Characterization of the Hookworm FOXO Transcription Factor DAF-16 and its Target Genes

By Xin Gao

B.S. in Biochemistry, June 1996, Fudan University
M.S. in Pathogenic Biology, June 1999, Chinese Center for Disease Control and Prevention

A Dissertation submitted to

The Faculty of
Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

May 15th, 2011

Dissertation directed by

John M. Hawdon
Associate Professor of Microbiology, Immunology and Tropical Medicine

The Columbian College of Arts and Sciences of The George Washington University certifies that Xin Gao has passed the Final Examination for the degree of Doctor of Philosophy as of March 21st, 2011. This is the final and approved form of the dissertation.

Characterization of the Hookworm FOXO Transcription Factor DAF-16 and its Target Genes

Xin Gao

Dissertation Research Committee:

John M. Hawdon, Associate Professor of Microbiology, Immunology, and Tropical Medicine, Dissertation Director

Paul Brindley, Professor of Microbiology, Immunology, and Tropical Medicine, Committee Member

Sydney Fu, Associate Research Professor of Biochemistry & Molecular Biology, Committee Member

Acknowledgements

I would like to thank my mentor Dr. John Hawdon for the time and effort he devoted to my research project throughout my graduate studies. His mentorship has been instrumental and influential not only to my research but also to my development as a scientist. He accepted me into the lab when I had zero experience with hookworms, for which I will be grateful for the rest of my life.

I would like to gratefully acknowledge the guidance, insight, and support of my dissertation committee: Drs. Paul Brindley, Susan Ceryak, and Sydney Fu. I would also like to thank Dr. Makedonka Mitreva (Washington Univeristy in St. Louis) for her willingness to invite me to work with her group. The bioinformatic expertise she provided was invaluable to my project.

I am grateful to all the former and current members of Dr. Hawdon's group and the entire MITM department for their help in countless ways and their forever friendship. I would like to thank Dr. Linda Werling and David Leitenberg, as academic advisors, for their enthusiasms and constant kindness.

Finally, I would like to express my gratitude to my parents and all my lifetime friends, for their unconditional love and support.

Abstract

Characterization of the Hookworm FOXO Transcription Factor DAF-16 and its Target Genes

Hookworm disease is an intestinal parasitic disease of humans and animals. It continues to rank among the most important neglected tropical diseases. Blood loss resulting from heavy hookworm infections has devastating impacts for children, pregnant women, and the elderly. As do many parasitic nematodes, hookworms undergo developmental arrest as free-living infective larvae (L3) prior to host entry, and their post infection development as a parasite is re-activated by host-specific signals. The transition to parasitism and subsequent development is poorly understood.

Accumulating evidence suggests that the insulin-like signaling (ILS) pathway is involved in hookworm infection and development. A key output of ILS pathway is the forkhead transcription factor DAF-16/FOXO, which has been implicated in diverse biological processes. In this study, an ortholog of DAF-16 from hookworm *A. caninum, Ac*-DAF-16, was characterized. The amino acid sequence of *Ac*-DAF-16/FOXO DNA binding domain (DBD) shared high identity (94%) with the b and c isoforms of *C. elegans* DAF-16. Other functional motifs including Akt/PKB phosphorylation sites and 14-3-3 protein binding sites common to DAF-16/FOXO, are also conserved in *Ac*-DAF-16/FOXO. We demonstrated that *Ac*-DAF-16/FOXO was transcriptionally active through a canonical DBE (DAF-16 family protein binding element) and negatively regulated by growth factor stimulation.

To better understand the roles of *Ac*-DAF-16/FOXO in hookworm parasitism, we combined molecular techniques and bioinformatics to identify *Ac*-DAF-16/FOXO binding sites in the hookworm genome and a group of its target genes. The DBD of *Ac*-

DAF-16 was used to select genomic fragments by in vitro genomic selection. Twenty four bound genomic fragments were analyzed for the presence of the DBE and putative alternative *Ac*-DAF-16/FOXO binding motifs. The 22 genes linked to those genomic fragments were identified as direct gene targets of *Ac*-DAF-16/FOXO and their developmental stage-specific expression patterns were examined. The results suggested that *Ac*-DAF-16/FOXO was involved in many biological processes throughout hookworm development.

Comprehensive characterization of *Ac*-DAF-16/FOXO gene targets will provide insights into the molecular mechanisms of its regulatory functions. For this reason, we selected an identified *Ac*-DAF-16/FOXO target gene in our genome wide screening studies, *Ac-snr*-3, for further investigation, since it was linked to a genomic fragment with strong binding affinity to *Ac*-DAF-16/FOXO. *Ac-snr*-3 encodes a core Sm protein, Sm-D1, in snRNP (small nuclear ribonucleoprotein) particles. It spans at least 5.6 kb in genome and contains 3 exons and 2 introns. The transcriptional profile of hookworm *snr*-3 for post infection developmental stages suggested a function in regulating hookworm maturation into the adulthood. Its 3′-end genomic fragment (Fragment 2.23) was able to drive transcription of a reporter gene in the presence of *Ac*-DAF-16/FOXO, regardless of the location relative to reporter gene, indicating that it might act as *cis*-regulatory element mediating the effects of *Ac*-DAF-16/FOXO in hookworm. The unresponsiveness of this fragment to serum stimulation suggested that co-acting protein factors were involved in its interaction with *Ac*-DAF-16/FOXO.

In overview, the findings presented in this dissertation identified a key transcription factor in hookworm ILS, *Ac*-DAF-16/FOXO, and demonstrated that it

regulated gene expression through the specific *cis*-regulatory elements in hookworm development.

Table of Contents

Acknowledgements ..... **iii**

Abstract ..... **iv**

List of Figures ..... **ix**

List of Tables ..... **xi**

List of Abbreviations ..... **xii**

Chapter 1. Background and General Introduction **1**

Hookworm disease and global public health *1*

Hookworm life cycle and nematode parasitism *3*

"Dauer hypothesis" for hookworm infectious process and insulin-like signaling (ILS) Pathway *6*

Forkhead transcription factor DAF-16/FOXO *9*

Regulatory mechanism of FOXO/DAF-16 transcriptional activity *12*

Consensus binding sequence for DAF-16 and downstream gene targets *19*

Hookworm Genome Project *23*

Dissertation outlines *27*

Chapter 2. Molecular cloning and DNA binding characterization of DAF-16 orthologs from *Ancylostoma* hookworms **34**

Abstract *35*

1. Introduction *36*

2. Materials and methods *38*

3. Results *47*

4. Discussion *55*

Chapter 3. Identification of hookworm DAF-16/FOXO response elements and direct gene targets **67**

Abstract *68*

1. Introduction *69*

2. Materials and Methods *71*

3. Results *79*

4. Discussion *85*

Chapter 4. Hookworm small nuclear ribonucleoprotein-3: gene structure, expression profile, and implications of 3'-untranslated region (3'-UTR) in the regulation of gene expression **101**

Abstract *102*

1. Introduction *103*

2. Materials and Methods *106*

3. Results *112*

4. Discussion *116*

Chapter 5. General Discussion **131**

References **141**

List of Figures

Figure 1.1 Hookworm life cycle 30

Figure 1.2 *C. elegans* life cycle. 31

Figure 1.3 Schematic representation of secondary structure elements within forkhead domain of FOXO protein. 32

Figure 1.4 Schematic diagram of insulin/IGF-1 signaling pathway in *C. elegans* dauer formation and exit. 33

Figure 2.1 The phylogenetic relationship between nematode DAF-16 and selected forkhead protein family members.. 62

Figure 2.2 Expression of *Ac-daf-16* mRNA during the hookworm life cycle. 63

Figure 2.3 Daf-16 expression in mammalian cells and antiserum characterization. 64

Figure 2.4 Streptavidin bead pull-down to detect DBE binding activity of recombinant *Ac*-DAF-16.. 65

Figure 2.5 Recombinant *Ancylostoma caninum* DAF-16 drives expression from the DBE. 66

Figure 3.1 Amino acid sequences of DNA binding domains from selected FOXO transcription factors 91

Figure 3.2 Expression of *Ac*-DAF-16 DBD (aa 214-314) and *Δ Ac*-DAF-16 DBD (aa 220-292) in *E. coli* Rosetta (DE3) cells... 92

Figure 3.3 Streptavidin bead pull-down to detect DBE binding activity of recombinant *Ac*-DAF-16 DBD and Δ*Ac*-DAF-16 DBD. 93

Figure 3.4 *In vitro* genomic selection of *A. caninum* DNA fragments containing *Ac*-DAF-16 binding sites 94

Figure 3.5 Identification of a novel DAF-16 DBD binding element. 95

Figure 4.1 *Ac-snr*-3 gene. 123

Figure 4.2 Phylogenetic relationship between Ac-SNR-3 and selected Sm-D1 proteins..... 125

Figure 4.3 Transcription of hookworm *snr*-3 during post infection developmental stages.. 126

Figure 4.4 Schematic representation of luciferase constructs. 127

Figure 4.5 *Ac*-DAF-16 drives expression of luciferase from Fragment 2.23. 128

Figure 4.6 Predicted transcription factor binding sites located on Fragment 2.23 130

List of Tables

Table 1. Proposed binding motifs found in fragments that bound to DAF-16 DBD *in vitro*. 96

Table 2. Expression pattern of genes that bound to DAF-16 DBD in vitro. 97

Table 3. GenBank Accession numbers and dbGSS identification numbers of *Ancylostoma caninum* genomic fragments isolated by genomic selection....98

Table 4. Frequencies of DBE containing genome fragments recovered from immobilized r*Ac*-DAF-16 DBD genomic selection and control genomic selection.....99

Table 5. Identity and stage-specific relative transcript levels of *Ac*-DAF16-DBD-selected genes.....100

List of Abbreviations

ABC transporter	ATP-binding cassette transporter
ASP	activation-associated secreted protein
BLAST	Basic Local Alignment Search Tool
BU	hookworm storage buffer
cAMP	cyclic adenosine monophosphate
cDNA	complementary (reverse transcribed) DNA
ChIP	chromatin immunoprecipitation
CK1	casein kinase-1
CPT-II	carnitine palmitoyl transferase
DAE	DAF-16 associated element
DALY	disability adjusted life years
DAO	dauer or aging adult overexpression
DBD	DNA binding domain
DBE	DAF-16 family protein binding element
DD-PCR	differential display PCR
DYRK1A	dual-specificity tyrosine-phosphorylated and regulated kinase
EST	expressed sequence tag
FKHR	forkhead homolog-1
FKTF	forhead transcription factor
FOX	forkhead family
FTT	14-3-3 protein
gDNA	genomic DNA

