Y	17	Y	n8-21c26/27	14
307	A0A044T088	Y	21	Y	20	-	-	-
308	A0A044RXY3	Y	22	Y	21	Y	n7-17c21/22	11
309	A0A044TDN4	Y	25	Y	24	Y	n9-20c24/25	12
310	A0A044T0M0	N	-	Y	26	-	-	-
311	A0A044S138	N	-	Y	44	-	-	-
312	A0A044SUS3	N	-	Y	55	-	-	-
313	A0A044SNZ2	N	-	Y	39	-	-	-
314	A0A044S650	N	-	Y	22	-	-	-
315	A0A044QXU9	N	-	Y	23	-	-	-
316	A0A044V445	N	-	Y	35	-	-	-
317	A0A044RNV9	N	-	Y	22	Y	n9-20c28/29	12
318	A0A044TZH9	N	-	Y	26	-	-	-
319	A0A044RTY5	N	-	Y	70	-	-	-
320	A0A044RWY4	N	-	Y	36	-	-	-
321	A0A044VIS7	N	-	Y	44	-	-	-
322	A0A044V9Y1	N	-	Y	35	-	-	-
323	A0A044VG14	N	-	Y	21	-	-	-
324	A0A044S7B9	N	-	Y	47	-	-	-
325	A0A044SWB1	N	-	Y	30	-	-	-
326	A0A044U7F2	N	-	Y	18	Y	n3-13c17/18	11
327	A0A044VAC7	N	-	Y	98	-	-	-
328	A0A044VGU6	N	-	Y	18	Y	n6-14c18/19	9

329	A0A044V4A7	N	-	Y	38	-	-	-
330	A0A044S696	N	-	Y	42	-	-	-
331	A0A044V7P6	N	-	Y	32	-	-	-
332	A0A044RSM7	N	-	Y	51	-	-	-
333	A0A044TSZ1	N	-	Y	22	-	-	-
334	A0A044U1F8	N	-	Y	21	-	-	-
335	A0A044RJC4	N	-	Y	33	-	-	-
336	A0A044RD51	N	-	Y	31	-	-	-
337	A0A044RT52	N	-	Y	32	-	-	-
338	A0A044VAN4	N	-	Y	26	Y	n10-20c25/26	11
339	A0A044R301	N	-	Y	35	-	-	-
340	A0A044R021	N	-	Y	15	-	-	-
341	A0A044RNY2	N	-	Y	30	-	-	-
342	A0A044QMC5	N	-	Y	37	-	-	-
343	A0A044SLX7	N	-	Y	46	-	-	-
344	A0A044U104	N	-	Y	83	-	-	-
345	A0A044V6D0	N	-	Y	24	-	-	-
346	A0A044TW88	N	-	Y	37	-	-	-
347	A0A044QPP3	N	-	Y	100	-	-	-
348	A0A044QVD2	N	-	Y	54	-	-	-
349	A0A044SWS4	N	-	Y	108	-	-	-
350	A0A044UPQ4	N	-	Y	23	-	-	-
351	A0A044VHZ6	N	-	Y	33	-	-	-
352	A0A044S0E3	N	-	Y	107	-	-	-
353	A0A044UL06	N	-	Y	33	-	-	-
354	A0A044TW32	N	-	Y	33	Y	n3-13c22/23	11
355	A0A044SCL2	N	-	Y	30	-	-	-
356	A0A044UXT5	N	-	Y	30	-	-	-

