PERMANENT REFERENCE

Structural and Functional Characterization of novel, parasitic nematode-specific protein from *Setaria digitata*

A thesis submitted for the Degree of Doctor of Philosophy

PERMANENT REFERENCE NOT TO BE REMOVED FROM THE LIBRARY

W. W. P. Rodrigo
Faculty of Science
University of Colombo

September 2013

ABSTRACT

Setaria digitata is an animal filarial parasite infecting cattle and other ungulates, causing mild to severe disease such as neuropathological disorders, lumber paralysis, blindness etc. imparting large economic losses for whose livelihood in livestock farming in the Far East and Asia. Further, infection of human by this nematode causes various pathological conditions. In spite of this, the biology and parasitic nature of this organism is largely unknown and only few investigations have been currently carried out at its molecular level. Expressed Sequence Tags (ESTs) are an effective approach for the discovery of novel genes in unraveling biology and this approach was implemented in this study. Approximately 250 ESTs were examined and identified a cDNA clone of *S. digitata* and its' coding sequence showed a high degree of similarity to functionally and structurally unannotated novel sequences of parasitic nematodes. Therefore, molecular characterization was undertaken to address the functional and structural properties of this group of proteins using *S. digitata* novel gene (*SDNP*)/protein (SDNP) as a model gene/protein taking the premise of the close biological and biochemical resemblance of *S. digitata* to human filarial parasites. Further, the paucity of adult materials of human filarial parasite in Sri Lanka for detail biological studies was warranted for *S. digitata* to be used in such studies. The bioinformatics analyses carried initially with *SDNP*/SDNP revealed that the SDNP consists of 205 amino acids with a predicted molecular weight and isoelectric point of 22.9 kDa and 9.81, respectively. A search carried out using SDNP over nucleotide, EST and protein databases at NCBI, NEMBASE4 and Parasite Genomes Database (PGD) identified homologous counterparts from the human filarial parasites *Wuchereria bancrofti*, *Brugia malayi*, *Onchocerca volvulus* and *Loa loa*, the mouse filarial worm *Litomosoides sigmodontis* and swine parasitic nematodes *Ascaris suum*; and also diverged counterparts from the plant parasitic nematode *Meloidogyne hapla* and free living nematodes *Caenorhabditis elegans*, *Caenorhabditis briggsae* and *Caenorhabditis remanei* protein TAG-267. Phylogenetic analyses revealed this group of novel proteins (NPs) is undergoing divergent evolution. A search of ESTs at PGD showed that NP is expressed in all the stages of *B. malayi*. Secondary structure analyses of multiply-aligned sequences of NPs using Jpred server indicated NPs to be rich in beta-pleated structures. TMHMM server and beta barrel finder programme indicated NPs are neither transmembrane nor beta barrels proteins but are likely to be globular proteins. Further, the Motif discovery tool of MEME identified three novel potential motifs for NPs, of which only two are present in *C. elegans*, *C. briggsae*, *C. remanei* and *M. hapla*. Analyses of NPs using Signal IP, TargetP and Psort indicated NPs are devoid of signal peptide cleavage sites, not mitochondrial targeting peptides and localized to the nucleus respectively. Further analyses of the NPs using ScanProsite server for phosphorylation, revealed potential sites for cAMP- and cGMP-dependent protein kinase, Protein kinase C and Casein kinase II. Putative functional analysis using ProtFun 2.1 Server indicated NPs are nonenzymatic, growth factor like protein.