GFP	green fluorescence protein
GSS	genome survey sequences
HCF-1	human cell factor-1
HSP	Heat shock protein
ILS	insulin-like signaling
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JNK	Jun-N-terminal kinase
L1	first stage larva
L2	second stage larva
L3	third stage larva
L4	fourth stage larva
LMM	laser microdissection microscopy
MGE	mobile genetic elements
mRNA	messenger RNA
MST	mammalian ste20-like kinases
NAD	nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
NLP-2	neuro-peptide like protein family-2
NTD	neglected tropical disease
ORF	open reading frame
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDK	phosphoinositide-dependent kinase

PGC-1α	peroxisome proliferator-activated receptor-γ coactivator
PI3K	phosphatidyl inositol 3-kinase
PIP2	phosphatidylinositol 4,5-bisphosphate
PIP3	phosphatidylinositol (3,4,5)-trisphosphate
PKA	protein kinase A
PKB	protein kinase B
PTEN	phosphatase and tensin homolog
RLE	regulation of longevity by E3
RNAi	RNA interference
RT	reverse transcriptase
RT-PCR	reverse transcriptase PCR
SELEX	systematic evolution of ligands by exponential enrichment
SGK	serum- and glucocordicoid-inducible kinase
SIR	silent information regulator
SLE	systemic lupus erythematosus
SNR/snRNP	small nuclear ribonucleoprotein
SOD	superoxide dismutase
SRT-42	serpentine receptor, class T-42
UTR	untranslated region
WHO	world health organization

Chapter 1. Background and General Introduction

Hookworm disease and global public health

Hookworm disease is an intestinal parasitic disease of humans and animals. The two primary hookworm species that infect humans are *Ancylostoma duodenale* and *Necator americanus,* originating in Asia and Africa respectively, but now globally distributed (Cox, 2002). Human hookworm infection is primarily acquired through skin contact with infective larvae that exist in contaminated soil. Following host entry, the larvae travel through vasculature, lungs and other tissues, before they enter small intestine to develop into adult parasites (Hotez et al, 2004).

The major pathology caused by hookworm infection is intestinal blood loss resulting from damage of the small intestine mucosa and sub-mucosa due to adult worm attachment and feeding behavior. Significant intestinal blood loss occurs in heavy infections, and iron-deficiency anemia and malnutrition develops if the blood loss exceeds nutritional intake and reserves of the host (Hotez & Pritchard, 1995; Stoltzfus et al., 1997; Albonico et al., 1998; Bethony et al., 2006). Chronic and severe iron-deficiency anemia has adverse and sometimes fatal consequences for children, pregnant women and the elderly. Additional hookworm-related pathological symptoms include ground itch, creeping eruption, and hookworm pneumonitis arising from larval penetration and migrations (Blackwell et al., 2001; Maxwell et al., 1987).

Hookworm disease has plagued the world for centuries and remains a major public health concern (Peduzzi & Piffaretti, 1983). The current prevalence of human hookworm infections is estimated at 740 million according to The World Health

Organization (WHO) report, predominantly in tropical and subtropical developing countries (WHO, 2010). Therefore, hookworm disease is collectively listed with sixteen other poverty-promoting infections as neglected tropical diseases (NTD) that largely affect populations with low visibility and little political voice. In many of those areas, hookworm infection commonly co-exists with other pathogen infections, including other helminths, bacteria, and viruses, in an individual patient (Lipner et al., 2006; Brooker et al., 2006; Fleming et al., 2006;Walson et al., 2010; Bhattacharya et al., 2010). The latest number of Disability Adjusted Life Years (DALYs) assigned to hookworm disease was over a million, ranking the global disease burden from hookworm among the most important tropical diseases (WHO, 2004). This concept of DALYs translates the long-term impact of infection on the overall life of the host, including states of poor health or disability, into lost years of healthy life so that the burden of individual disease can be assessed quantitatively and comparatively (Anand & Hanson, 1997; Murry, 1994).

Treatment of hookworm disease involves deworming with anthelmintic drugs plus iron supplements to correct iron deficiency (Horton, 2002; Stoltzfus et al., 1997). However, re-infection after treatment occurs rapidly in endemic areas, tissue-arrested larvae are relatively resilient to the effects of anthelmintics, and there always exists a potential for the development of drug resistance (Arasu, 1998). Substantial reduction of hookworm infection will obviously need a combination of traditional control strategies with novel drugs and/or vaccines as well as improved sanitation.

Hookworm life cycle and nematode parasitism

Hookworms belong to the phylum Nematoda. In addition to two major species of hookworms infecting humans, there are hookworms of dogs (*Ancylostoma caninum, A. ceylanicum, Uncinaria stenocephala*), cats (*A. tubaeforme, A. braziliense*), cattle (*Bunostomum phlebotomum*), goats and sheep (*Gaigeria pachyscelis*), and pigs *(Globocephalus urosubalatus)*(Kalkofen, 1987; Kaufmann, 1996; Flynn & Baker, 2007; Bowman et al., 2010). *Ancylostoma ceylanicum* also infects humans naturally in some areas (Landmann & Prociv, 2003).

As many other parasitic members in phylum Nematoda, hookworms are dioecious and have a direct life cycle involving free-living stages in the environment and parasitic stages in the definitive host (Figure 1.1). Adult hookworms live and mate in the small intestine of the host. The female worms produce up to 30,000 eggs per worm per day, which are passed out of the body to the environment with feces. Eggs hatch in the warm, moist environment and proceed through a first-stage Larva (L1) and L2 to the infective L3. This development occurs in 5-10 days, depending on temperature and humidity. L1 and L2 are rhabditiform stages and feed on surrounding soil microbes. The infective L3 is an ensheathed, non-feeding filariform dispersal stage, and remains developmentally arrested in the environment for weeks or months until it encounters the host (Roberts & Janovy, 2005).

The infective L3 of different hookworms enter the host differently, either through skin penetration or orally. *N. americanus* L3 only infect by penetration of skin, *Uncinaria stenocephala* primarily infect through oral infection, while *A. duodenale*, *A. caninum*, *A. ceylanicu*, *A. tubaeforme*, *A. braziliense*, *B. phlebotomum*, *G. pachyscelis*, and *G.*

urosubalatus are capable of infecting by either route (Yoshida, 1973; Epe, 2009). Skin-penetrated L3 enter the circulation system, migrate to lungs, break out into the bronchial lumen, and then ascend the respiratory tree to the trachea to be coughed into the mouth, where they are swallowed. Ingested L3 undertake a relatively simple journey by directly passing through the stomach and traveling to the small intestine. Within the definitive host, specific signals reactivate suspended development of the L3, causing the molt to the L4 and subsequent maturation to the adulthood. The timing of post infection molts differs for individual species.

During its life cycle, hookworms are exposed to different environmental conditions, and therefore undergo morphological and physiological transformation. The critical factors for hookworm survival and transmission appear to be its ability to persist in the external environment as a developmentally arrested L3 for long enough periods until transmission occurs, and then resume development in the host by properly responding to host-specific signals and faithfully executing the developmental programs. The existence of an arrested development larval stage is common in life cycles of many nematodes, including both free-living and parasitic species.

The best-studied developmental arrest stage in nematodes is the dauer stage of the free-living nematode *Caenorhabiditis elegans* (Félix & Braendle, 2010). *C. elegans* transiently enters dauer stage in response to adverse environmental conditions such as high population density, increased temperature and lack of food. When optimal growth conditions are restored, *C. elegans* dauer larvae exit from the dauer stage and re-enter the continuous reproductive developmental cycle (Cassada & Russell, 1975). The length of

time spent in the dauer stage seems to have no effect on the life span of post-dauer adults; therefore, dauer larvae are considered non-aging (Klass & Hirsh, 1976).

In contrast to many parasitic nematodes, developmental arrest as a dauer larva in *C. elegans* is facultative, and can be bypassed if living conditions are suitable for the worm to molt through larval stages from egg to adulthood without being interrupted (Figure 1.2). L3 arrest in parasitic nematodes is instead an obligatory arrested stage prior to host entry. Exit from this programed arrest stage after host entry represents the return to the reproductive growth as well as establishment of a parasitic relationship with the host. The process involves detecting and processing host-specific signals, resuming feeding and development, evading the host's immune system, and modulating the local microenvironment. Distinct biochemical adaptations, including a shift from aerobic metabolism to anaerobic metabolism, occur during the transition to parasitism in nematodes (Fry & Jenkins, 1984). The extent of utilizing anaerobic metabolism within the host depends on the parasites' body diameter and oxygen availability in their habitat. In hookworms, aerobic pathways are favored in infective L3s, whereas both aerobic and anaerobic pathways operate effectively in adult worms (Onwuliri, 1985; Goyal et al., 1991).

The infection, migration and colonization of the definitive host by parasitic nematodes are closely correlated with the pathogenesis of the parasitic infections. Therefore, dissecting the molecular events of the parasitic infection might have significant implications for development of novel and more effective strategies for parasitic nematode disease control. Until recently, difficulties in the *in vitro* cultivation of

parasitic nematodes and inability to apply advance genetic approaches have severely hampered these studies.

In the past decade, considerable advances in understanding of *C. elegans* development have revealed many parallel aspects between parasitic and non-parasitic development, and created an ideal framework for in-depth investigations of parasitism mechanisms. Specifically, the signal transduction pathways involved in regulating *C. elegans* developmental arrest have been well characterized and widely adopted to the research on parasitism (Beall & Pearce, 2002).

"Dauer hypothesis" for the hookworm infectious process and the insulin-like signaling (ILS) Pathway

The term "dauer" was first coined to describe the enduring stage of insect-parasitic nematodes and later applied to other parasitic nematodes and *C. elegans* (Webster, et al 2008). The normal life cycle of *C. elegans* from embryo to adult takes ~3.5 days at 20°C with adequate food supply and minimal growth competition. Adult worms typically live another 2~3 weeks. The dauer larva of *C. elegans* is an alternative, resistant, and reversible developmental stage that initiates at late L1 stage and enables the worm to extend survival for several months upon exposure to harsh living conditions. It is characterized by vast remodeling changes throughout the entire body, such as radial shrinkage of hyperdermis, neuron alteration, mouth plugging, body fat accumulation, growth reduction, and reproduction inhibition (Riddle & Albert, 1997).

The infective L3 of parasitic hookworm has very similar biological requirements for survival prior to finding the definitive host, and bear a striking resemblance to *C.*

elegans dauer stage in terms of behavior, morphology and biochemical features. Therefore, the "dauer hypothesis" proposed that hookworm L3s are "dauer constitutive", and the same cellular and molecular mechanisms are utilized by both infective hookworm L3 and *C. elegans* dauer to control the arrest-related morphogenesis (Hawdon & Schad, 1991a; Hotez et al., 1993).