357	A0A044T655	N	-	Y	28	-	-	-
358	A0A044RXR2	N	-	Y	37	-	-	-
359	A0A044QW19	N	-	Y	19	-	-	-
360	A0A044RIG3	N	-	Y	40	-	-	-
361	A0A044V136	N	-	Y	21	-	-	-
362	A0A044V162	Y	**20**	Y	**20**	Y	n2-13c20/21	12
363	A0A044S1I6	Y	32	Y	31	-	-	-
364	A0A044VG75	Y	24	Y	23	Y	n7-18c23/24	12
365	A0A044RMM8	Y	19	Y	18	Y	n5-13c18/19	9
366	A0A044T009	Y	23	Y	22	-	-	-
367	A0A044QSB0	Y	21	Y	20	Y	n6-16c20/21	11
368	A0A044UB98	Y	22	Y	21	-	-	-
369	A0A044SEB3	Y	17	Y	16	Y	n4-12c16/17	9
370	A0A044R521	Y	24	Y	23	Y	n8-18c23/24	11
371	A0A044TJF8	Y	24	Y	23	Y	n8-18c23/24	11
372	A0A044R0R1	Y	21	Y	20	Y	n4-16c20/21	13
373	A0A044T193	Y	20	Y	19	Y	n5-15c19/20	11
374	A0A044QZD1	Y	23	Y	22	Y	n7-17c22/23	11
375	A0A044QQD0	Y	22	Y	21	Y	n7-17c21/22	11
376	A0A044SBT5	Y	**30**	Y	**24**	Y	n11-19c24/25	9
377	A0A044T0J7	Y	**17**	Y	**21**	Y	n3-13c21/22	11
378	A0A044QSP0	Y	23	Y	22	-	-	-
379	A0A044RGR8	Y	24	Y	23	Y	n4-15c23/24	12
380	A0A044UBD0	N	-	Y	26	-	-	-
381	A0A044S9V7	N	-	Y	25	-	-	-
382	A0A044V8Y1	N	-	Y	20	-	-	-
383	A0A044RTY8	N	-	Y	28	-	-	-
384	A0A044UTQ0	N	-	Y	42	-	-	-

385	A0A044QPW3	N	-	Y	27	-	-	-
386	A0A044RJ90	N	-	Y	44	-	-	-
387	A0A044RW31	N	-	Y	15	-	-	-
388	A0A044U3T2	N	-	Y	22	-	-	-
389	A0A044RNY8	N	-	Y	40	-	-	-
390	A0A044QVL5	N	-	Y	30	-	-	-
391	A0A044TUS5	N	-	Y	34	-	-	-
392	A0A044RP61	N	-	Y	33	-	-	-
393	A0A044S9K0	N	-	Y	44	-	-	-
394	A0A044RUX0	N	-	Y	20	-	-	-
395	A0A044T4L7	N	-	Y	34	-	-	-
396	A0A044U4D4	N	-	Y	16	-	-	-
397	A0A044VF19	N	-	Y	26	-	-	-
398	A0A044SQ43	N	-	Y	30	-	-	-
399	A0A044SKR1	N	-	Y	30	-	-	-
400	A0A044VEJ7	N	-	Y	33	-	-	-
401	A0A044R6Z6	N	-	Y	26	-	-	-
402	A0A044RVU1	N	-	Y	27	-	-	-
403	A0A044V9B1	N	-	Y	21	-	-	-
404	A0A044T7U3	N	-	Y	30	-	-	-
405	A0A044RVD3	N	-	Y	34	-	-	-
406	A0A044RM49	N	-	Y	38	-	-	-
407	A0A044SQV9	N	-	Y	45	-	-	-
408	A0A044U337	N	-	Y	18	-	-	-
409	A0A044SRJ6	N	-	Y	24	-	-	-
410	A0A044REF1	N	-	Y	18	-	-	-
411	A0A044V915	N	-	Y	40	-	-	-
412	A0A044VE43	N	-	Y	27	-	-	-

413	A0A044U9I9	N	-	Y	23	-	-	-

'Y' = yes signal peptide present and 'N' = no signal peptide. TargetP and SignalP detect signal peptide in the N-terminal presequence and signal peptide cleavage site in the C-terminal end of the presequence respectively. '_' = no prediction. Phobius was used to confirm predictions (i.e. to eliminate false positives). The lower case 'n' is positively charged region of signal peptide, 'c' is the neutral but polar region and '/' indicates the cleavage site or point.