Gene expression analysis using RT-PCR indicated that this gene is ubiquitously expressed in adult female & male and microfilarial stages. Southern hybridization studies revealed that this gene is a single-copy gene. Sequence analysis of the genomic region obtained by overlapping PCR amplification using primers derived from exonic regions indicated that the genomic region is 1819 bp. Gene has four exons encoding 205 amino acids were interrupted by three introns of varying lengths of 389, 659 and 123 bp respectively. Expansion of the size of the introns of *S. digitata* compare to its othologues by integrating micro and mini-satellite containing sequence of $(TA)_n$, $(CA)_n$, $(AT)_n$, $(GT)_n$, $(AC)_n$, $(CC)_n$, $(TG)_n$, $(CAA)_n$, $(AAT)_n$,

$(CAC)_n$, $(CCA)_n$, $(CAG)_n$, $(ACG)_n$, $(ACA)_n$ ect. into intronic regions were also observed. Sequences around the splice junctions were fairly conserved and agreed with the general GU-AG splicing rule. The gene was found to be AT rich with a GC content of 38.1%.

Protein expression was carried out by means of heterologous expression in both *Pichia pastoris* and pET expression systems to understand the structure and function of the SDNP. In *Pichia* expression system, SDNP was expressed both as secretory and intracellular protein. SDS-PAGE and Western blot analyses revealed low level of expression of recombinant SDNP (r-SDNP) in *P. pastoris*, which was a single band in Western blot analyses with a molecular weight of ~23.8 kDa that was similar to the theoretical molecular weight. Recombinant protein expressed in *P. pastoris* was successfully purified using Ni affinity chromatography. Since the achieved expression of r-SDNP was low in *P. pastoris* system, the bacterial expression was carried out and the considerable level of expression was seen when *E. coli* strain BL21(DE3) was used. The recombinant protein was largely concentrating into the insoluble fraction (inclusion bodies) suggesting that the protein to be more hydrophobic than hydrophilic. Therefore, the co-expression of SDNP was carried out using chaperone plasmids (pG-KJE8, pGro 7, pKJE7, pG-Tf2 and pTf16) mediated system in BL21(DE3). This revealed a significant increase of the recombinant protein in soluble fraction when the pGro7 was used as the chaperone plasmid with 2 mg/ml L-arabinose. r-SDNP was purified both from soluble and insoluble fractions using Ni-NTA resin and the purity of the r-SDNP was confirmed by both SDS-PAGE and Western blot analyses by obtaining a single band with the expected size of ~24 kDa. The purified r-SDNP from the soluble and insoluble fractions was used for structural studies and for the development of polyclonal antibodies for functional studies respectively.

Immunohistochemical staining of body sections of *S. digitata* was carried out using polyclonal antibody prepared against recombinant SDNP; the highest expression of SDNP was seen in the longitudinal muscles of the body wall of adult males and females indicating its possible involvement in parasite locomotion and tissue modeling. Moderate expression was observed in the reproductive organs of both sexes suggesting its role in male and female reproduction. A low level of expression was observed in the cuticle, syncytial hypodermis region, lateral line and the intestinal wall. Further, the expression of SDNP was also seen in developing microfilariae within the uterus of female worms, developing spermatozoa of males and different developmental stages of embryos implicating its involvement in nematode growth and development. Subcellular localization of SDNP carried out in yeast, *Pichia pastoris* using green fluorescence construct revealed that this protein localized mainly in nucleus and partly in the cytoplasm. Comprehensive bioinformatics analyses on what was seen in immunohostochemical staining revealed that this protein contains a nuclear localization signal, RNAP_Rpb7_N_like domain, regions that are homologous to a part of the nuclear factor localization-like domain, interdomain linkers of muscle specific twitchin kinase of *C. elegans* and calcium-dependent protein kinase isoform CDPK1 of *Arabidopsis thaliana*. Therefore, considering all these outcomes together, it can be concluded that the *SDNP* is a monocopy gene having four exons and three introns and ubiquitously expressed in all stages of *S. digitata* life cycle while the SDNP is a growth factor like, parasitic nematode specific, beta-pleated strands rich, muscle and reproductive system protein that may be localized in both the nucleus and cytoplasm, and regulated via phosphorylation and dephosphorylation, and involved in growth and development as it was profoundly found in all the developmental stages investigated in this study. The finding reported in this study shed light on the biology of hither to unannotated parasitic nematode specific novel genes/proteins.