The infective L3 initiates infection once a definitive host is located. Events of the infectious process are undoubtedly essential for the hookworm's subsequent parasitic stages; however, its fundamental aspects are still largely unexplored. The canine hookworm *A. caninum* has long been used as a lab model for investigation of human hookworm diseases. An *in vitro* serum stimulating system using resumption of feeding by *A. caninum* L3 as marker of activation was developed to mimic early events of hookworm infection, and has led to the identification of several molecules that are released during the infectious process (Hawdon, et al, 1990; Hawdon et al., 1996). This system has also been employed to demonstrate that cGMP signaling and a muscarinic pathway are involved in larval activation (Hawdon et al., 2003; Tissenbaum et al., 2000). However, recent analyses of the gene expression profile of *in vitro* serum-stimulated L3 showed that mRNAs encoding extracellular proteins with putative roles in host-parasite interactions are over-represented, but not those involved in growth and development, indicating that this *in vitro* system is limited for fully understanding the hookworm infectious process (Moser et al., 2005; Datu et al., 2008).

Given the relatively close phylogenetic relationship between hookworm and *C. elegans* as well as similarities between resumption of hookworm L3 development in the host and dauer exit of *C. elegans*, it stands to reason that the mechanism of dauer exit in

C. elegans might be conserved in the initiation of hookworm infection (Rogers & Sommerville, 1963; Hawdon & Schad, 1991a). Genetic mutations in *C. elegans* have produced numerous abnormal dauer phenotypes, many of which were used to define the insulin-like signaling pathway (ILS) as one of the critical cellular signal transduction pathways that regulate dauer formation and recovery (Riddle et al., 1981; Hu, 2007). Constitutive *C. elegans* dauer mutants (Daf-c) always enter the dauer, whereas defective ones (Daf-d) bypass dauer even in poor conditions (Riddle et al. 1981; Albert and Riddle, 1988). More importantly, the availability of the *C. elegans* genome sequence has greatly facilitated the isolation and analysis of genes involved in *C. elegans* dauer formation (*daf*) (The *C. elegans* Sequencing Consortium, 1998).

The first gene described in the *C. elegans* ILS pathway was *daf*-2, which encodes the only ortholog of the mammalian insulin receptor. Strong *daf*-2 mutants constitutively entered dauer diapause (Kimura et al, 1997). Subsequent ILS components identified in *C. elegans* include phosphatidyl inositol 3-kinase (PI3K/AGE-1), phosphoinositide-dependent kinase-1 (PDK-1), DAF-18 /PTEN lipid phosphotase, Akt/Protein kinaseB (PKB), Serum- and glucocorticoid-inducible kinase-1 (SGK-1), and forkhead/FOXO transcription factor DAF-16 (Dorman et al., 1995; Paradis et al., 1999; Paradis & Ruvkun, 1998; Ogg & Ruvkun, 1998; Hertweck et al., 2004; Lin et al., 1997; Ogg et al., 1997).

In addition to regulating the dauer stage, ILS also controls many other biological processes such as life span, fat storage, reproduction, and stress responses in *C. elegans* (Barbieri et al; Honda & Honda, 1999; Lee et al., 2003; Murphy et al., 2003; Murphy, 2006). ILS also interacts with multiple cellular signaling pathways, together coordinating

a complex network of gene expression and cellular responses to affect worm behavior as a whole (Vowels & Thomas, 1992; Jia et al., 2004; Walker et al., 2003).

In parasitic hookworms, the specific inhibitor of PI3K, LY294002 (2-(4-morpholinyl)-8-phenyl-4H-1-benzopryan-4-one), inhibits *in vitro* activation of L3 (Brand et al., 2004). In addition, an expressed sequence tag (EST) encoding DAF-16 ortholog from hookworm *A. ceylaynicum* was reported (Mitreva et al., 2005). Taken together, these results strongly suggest that ILS pathway might be similarly critical for hookworm development and its infectious process.

Forkhead transcription factor DAF-16/FOXO

daf-16 is one of the over 30 genes that regulate dauer larva formation in *C. elegans* (Riddle et al., 1981; Albert & Riddle, 1988). Extensive genetic screens for *C. elegans* daf (dauer formation) mutants and the following epistasis analysis suggest that *daf-16* promotes dauer formation, and might represent the principal negatively regulated output of ILS pathway (Gottlieb et al., 1994; Riddle et al., 1981). *daf-16* has been genetically mapped to cosmid R13H8. Its genomic DNA sequence spans at least 14 kb and its cDNA consists of 10 exons. The predicted coding region is homologous to members of the forkhead family (Lin et al., 1997; Ogg et al., 1997).

The forkhead family (Fox) is a large superfamily of transcription factors with more than 100 structurally related proteins. They exist in both eukaryotes and prokaryotes, but not in plants, and play important roles in regulating expression of genes involved in embryogenesis, tumorigenesis, and differentiation (Tuteja & Kaestner, 2007a; Tuteja & Kaestner, 2007b). Each member shares a conserved DNA binding domain of

appromixmately 100 amino acids, known as the forkhead domain. This domain contains three major α-helices and two large wing-like loops that form a "winged helix" structure (Figure 1.3) (Clark et al., 1993). Fox genes have been grouped into 18 subclasses, named from A to R (Kaestner et al., 2000; Hong et al., 2001; Shu et al., 2001; Schubert et al., 2001). DAF-16 and its mammalian counterparts, AFX, FKHR, and FKHRL1, belong to the "O" class (FOXO). The distinguishing feature of those FOXO proteins is the presence of highly conserved Akt / PKB phosphorylation sites within and near the forkhead domain.

Molecular and genetic data have revealed that there are two major DAF-16 isoforms, DAF-16c (formerly DAF-16a1) and DAF-16a (formerly DAF-16b), in *C. elegans,* which share a common C-terminal 319 amino acids and are probably generated by alternative promoters (Ogg et al., 1997). Both DAF-16a and DAF-16c contain a characteristic forkhead domain and four consensus Akt / PKB phosphorylation sites [RXRXX(S/T)]. Their DNA binding domains are not identical, but highly related. The Akt/PKB sites are located in the N-terminal, C-terminal, as well as forkhead domain, and three of them are found to be conserved in all the members of the mammalian FOXO family. The phenotypes of mutant strains containing different *daf-16* alleles as well as transgenic strains with *daf-16*-specific RNAi suggested that these two isoforms have different tissues-specific expression patterns and distinct functions, and DAF-16c has a more central function than DAF-16a in dauer arrest (Ogg et al., 1997; Lee et al., 2001; Lin et al., 2001). Recently, a new isoform (DAF-16d/f) was also reported to regulate *C. elegans* dauer formation. DAF-16d/f showed distinct tissue expression pattern and nuclear/cytosolic distribution pattern compared with DAF-16a and DAF-16c (Kwon et

al., 2010). The presence of other DAF-16 isoforms (e, g, and h) was also confirmed by cDNA analysis (www.wormbase.org) and their functions are under investigation. Multiple isoforms obviously favor the fine-tuning of biological processes mediated by ILS within the worms.

Many of the prevailing views about *C. elegans* DAF-16 are deduced from its analogy to other FOXOs (Figure 1.4). Dauer inducing conditions inactivate ILS and cause DAF-16 to enter the nucleus and bind to regulatory elements of the dauer-formation related genes as a transcriptional activator or repressor. Environmental cues signaling favorable growth conditions are sensed by *C. elegans* ciliated sensory neurons located in head and tail, and then converted into the internal tissue-specific stimuli (Ailion et al., 1999; Munoz & Riddle, 2003). Those stimuli activate ILS and trigger an intracellular phosphorylation cascade, starting with the activation of the receptor tyrosine kinase, DAF-2. This leads to recruitment of PI3K to the membrane and generation of signaling molecule phosphatidylinositol 3, 4, 5-triphosphate (PIP_3) from phosphatidylinositol 4,5-biphosphate (PIP_2) (Dorman et al., 1995). PIP_3 recruits and activates downstream protein kinase PDK1, which subsequently activates the protein kinases, AKT-1, AKT-2 and SGK-1, by phosphorylation (Paradis et al., 1999; Paradis & Ruvkun, 1998). Finally, activated AKTs and SGK-1 phosphorylate DAF-16 to create the binding sites for chaperone protein 14-3-3 (FTT-2), which promotes translocation of DAF-16 from the nucleus to the cytoplasm to suppresse its transcriptional activity, resulting in dauer exit and reproductive growth (Lin et al., 1997; Ogg et al., 1997; Wang et al., 2006; Cahill et al., 2001). DAF18/PTEN is capable of counterbalancing ILS by

specifically dephosphorylating the D3 position of PIP3, thereby preventing the series of downstream activations. (Ogg & Ruvkun, 1998; Machama & Dixon, 1998).

The first identified DAF-16 ortholog in a parasitic nematode was FKTF-1 from *Stongyloides stercoralis*, the threadworm of humans (Massey et al., 2003). *S. stercoralis fkth*-1 gene is organized much like *C. elegans daf*-16 gene, except with fewer and often smaller introns. Genomic and cDNA sequences predict that two isoforms, FKTF-1a and FKTF-1b, are present in *S. stercoralis*. Functional motifs common to *C. elegans* DAF-16, such as the forkhead DNA binding domain and Akt/PDB sites, are conserved in FKTF-1. Strong identity was shown in forkhead DNA binding domains of FKTF-1 and DAF-16. When expressed under the control of the *daf-16* promoter in *C. elegans daf-2 (e1370): daf-16(mg54)* double mutants, one of the FKTF-1 isoforms, FKTF-1b partially restored DAF-16 function (Massey et al., 2006). These data suggested that FKTF-1 and DAF-16 are orthologues and support the hypothesis that ILS in parasitic nematodes might control infective larval arrest and resumption of development similar to the regulation of the dauer development in *C. elegans*.

Regulatory mechanism of FOXO/DAF-16 transcriptional activity

Transcriptional activities of FOXO factors are negatively regulated by ILS through a shuttling mechanism that regulates FOXO translocation between nucleus and cytosol. The main regulator of FOXO transport is Akt/PKB kinase, which phosphorylates FOXOs at the specific sites and allows their interactions with 14-3-3 protein (Bruet et al., 1999). Binding of phosphorylated FOXO proteins to 14-3-3 protein sequesters them in

the cytoplasm presumably by masking FOXO protein's nuclear localization signals, preventing their re-entry to the nucleus (Carlsson & Mahlapuu, 2002).

14-3-3 proteins, highly conserved and abundant chaperone proteins within the cells, are involved in a broad range of biological processes. They belong to a group of scaffolding proteins that specifically bind to certain phospho-serine/threonine-containing proteins. The name refers to their elution location in a systematic attempt to classify bovine brain proteins by DEAE-cellulose chromatography and starch gel electrophoresis (Yaffe, 2002). 14-3-3 proteins have a molecular mass of approximately 30 kDa and function as a homo- or hetero- dimerized U-shaped structures (Liu et al., 2002). Peptide library screening led to identification of two optimal motifs that are recognized by 14-3-3 proteins, RSXpSXP and RXXXpSXP, where pS denotes a phosphorylated serine residue or a phosphorylated threonine residue. FOXO proteins all contain an optimal 14-3-3 binding site, overlapping the N-terminal Akt/PKB motif. Another conserved Akt/PKB site in the forkhead domain is also required for interacting with 14-3-3 proteins, as demonstrated by FKHRL1's role in cell survival (Brunet et al., 1999).

In addition to Akt/PKB kinase, SGK, CK1 (casein kinase), and DYRK1A (dual-specificity tyrosine-phosphorylated and regulated kinase), also phosphorylate FOXO proteins and promote it their nuclear exclusion within the cells (Brunet et al., 2001; Rena et al., 2002; Woods et al., 2001). However, interaction of 14-3-3 proteins with the phosphorylated sites from those kinases has generally not been investigated. Phosphorylation by JNK (Jun-N-terminal kinase) signaling and MST (mammalian Ste20-like) kinases signaling have the opposite effects on FOXO cellular location. Both JNK and MST family members are parts of a signaling transduction cascade that is activated

by exposure to environmental stresses. Activation of JNKs and MSTs has been shown to phosphorylate cytoplasmic FOXO and reverse its nuclear-cytoplasmic shuttling in response to ILS, resulting in enhanced nuclear localization of FOXO (Wang et al., 2005; Lehtinen et al., 2006).

Although nuclear-cytoplasmic shuttling of FOXO proteins is clearly necessary for their functioning, multiple lines of evidence indicate that it is not sufficient. Additional molecular mechanisms relevant to FOXO protein transcriptional activities include modulation of protein stability, modification of DNA-binding potential, and co-operation with associated proteins (Aoki et al., 2004; Zhang et al., 2002; Christian et al., 2002). Polyubiquitylation-dependent degradation of FOXO1 and FOXO3a proteins has recently emerged as an important event. It may occur in the nucleus or cytoplasm, coupled with acetylation-deacetylation or phosphorylation (Kitamura et al., 2005; Hu et al., 2004; Aoki et al., 2004; Huang et al., 2005). Direct disruption of FOXO protein binding to DNA has been linked to phosphorylation of Akt/PDB site as evidenced by gel shift and fluorescence anisotrophy experiments with a FOXO1 mutant in which Ser256 was replaced by negatively charged Asp (Zhang et al., 2002). A luciferase reporter assay also showed that FOXO1 transcriptional activity could be inhibited in the presence of insulin even though its nuclear exclusion was prevented by mutation (Tsai et al., 2003). Interactions between FOXO proteins and other associated proteins, such as nuclear hormone receptor, PGC-1α (proliferative-activated receptor-γ co-activator 1), and C/EBPβ (CCAAT/enhancer binding protein) depend on FOXO localization and possibly post-transcriptional modification status (Rochette-Egly, 2003; Pulgerver et al., 2003; Christian et al., 2002). FOXO proteins and co-activators might simultaneously bind to a

specific target gene sets so that signaling events can be integrated with the transcriptional machinery, or they might co-operate in a manner independent of DNA binding. Apparently, FOXO proteins serve as an endpoint of several signaling cascades and are regulated at different levels and in different cellular compartments.

In the case of *C. elegans*, functional fusion protein of DAF-16 and green fluorescence protein (GFP) has been constructed (*daf*-16::*gfp*) and used to study DAF-16 intracellular localization in a variety of genetic backgrounds (Lin et al., 2001; Hertweck et al., 2004). Findings from several *in vitro* experiments have established that DAF-16 physically interacts with 14-3-3 protein, SIR2.1, SGK-1, JNK-1, HCF-1 (host cell factor-1), and ubiquitin ligase, and provided considerable insights into the mechanism of DAF-16 function in worms (Wang et al., 2006; Berdichevsky et al., 2006; Oh et al., 2005; Li et al., 2007a; Li et al., 2008).

In wild-type worms, the DAF-16::GFP fusion protein is exclusively present in the cytoplasm and its presence there is dependent on the intact Akt / PKB sites. Upon interruption of the ILS pathway, such as in mutant strains [*daf-2 (e1370), pdk-1 (sa680), or akt-2(ok393)*] or in strains with gene-specific RNAi knock-down of *akt-1* or *sgk-1*, DAF-16 migrates to the nucleus and worms display the dauer phenotype. In the *daf-2 (e1370);daf-18 (mu397)* double mutant strain, nuclear localization of DAF-16 was blocked and dauer phenotypes were suppressed, consistent with the antagonistic effect of PTEN on the ILS pathway (Lin et al., 2001; Hertweck et al., 2004).

In *C. elegans*, there are two 14-3-3 encoding genes, *par5/ftt-1* and *ftt-2*, and their transcripts share ~78% sequence identity at nucleotide level and ~86% sequence identity at the amino acid level (Wang & Shakes, 1994; Wang & Shakes, 1996; Wang & Shakes,

1997). RNAi feeding experiments targeting unique fragments corresponding to the 3′-coding and UTR regions of either *ftt-2* or *par-5* indicated that FTT-2, but not PAR5, was able to regulate DAF-16 activities by forming a protein complex with DAF-16 and preventing its translocation from the nucleus to the cytoplasm (Li et al., 2007b).

SIR2 was originally isolated from *Saccharomyces cerevisiae* as a silent information regulator, and possessed a NAD (nicotinamide adenine dinucleotide) mediated histone deacetylase activity. In *C. elegans*, there are four SIR2 related genes present, and SIR2.1 was the most similar to yeast SIR2 due to its relatively high identity (31%) in the conserved core domain (Tissenbaum & Guarente, 2001). In transgenic worm lines over-expressing both SIR2.1 and DAF-16::GFP, co-immunoprecipitation using anti-SIR-2.1 antibody indicated that SIR-2.1 specifically bound to DAF-16 (Wang et al., 2006). However, others found that the interaction between SIR2.1 and DAF-16 only occurred during heat shock treatment and in 14-3-3 protein dependent manner (Berdichevsky et al., 2006). Whether SIR2.1 can directly deacetylate DAF-16 *in vivo* or *in vitro* is of great interest.

SGK-1 is the only open reading frame in *C. elegans* genome with strong sequence similarity to the human SGK. SGK-1 contains a kinase domain and an X domain with unknown function (Hertweck, et al., 2004). Transgenic worms expressing the *sgk-1::gfp* fusion construct indicated that SGK-1 is expressed in worm neurons and intestinal cells. Purified SGK-1 from wild type worms, but not *pdk-1(sa680)* mutant worms, was capable of phosphorylating DAF-16 *in vitro*, indicating this phosphorylation depends on the presence of PDK-1 and the activation status of SGK-1. *In vitro* pull-down assay and yeast two-hybrid screening showed that *C. elegans* SGK-1 was capable of interacting with

AKT-1 and AKT-2 to form a multimeric protein complex, which is a distinct characteristic never reported from its human counterpart (Hertweck, et al., 2004).

JNK-1 is one of the several JNK components that have been characterized in *C. elegans,* and was initially implicated in neuronal functions regulating coordinated locomotion (Kawasaki et al., 1999). Years later, an *in vitro* kinase assay using mammalian cell expressed *C. elegans* JNK-1 determined that DAF-16 serves as its phosphorylation substrate. Those phosphorylation sites reside only in the N-terminal region of DAF-16, and are presumably different from the Akt/PKB sites. In worm lines over-expressing JNK-1, enhanced nuclear translocation of DAF-16 was observed (Oh et al., 2005b), suggesting it affects the DAF-16 sub-cellular repositioning.

HCF-1 is an abundant nuclear transcriptional co-regulator that binds to chromatin and plays a critical role in both cell growth and division (Herr, 1998). In *C. elegans*, there is a single homologue of the human HCF with a much shorter open reading frame, and it is primarily expressed in embryos and adults (Liu et al., 1999). Immunostaining assays using anti-HCF-1 antibody showed that HCF-1 localizes to the nuclei in most, if not all, somatic and germline cells in wild type worms, and in all developmental stages. A specific HCF-1/ DAF-16 protein complex was detected in worms that lacked endogenous DAF-16 but carried a low-copy number *gfp::daf-16* transgene (*daf-16(mu86);muIs71*). Loss of *hcf-1* had significant effects on dauer formation, causing either no dauer phenotype or a weak dauer exit phenotype, and also led to increased recruitment of some DAF-16 to its target gene promoters and an altered gene expression profile (Li et al., 2008).

C. elegans potentially has hundreds of E3 ubiquitin-protein ligases (Kipreos, 2005). RLE-1 is the one that interacts with DAF-16 (Li et al., 2007). RLE-1 contains an N-terminal RING finger, a characteristic of E3 ubiquitin ligase, and has high homology with mammalian Roquin. Disruption of its expression results in partial dauer formation at 25°C and an increased level of DAF-16 protein, but not its mRNA, suggesting its involvement in worm development through regulating DAF-16 synthesis and turnover. RLE-1 might also regulate DAF-16 sub-cellular location, as evidenced by delayed DAF-16 redistribution in *rle-1* mutant worm when temperature was downshifted. Co-immunoprecipitation of DAF-16 from DAF-16 and RLE-1 co-transfected HEK293 cells indicated that there was direct interaction between DAF-16 and RLE-1, and confocal microscopy confirmed that these two proteins colocalized both in worms and in cultured cells. The interaction was completely abolished when the C-terminal 491 amino acids were deleted from RLE-1. An *in vivo* ubiquitin assay with or without the proteosome inhibitor MG132 demonstrated that an ubiquitylation mediated proteosome pathway was involved in DAF-16 degradation.

All of the above studies highlight a central role of DAF-16 at the interface of multiple signaling pathways and a high degree of conservation in its activity throughout metazoan phylogeny. DAF-16 orthologues have been identified from parasitic nematode *S. stercoralis*, and there is evidence indicating that DAF-16 orthologues are present in hookworms (Massey et al., 2003; Mitreva et al., 2005). Characterization of DAF-16s in parasitic nematode will ultimately contribute to our understanding the roles of DAF-16 in broader biological contexts.

Consensus binding sequence for DAF-16 and downstream gene targets

Genes are often flanked by distinct binding sites for different transcription factors so that efficient and selective transcription can be achieved (Berg et al., 2002). Therefore, there is great interest in characterizing transcription factor binding sites and their direct gene targets, as they may help to unravel the underlying molecular basis for transcription factor function.

DAF-16, categorized as an insulin-sensitive transcription factor, has been the focus of much attention because it acts as a signaling master influencing dynamic gene expression programs in various biological processes, and is one of the best-studied transcription factors in *C. elegans*. Early investigations demonstrated that DAF-16 and all other FOXO proteins directly interacted with insulin-responsive sequences (IRS), which are T/A rich and has a core sequence of T(G/A)TTT(T/G)(G/T), to mediate the regulatory effects of insulin on gene expression (Kops et al., 1999; Guo et al., 1999; Brunet et al., 1999; Paradis & Ruvkun, 1999). Later on, an *in vitro* iterative assay using degenerate oligonucleotides was performed using the known FOXO proteins, and after four rounds of selection/amplification, a slightly different 8-bp consensus sequence, TTGTTTAC, was identified as the consensus DAF-16 family protein element (DBE) (Furuyama et al., 2000). All four known FOXO proteins bound more strongly to this DBE than to IRS, and different FOXO proteins exhibited different preferences in the 5′ flanking sequence of the DBE. A luciferase reporter assay in NIH3T3 cells was used to test if the DBE was recognized by FOXO protein *in vivo.* The DBE was active, and the transcriptional activity was obviously much higher than the IRS. The DBE activity depended on copy number, and was negatively regulated by the serum stimulation. In the process of

searching the *C. elegans* genome database, one copy of the DBE was found upstream of the first exon of the *sod-3* gene, which encodes a mitochondrial Mn-superoxide dismutase (Mn-SOD), SOD-3. SOD-3 has been postulated to be a direct DAF-16 gene target, supporting a model that *sod-3* expression was regulated by DAF-16 through its direct interaction with its DBE (Furuyama et al., 2000; Honda & Honda, 1999).

The existence of a consensus DBE undoubtedly provided compelling evidence for the evolutionary conservation of DAF-16 functions. This knowledge also led to systematic prediction of putative DAF-16 targets in *C. elegans* using comparative genomics (Lee et al., 2003). A survey of 1 kb upstream of the start codon of *C. elegans* and *Drosophila melanogaster* genes identified thousands of genes that contained at least one perfect-match DBE, and among them, 17 genes were orthologous between *C. elegans* and *D. melanogaster*, highlighting a conserved group of gene targets in DAF-16 transcriptional cascade. The expression levels of each individual gene was examined in wild type, *daf-2 (e1370)*, and *daf-2 (e1370); daf-16 (mgDf47)* mutant worms, and six of the 17 genes (~35%) were found to be differently expressed in *daf-2 (e1370)* and *daf-2 (e1370); daf-16 (mgDf47)* mutant worms. This fraction was much higher than expected by chance. Both up-regulated and down-regulated targets were identified, suggesting DAF-16 could act as a transcriptional activator or transcriptional repressor. RNAi experiments showed 4 of the 17 genes significantly affected worm phenotypes. More DAF-16 downstream gene candidates were obtained when the informatic search was expanded to cover longer promoter regions. A similar search for genes with the DBE within putative promoter regions followed by comparison with human homologs exacted a list of 159 genes and one (*scl-1*) was studied in detail (Ookuma et al., 2003). *scl-1*

encodes a SCP-like secretory protein and its expression pattern was highly correlated with developmental stage of the worm and the expression of the components of the ILS pathway.

Expression microarray allows the parallel study of the expression of thousands of genes, and has succeeded in identifying DAF-16 downstream targets since completion of the *C. elegans* genome sequence. The data obtained so far include comparison between *daf-2* and *daf-2; daf-16* mutant worms, and comparison between *daf-2* RNAi and *daf-2* plus *daf-16 RNAi* (Murphy et al., 2003; McElwee et al., 2003; McElwee et al., 2004). Because normal DAF-2 functions to inhibit DAF-16 activity, those microarray comparisons were examining transcriptional differences between a *daf-16* active state (when *daf-2* was mutant) and a *daf-16* inactive state (when both *daf-2* and *daf-16* were mutant). The number of DAF-16 gene targets identified by microarray analysis was surprisingly large and those gene targets only partially overlapped between different microarray data sets. Homology-based functional analysis suggested that they represented diverse biological processes such as metabolism, development, stress responses, and longevity. However, when a subset of identified genes were subject to RNAi studies to test their functions, only a portion of them showed correlation between expression levels and predicted phenotypes (McElwee et al., 2003). In addition to functional information, regulatory information in promoters of the identified gene targets were also mined, and indicated that DBE was indeed over-represented. However, not all of the promoter regions of the gene targets contained the canonical DBE. A new over-represented motif, the DAF-16 associated element (DAE) with the sequence of CTTATCA, was found, suggesting the involvement of an unknown co-factor (Murphy et al., 2003).

Both bioinformatic analysis and microarray analysis extracted large number of DAF-16 candidate gene targets without distinguishing the direct ones from the indirect ones. To solve those problems, ChIP (chromatin immunoprecipitation) was used. ChIP combined with subsequent cloning is a powerful and relatively new approach designed to identify the direct *in vivo* gene targets of a transcription factor. The availability of anti-DAF-16 antiserum made ChIP feasible for identification of DAF-16 downstream gene targets (Oh et al., 2005a). DAF-16 and crosslinked genomic DNA was co-immunoprecipitated from *daf-2 (e1370)* mutant worms, which strongly sequestered DAF-16 in the nucleus and potentially provided an enriched pool of DAF-16 target genes. The DNA was recovered and sequenced. One hundred and thirty genes were identified, and among them, 88 contained a least one copy of either the DBE or DAE within a 5 kb region upstream, 15 had DBE/DAE in the coding region, and the remaining lacked either of the consensus sites. Only one of those 130 genes had been identified in previous studies, which was much lower than expected. Consistent with previous studies, those DBE/DAE containing gene targets were implicated in a variety of biological and physiological processes. Thirty-three genes were studied further for phenotype analysis using RNAi, and eighteen of them showed significant effects.

Another profiling based technique employed to identify DAF-16 gene targets was differential display PCR (DD-PCR) (Yu & Larsen, 2001). This technique was initially designed to compare the transcripts between *daf-2 (m41ts)* mutant worms and wild type worms, and a group of nine genes named *dao* were identified that displayed *daf-2* pathway-dependent expression patterns. Since *daf-16* suppresses *daf-2*, the expression level of the DAF-2 gene target in *daf2 (m41ts);daf-16 (m26)* double mutant worms was

expected to be different from the levels in *daf-2(m41ts)* mutant worms if it was also subject to the regulation by *daf-16*. RNA quantification showed that all of the identified *dao* genes except *dao-3* were regulated or partially regulated by *daf-16*.

Candidate gene approaches were also helpful in identification of DAF-16 target genes. Because the DAF-2/DAF-16 pathway is linked to worm responses to oxidative stress, heat, and heavy metals, genes that confer resistance to these stresses are good candidates for DAF-16 downstream targets. For examples, *sod*-3 (superoxide dismutase-3), *mtl*-1 (metallothionein-1), and *hsp*-16 (heat-shock protein-16) were all shown to be up-regulated in *daf*-2 mutants, but those increased levels were suppressed by introduction of the *daf-16* mutation into *daf-2* mutants (Honda & Honda, 1999; Barsyte et al., 2001; Walker et al., 2001).

The advances in identifying DAF-16 target genes highlight the power of a complete genome as a tool for gene discovery and gene function studies. This ever-improving genomic resource will continue to shape how the fundamental aspects of DAF-16 functions are interpreted.

Hookworm Genome Project

In the past decade, molecular parasitology has been revolutionized by increasing availability of genomic data from nearly 40 parasitic nematode species, including hookworms (McCarter et al., 2005). Inventories of hookworm genes have been created and are being expanded rapidly. Quantitative and comparative genomic analyses of different hookworm species and different developmental stages are now possible,

providing new insights into fundamental biology of this parasite and allowing connections to be established between hookworm and other parasites.

Like many other genomic studies, hookworm genomics began with the generation and analysis of expressed sequence tags (EST), which are derived from randomly selected transcribed sequences and have proven to be rapid and cost-effective tools for gene discovery (Boguski et al., 1994; Daub et al., 1999). The first survey of 259 clones selected from mixed adult *Necator americanus* cDNA library identified 161 new hookworm genes, and functions were assigned by homology for the majority of them. These genes included activation-associated secreted proteins (ASPs), small secreted effecter molecules, and many other functionally defined gene products. Since representation in a cDNA library generally correlated with abundance in the original biological sample, the EST clustering process also permitted preliminary abundance analysis even though the dataset was relatively small and the clones selected for sequencing was highly random. Tissue-specific EST survey was also used to identifiy hookworm gene products. Laser microdissection microscopy (LMM) enabled accurate extraction of gut tissue from adult *N. americanus* and *A. caninum*, and subsequent expression profiling analyses resulted in identification of nearly 300 transcripts. The predominant transcripts corresponded to mRNAs encoding cholesterol/fatty acid binding proteins, C-type lectins, ASPs, and proteases (Ranjit et al., 2006).

As increasing number of hookworm ESTs became available, especially for two animal-infective species, *A. caninum* and *A. ceylanicum*, a comparative analysis was made feasible (Mitreva et al., 2005). Analysis of nearly 20, 000 ESTs for transcripts originated from *A. caninum* and *A. ceylanicum* cDNA libraries representing pre-parasitic

and parasitic larval stages, as well as adults, showed that 80% of the genes had homologs in *C. elegans*. The coding sequences of these two species featured similar GC content, which was also consistent with that of *C. elegans*. The phylogenetic distance between *A. caninum* and *A. ceylanicum* is almost identical to the distance between *C. elegans* and *C. briggsae*. Transcripts with stage-specific expression representing the very different biological processes were present in different points of the lifecycle. For instance, *hsp12.6*, a dauer-enriched gene in *C. elegans*, was highly represented in infective L3 and tissue-arrested L3; cathepsin D-like aspartate protease was abundant in serum stimulated L3, consistent with its role in degradation of serum proteins and collagens; adult-specific proteins such as VIT-3 or chitinase-like protein were more likely to be involved in reproduction. Interestingly, ASP-1 cDNA was demonstrated to be abundant in both infective L3 and serum stimulated L3 even though the protein was noted as a marker for the latter only, supporting a hypothesis that arrested larvae were transcriptionally prepared at some level prior to infection.

A better understanding of hookworm biology and parasitism came with the advent of high throughput sequencing technologies (Wang et al., 2010). Over 1.5 million ESTs were generated from *A. caninum* infective L3, serum stimulated L3, and adult males and females, covering an estimated of 93% of the *A. caninum* transcriptome. Surprisingly, this near complete transcriptome analysis showed that the vast majority (77%) of the *A. caninum* transcripts were species-specific, and a significant number of transcripts shared homology with genes from *Brugia malayi,* which is a distantly related parasitic nematode species, demonstrating an impact of parasitism on the transcriptome. Only half of the *A. caninum* transcripts were expressed through all the stages, indicating a highly dynamic

gene expression profile. A dramatic difference was found in expression profiles between larvae and adults, but not between infective L3 and serum stimulated L3, suggesting exposure to serum stimulation did not trigger significant and immediate gene expression changes. Massive sequencing and transcriptome analysis were also conducted on 116,849 ESTs from adult stage of *N. americanus*, and similarly large percentage (67%) of species-specific transcripts were revealed. The most abundant proteins included proteins with WD40 repeats, protease inhibitors, and calcium-binding EF-hand proteins. The large number of newly predicted proteins also provided an attractive starting point for discovery of hookworm drug targets (Cantacessi et al., 2010). EST datasets also provide an excellent resource for characterization of codon usage preference in hookworm, which serves as a useful tool for open reading frame prediction and evolutionary analysis (Mitreva et al., 2006)

Following the generation and analyses of a large number of EST data was a shift toward the full-scale hookworm genome project to include the sequence information for non-transcribed and intergenic regions (Abubucker et al., 2008). Hookworm has a substantially large genome, compared with other clade V strongylid species, estimated to be 347.2 Mb, as measured by flow cytometry. A published hookworm genome draft spanning 57.6 Mb was assembled from 104,000 genome survey sequences (GSS) with ~27% as repetitive elements, and approximately 5500 protein coding genes were identified. Direct comparison of the *A. caninum* genes with the corresponding *C. elegans* homologs indicated that the median exon size and median intron size were smaller in *A. caninum*. As many other eukaryotic genomes, mobile genetic elements (MGEs), such as *transib* elements, *dingo* elements, and *bandit* elements, are present in the hookworm

genome (Kapitonov & Jurka, 2005; Laha et al., 2006; Laha et al., 2007). MGEs have implications in genome evolution and might also offer potential applications as transgenesis and gene silencing vectors.

The complete *C. elegans* genome has effectively accelerated the already rapid characterization of individual genes, molecular pathways and developmental mechanisms in this non-parasitic nematode. A complete hookworm genome is expected to advance the knowledge of hookworm research community and serve as an ongoing reference for the development of sustainable hookworm control strategies.

Dissertation outlines

Accumulating evidence suggests that the ILS pathway is involved in hookworm infection and development processes, and a DAF-16 ortholog from hookworm has been identified. These findings have led to the hypothesis that the hookworm DAF-16 downstream gene network is critical for resumption of development during infection, and therefore might represent targets for therapeutic intervention in the hookworm life cycle. In this dissertation, the transcriptional activity of hookworm DAF-16/FOXO is investigated, several of its downstream gene targets are identified, and a *cis*-regulatory element mediating the effects of hookworm DAF-16/FOXO is characterized.

Chapter 1 of this dissertation has provided a survey of the scientific literature relevant to hookworm disease, nematode development, and current views about the ILS pathway and its key transcription factor, DAF-16.

The molecular characterization of hookworm DAF-16 orthologs is presented in Chapter 2. Analysis of the amino acid sequence of hookworm DAF-16 showed that it was

closely related to *C. elegans* DAF-16 and contained the conserved functional domains common to all forkhead transcription factors. Our *in vitro* binding and *in vivo* reporter gene assays suggested hookworm DAF-16 acted as a transcription factor and drove gene expression through the consensus DBE. The data in Chapter 2 has been published in the peer-reviewed journal, *International Journal for Parasitology* (Gao et al., 2009).

Identification of the hookworm DAF-16 binding sites, response elements, and target genes in genomic DNA is presented in Chapter 3**.** In *C. elegans*, identification of DAF-16 target genes has been the focus of much attention because they mediate the formation of dauer stage, determine longevity, and regulate fatty acid metabolism. Therefore, hookworm DAF-16 was hypothesized to be critical for the transition to parasitism and the resumption of the development. We identified hookworm DAF-16 binding sites, response elements and target genes by a genomic selection method. Hookworm DAF-16 binding elements were examined for the presence of the consensus DBE and other over-represented motifs and linked target genes were analyzed. The information about those important genes and their functional profiles helped to define biological processes regulated by hookworm DAF-16. The data in Chapter 3 has been published in the peer-reviewed journal *PLoS One* (Gao et al., 2010).

Chapter 4 investigates the gene structure of a hookworm DAF-16 downstream gene target, hookworm *snr*-3, and its expression profile in different post infection developmental stages. Real time PCR was be used to measure changes in the expression level of hookworm *snr-3* over different stages, and suggested that it contributed to hookworm maturation into adulthood. The reporter assays suggested that the 3′-UTR of

hookworm *snr*-3 was implicated in regulation of gene expression mediated by hookworm DAF-16.

In chapter 5, the last chapter, the overall findings of the dissertation research and their significance are discussed.

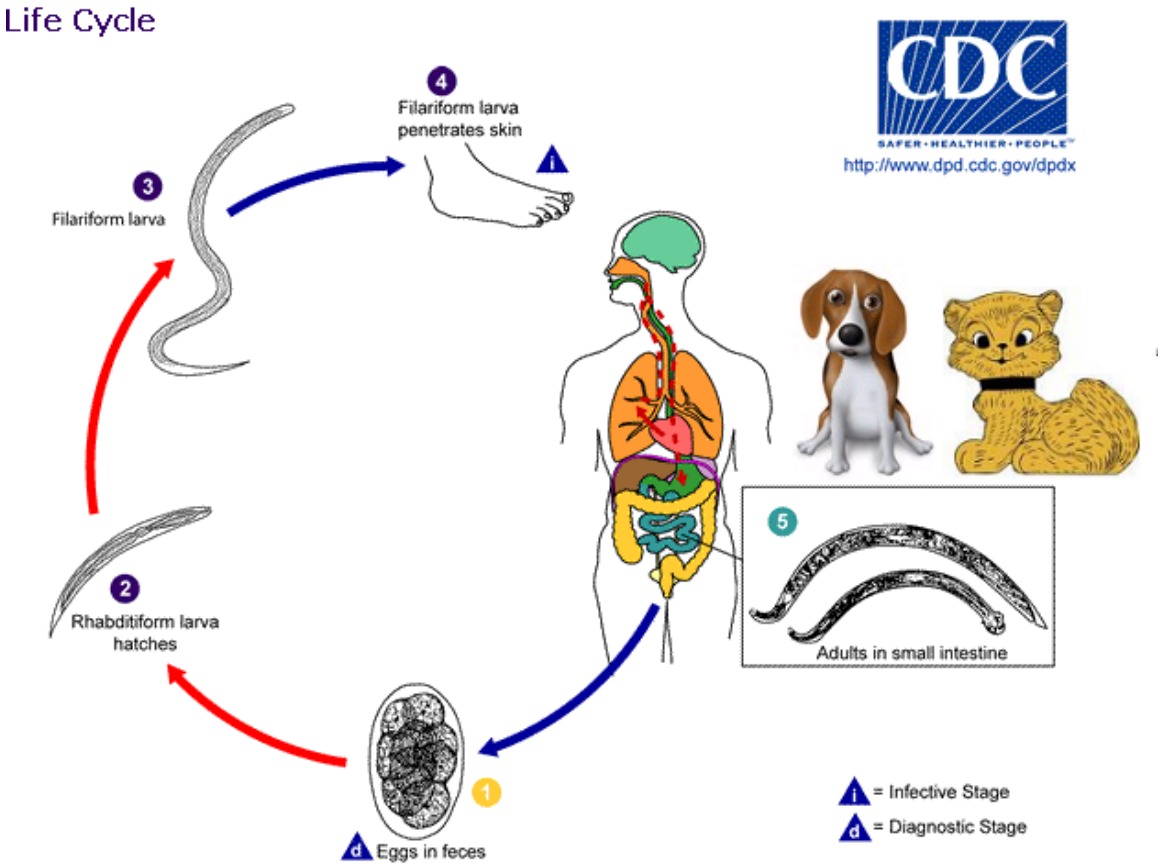


Figure 1.1 Hookworm life cycle. Figure modified from CDC website (http://www.dpd.cdc.gov/dpdx/html/hookworm.htm). Eggs are passed in the feces1, and under favorable conditions (moisture, warmth, shade), larvae hatch in 1 to 2 days. The released rhabditiform L1/L2 grow in the feces and/or the soil2, and after 5 to 10 days (and two molts) they become infective L33. On contact with the host, the L3 larvae may be ingested or penetrate the skin4. The larvae reach the small intestine, where they reside and mature into adults. Adult worms live in the lumen of the small intestine, where they attach to the intestinal wall with resultant blood loss by the host 5.

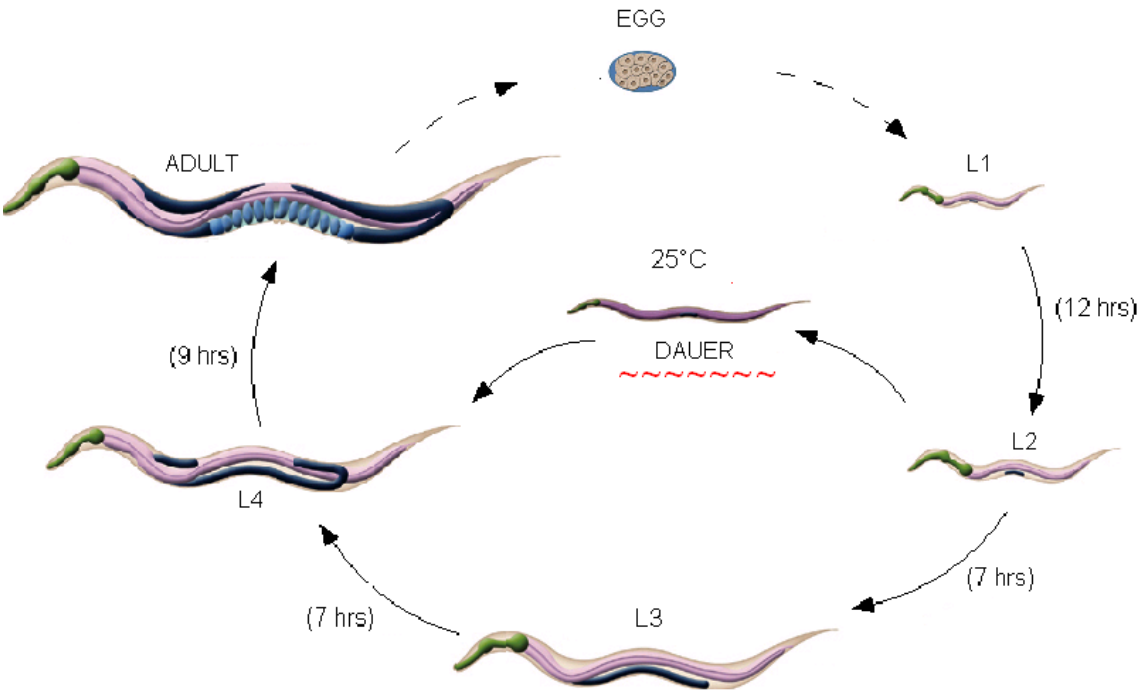


Figure 1.2 *C. elegans* life cycle. Numbers along the arrows indicate the length of time the animal spends at a certain stage. Figure modified from WORMATLAS website (http: //www.wormatlas.org/hermaphrodite/introduction/mainframe.htm).

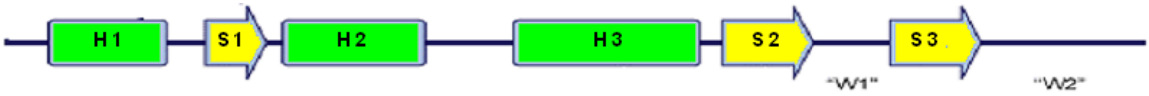


Figure 1.3 Schematic representation of secondary structure elements within the forkhead domain of FOXO proteins. α-helices (H1, H2, H3) and β-strands (S1, S2, S3) are highly conserved; wing regions (W1, W2) are variable.

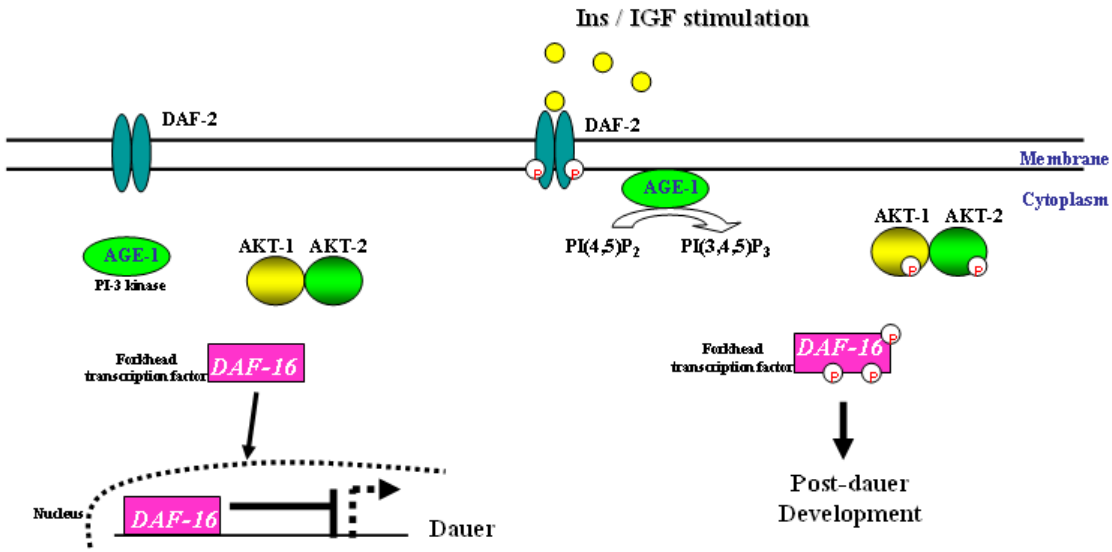


Figure 1.4 Schematic diagram of insulin/IGF-1 signaling pathway in *C. elegans* dauer formation and exit. Dauer inducing conditions cause DAF-16 to enter the nucleus and binds to the regulatory elements of specific genes. ILS stimulates the pathway in worms by a phosphorylation cascade.. The culmination of these phosphorylation events results in translocation of DAF-16 from the nucleus to the cytoplasm and insulin-induced dauer recovery.

Chapter 2. Molecular cloning and DNA binding characterization of DAF-16 orthologs from *Ancylostoma* hookworms

Xin Gao, Daniel Frank, and John M. Hawdon

Department of Microbiology and Tropical Medicine, The George Washington University Medical Center, Washington, D.C.

Abstract

Third-stage infective larvae (L3) of hookworms encounter a host-specific signal during infection that re-initiates suspended developmental pathways, resulting in development to the adult stage. This resumption of development in the host is analogous to recovery of developmentally arrested *Caenorhabditis elegans* dauer larvae in response to favorable environmental signals. Dauer recovery in *C. elegans* dauer and hookworm L3s is mediated by insulin-like signaling (ILS). A key output of ILS in *C. elegans* is the forkhead transcription factor DAF-16, which controls the expression of genes required for maintenance of the dauer stage. The similarity between recovery pathways of L3 and dauers suggests that DAF-16 functions similarly in hookworm L3 activation. To test this, orthologs of *Ce*-DAF-16 were isolated from the hookworms *Ancylostoma caninum* and *A. ceylanicum*. The protein sequences of hookworm DAF-16 DNA binding domains were identical, and shared 94% identity with the b and c isoforms of *Ce*-DAF-16. *Ac*-DAF-16 expressed in HEK293 cells bound strongly to the conserved DAF family binding element (DBE), but much more weakly to a similar sequence found upstream of a developmentally regulated hookworm gene. *Ac*-DAF-16 was able to drive transcription of a reporter gene located downstream of six copies of the DBE (6x DBE) in NIH3T3 cells under starved conditions. Addition of serum caused a decrease in reporter gene expression, indicating that DAF-16 is negatively regulated by growth factor stimulation. These data confirm the presence of DAF-16 orthologs in hookworms, and demonstrate that *Ac*-DAF-16 binds to and drives transcription from a conserved DAF-16 family DNA binding element.

Key words: *Ancylostoma*, hookworm, insulin signaling, dauer, forkhead, transition to parasitism,

1. Introduction

Hookworms infect an estimated 740 million people, mostly in tropical and subtropical regions, making it among the most common infectious diseases worldwide (Bethony et al., 2006). Heavy infection with hookworm can create severe health issues for newborns, growing children, pregnant women and malnourished persons (Lozoff et al., 1991; Bundy et al., 1995; Dreyfuss et al., 2000; Bethony et al., 2002). The critical step for hookworm development and survival is to successfully establish a parasitic relationship. The infective form of hookworm is a free-living third-stage larva (L3) that remains developmentally arrested in the environment until it enters a definitive host. In this host, it encounters host-specific signals that initiate development and maturation to adulthood (Rogers and Petronijevic, 1982; Hawdon and Schad, 1991a). An *in vitro* system using the canine hookworm *Ancylostoma caninum* has allowed the investigation of the early events of reactivation of hookworm L3s (Hawdon and Schad, 1990; Hawdon and Hotez, 1996). This system uses the resumption of feeding by infective L3 as a marker for activation (Hawdon and Hotez, 1996). Using this system, several molecules that are released during the infectious process have been identified (Hawdon et al., 1995b; Hawdon et al., 1996; Hawdon et al., 1999). More recently the systems was used to demonstrate that cGMP signaling is involved in larval activation (Hawdon and Datu, 2003) However, very little is

known about the molecular mechanisms underlying this transition to parasitism in large part due to the complexity of hookworm life cycles and the lack of genetic tools.

The free-living nematode *Caenorhabditis elegans* has served as a useful model for understanding the mechanisms of development and genetics of hookworm larval stages. *C. elegans* adopts a non-feeding, developmentally arrested dauer stage under conditions of overcrowding and starvation (Cassada and Russell, 1975). The dauer stage is remarkably similar to infective L3s of hookworm in terms of behavior, morphology and function, except that development into L3 is obligatory for hookworm (Rogers and Sommerville, 1963; Hawdon and Schad, 1991). Exit from the dauer stage in *C. elegans* in response to improved environmental signals is hypothesized to be analogous to the reactivation of the arrested hookworm L3 in response to the host-specific signals during infection (Rogers and Sommerville, 1963; Hawdon and Schad, 1991; Hotez et al., 1993). Studying the processes of signaling transduction in *C. elegans* has yielded important information about the development of hookworm.

Insulin-like signaling (ILS) has been implied in regulation of *C. elegans* dauer arrest and recovery by epistasis analysis coupled with cloning of genes related to dauer formation (Gottlieb and Ruvkun, 1994; Morris et al., 1996; Kimura et al., 1997; Ogg et al., 1997; Ogg and Ruvkun, 1998; Paradis and Ruvkun, 1998; Paradis et al., 1999). A key target of the ILS signaling pathway is a forkhead/FOXO transcription factor DAF-16, which is located in the nucleus under dauer inducing conditions, where it binds to insulin response sequences in the promoters of target genes and causes entry into dauer (Lin et al., 1997; Ogg et al., 1997). Phosphorylation by the activated serine-threonine protein kinase Akt/protein kinase B (Akt/PKB) under permissive growth conditions results in its nuclear

exclusion and cytoplasmic retention of DAF-16, leading to reproductive growth (Brownawell et al., 2001; Cahill et al., 2001). An 8-bp consensus DAF-16 family member binding element (DBE) has been identified that binds both *C. elegans* and mammalian forkhead transcription factors (Furuyama et al., 2000).

Involvement of an ILS pathway in hookworm L3 activation was supported by experiments with specific inhibitors for PI3 kinase and Akt/PKB kinase (Brand and Hawdon, 2004). In the present study, *C. elegans* DAF-16 orthologs were cloned and sequenced from *A. caninum* and *A. ceylanicum*, two close relatives of human hookworm, and shown to bind to and drive transcription from a conserved binding element found in the promoters of FOXO/DAF-16 target genes in *C. elegans* and mammals. These data represents the first identification of a molecule involved in ILS from a hookworm, further support a role for insulin signaling in the hookworm activation, and suggest that DAF-16 plays a critical role in gene expression associated with the transition to parasitism.

2. Materials and methods

2.1 Nomenclature

Molecules are named in accordance to the guidelines established by the Society of Nematology (Bird and Riddle, 1994). The *C. elegans* name for the forkhead transcription factor DAF-16 was retained to indicate orthology. The gene names and DNA sequences are written in lowercase italics, with the species of origin as a prefix (*Ac-daf-16*). When referring to the protein, the name is written in uppercase letters (*Ac*-DAF-16).

2.2 Parasites

The Baltimore strain of *A .caninum* maintained as previously described in beagles (Schad and Page, 1982) and *A. ceylanicum* was maintained in Syrian hamsters (Garside and Behnke, 1989). Animals were housed and treated in accordance with institutional care and use committee guidelines. Infective L_3 are recovered from coproculture by a modified Baermann technique and frozen until used. Adults were collected from intestines of infected animals following euthanasia, and frozen at -80 °C until nucleic acids were isolated.

2.3. Cloning and characterization of hookworm DAF-16

An *A. ceylanicum* (*Ae*) expressed sequence tag (pk84h05) (Mitreva et al., 2005) with similarity to the *C. elegans* forkhead transcription factor DAF-16 was used to design specific forward and reverse oligonucleotide primers for PCR. The *Ae-daf-16* cDNA ends were amplified in two separate hemi-nested PCR reactions. The 3′ end was amplified from an *A. ceylanicum* L3 directional cDNA library constructed in Lambda ZAP II (Hawdon et al., 1995a), with the nested forward primers DAF16-F1 (5′ -ACAATCTCTGAGTGCCGT GCACC-3′) and DAF16-F2 (5′ -CATTGACATACACACAT CCATCCAC-3′) together with a primer complementary to the T7 promoter sequence flanking the 3′ end of the cDNA insert site in the pBluescript vector. In the first round PCR, DAF16-F1, T7, and the cDNA library were incubated with *Taq* polymerase (USB, Cleveland, OH) for 35 cycles of 1 min at 94° C, 1 min at 57° C, 2 min at 72° C, followed by a final extension for 5 min at 72° C. One microlitre of the first round reaction was used as template for the second round PCR, containing the reverse nested primer DAF16-F2 and T7, and incubated with

JumpStart Taq polymerase (Sigma, St. Louis) for 30 cycles as above except the annealing temperature was lowered to 55° C.

The 5′ end of *daf-16* was amplified from first strand cDNA synthesized from infective L3 of *A. ceylanicum* by standard methods (Hawdon et al., 2003). The nested reverse primers DAF16-R1 (5′ -CTGCGACTGATTACGGACACTGC-3′) and DAF16-R2 (5′ -GACAGGCGATTCCGTCTTGATTGG-3′) and the forward conserved nematode spliced leader primer SL1 (5′ -GGTTTAATTACCCAAGTTTGAG-3′) (Bektesh et al., 1988; Hawdon et al., 1995a) were used in successive reactions. In the first PCR, DAF16-R1, SL1, and cDNA were incubated with JumpStart *Taq* for 2 min at 95 C, followed by 30 cycles of 1 min at 94° C, 1 min at 55° C, 2 min at 72° C, followed by a final extension for 5 min at 72° C. One microlitre of the first round reaction was used as template for the second round PCR, containing the reverse nested primer DAF16-R2 and SL, and incubated as above. Amplicons of approximately 850 bp and 400-bp from the 5′ and 3′ reactions, respectively, were gel purified and cloned into pGEM-T Easy vector (Promega, Madison, WI), and the DNA sequences determined (see below).

To complete the sequences, the section of the *daf-16* cDNA connecting the 5′ and 3′ ends was amplified. Primers DAF16-F3 (5′ -GAAACCAATCAAGA CGGAATCGC-3′) and DAF16-R3 (5′ -TGCACGGCACTCAGAGATTGTCG-3′) were incubated with L3 cDNA library (1 μL) and *Taq* polymerase for 35 cycles of 1 min at 94° C, 1 min at 55° C, 2 min at 72° C, followed by a final extension for 5 min at 72° C. An amplicon of approximately 350 bp was cloned in pGEM and sequenced as described above.

The oligonucleotide primers designed from the *Ae-daf-16* sequence were used to amplify the full-length coding sequence of the *A. caninum daf-16* cDNA. Primers

DAF16pic-FX (5′ -ATCCTCGAGCATAATGTCTATGTCGAATCACCTCTCGCAA ACG-3′) and DAF16pic-RX (5′ -TCATTCTAGAAGATTGTCGAAATTCAGGCG-3′) were incubated together with *A. caninum* L3 first-strand cDNA and *Taq* polymerase for 35 cycles of 1 min at 94° C, 1 min at 55° C, 4 min at 72° C, followed by a final extension for 4 min at 72° C. An amplicon of approximately 1.8 kb was gel purified, cloned in pGEM and sequenced.

2.4. Stage specificity

Ancylostoma caninum eggs were isolated from infected dog feces, and *A. ceylanicum* eggs from infected hamster feces, by sucrose floatation (Nolan et al., 1994), axenized by treatment with NaOCl, and plated on NGM (nematode growth medium) agar plates (Sulston and Hodgkin, 1988). Following incubation at 28 C for 16-18 h, the hatchlings (L_1) were washed from the plates with BU (50 mM Na_2HPO_4/22 mM KH_2PO_4/70 mM NaCl, pH 6.8) (Hawdon and Schad, 1991b) and snap-frozen in a dry ice/ethanol bath.

Reverse transcriptase (RT)-PCR was performed to determine the stage specificity of *Ac-daf-16* expression. Total RNA was isolated from *A.caninum* eggs, L1/L2, L3, and adult stages by homogenization in TRIzol Reagent according to manufacturer's instructions (GibcoBRL). First-strand cDNA was synthesized from approximately 1 µg of total RNA using the Superscript First-Strand Synthesis System (GibcoBRL). Primers spanning a 360 bp intron were chosen so that products amplified from contaminating genomic DNA could be distinguished from those amplified from the stage-specific cDNA. PCR reactions contained gene specific primers DAF16-F3 and DAF16-R3 and 1 µl of first

strand cDNA from each *A. caninum* developmental stage. These primers are predicted to generate a product of ~ 384 bp from cDNA. Primers specific for the 3' untranslated region of the catalytic subunit of protein kinase A (Hawdon et al., 1995a), which is expressed in all stages, were used in positive control reactions.

2.5. Bacterial expression and antiserum production

A 492 bp fragment corresponding to amino acids 365 to 528 was amplified from a cDNA clone of *Ac-daf-16* for expression in bacteria. Primers DAF16E-FB (5′ - TTAGGATCCGA GGAAGCCATG CAAGCATCATTCG-3′) and DAF16E-RX (5′ - TATACTCGAGAGTAAT CAGTGGTGTATGTGTCGGCCACGG-3′), containing the 5′ restriction sites *Bam*HI and *Xho*I, respectively (underlined), were incubated in a PCR containing plasmid DNA and *Pfu* Ultra polymerase (Stratagene, La Jolla, CA) for 2 min at 95 C, followed by 30 cycles of 30 sec at 95 C, 30 sec at 55 C, 1 min at 72, and a final extension for 4 min at 72 C. The 492 bp amplicon was digested with *Bam*HI and *Xho*I (New England Biolabs Beverly, Massachusetts), ligated into the bacterial expression vector pET28a (Novagen, San Diego, CA) cut with the same restriction enzymes, and transformed into DH5α by standard methods. Clones containing *Ac-daf-16*$_{365\text{-}528}$ were confirmed by DNA sequencing. For expression, plasmid DNA was transformed into *E. coli* BL-21 (DE3) competent cells, and protein expression induced by the addition of 1 mM isopropyl-beta-D-thio-galacto-pyranoside (IPTG) to log-phase bacterial cultures. An aliquot of cells was removed from the culture prior to induction to serve as a pre-induction control, and from an un-induced culture grown under the same conditions. The induced cultures were grown for 3 h at 37° C with continued shaking at 220 revolutions per min.

Aliquots (1 ml) were withdrawn at 0, 1, 2, and 3 h of induction, and soluble and insoluble fractions of the lysates analyzed on 12% SDS-PAGE as described (Hawdon et al., 1999). Recombinant DAF-$16_{365-528}$ (rDAF-$16_{365-528}$), which was expressed in an insoluble form, was solubilized by the addition of 1% SDS and 0.5% 2-mercaptoethanol, and purified as described elsewhere (Hawdon et al., 1996).

Polyclonal antiserum was raised in New Zealand white rabbits by immunization with purified rDAF-$16_{365-528}$. Approximately 150 μg of rDAF-$16_{365-528}$ were injected subcutaneously after emulsification with Freund's complete adjuvant. Additional boosts with recombinant protein (150 μg each) in incomplete Freund's adjuvant were administered at 3, 6, and 9 weeks.

2.6. Cell culture, expression of recombinant Ac-DAF-16 and antiserum characterization

HEK293T cells were maintained in 25-cm^2 flasks in RPMI1640 supplemented with 10% fetal bovine serum (Biowhittaker, Walkersville, MD), 100 I.U. /ml penicillin (Mediatech, Herndon, VA), 100 μg/ml streptomycin (Mediatech), and 2mM L-glutamine (Mediatech), at 37°C in an atmosphere of 5% CO_2. The entire coding region of *Ac*-DAF-*16* was cloned into the mammalian expression vector pCMV-TAG4 (Stratagene) at the *Xho* I and *Bam*HI restriction sites to generate pCMV4-*daf*-16 construct containing an in-frame C-terminal FLAG octapeptide tag (DYKDDDDK). The construct was confirmed by DNA sequencing. HEK293T cells grown in 6-well dishes were transfected with 4 μg of pCMV4-*daf*-16 using Metafectene (Bionex, Munich) according to manufacturer's instructions, and cultured for 48 hrs. Empty vector controls received 4 μg pCMV without an insert, and mock-treated cells received no DNA. The cells were disrupted in lysis

buffer (20 mM Tris-HCl (pH7.5), 50 mM NaCl, 5 mM 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), 1% Triton X-100, 0.075% SDS), and the protein concentration determined by micro-bicinchoninic acid (BCA) assay (Pierce, Rockford, IL). Samples containing equal amounts of protein were separated by electrophoresis through 4-20% Tris-glycine pre-cast Novex gradient gels (Invitrogen, Carlsbad, CA) and transferred to PVDF membrane (Immobilon-P, Millipore, Billerica, MA). Membranes were blocked overnight with 5% nonfat milk in phosphate buffered saline containing 0.2% Tween 20 (PBST) and then incubated with anti-FLAG antibody M2 (1:2000 dilution, Sigma) or DAF16 antiserum (1:20,000) for 2 hrs at 22 °C. Membranes were washed twice with PBST and probed with the corresponding secondary antibodies conjugated with horseradish peroxidase (anti-mouse IgG, Sigma; anti-rabbit IgG, Boehringer Mannheim) at room temperature for 1 hr. Lastly, the membranes were washed several times with PBST and visualized using the ECL detection kit (Pierce).

2.7. Streptavidin bead pull-down assay

The following 5'-end biotin labeled forward oligonucleotides and their unlabeled complements were synthesized by Integrated DNA Technologies (Coralville, IA): forward DBE oligonucleotide (5′-/5Bio/TGCCAGTGCG GGTTGTTTACGCGTTTTCAAGA-3′), reverse DBE oligonucleotide (5′-AAAATCTTGAAAACGCGTAAACAACCCGCACT-3′), forward DAO binding element oligonucleotide (5′-/5Bio/CACTCCACCGT ATTGTTAATCGTTG ATGTAGT-3′), reverse DAO binding element oligonucleotide (5′-ACTACATCAAC GATTAACAATACGGTGGAGTG-3′), forward random oligonucleotide (5′-/5Bio/CATACGGGCGCATG AATTACCGCAAGCACAGG-3′),

reverse random oligonucleotide (5′-CCTGTGCT TGCGGTAATTCATGCGCCC GTATG-3′). Forward and reverse oligonucleotides were annealed to form double-stranded (ds) DNA and used in the pull-down assays. Concentrated complementary oligonucleotides (200uM forward and reverse DBE oligonucleotides, 200uM forward and reverse random oligonucleotides, 130 uM forward and reverse DAO-DBE oligonucleotides) were mixed at 1:1 molar ratio in microcentrifuge tube and annealed in a thermocycler. The oligonucleotides were incubated for 1 min at 95 ºC, and then the incubation temperature was gradually reduced (-1 ºC/ min) for 70 cycles, followed by a final hold at 4 ºC. Annealed oligonucleotides were stored at -20ºC. Twenty micrograms of HEK293T cell lysate was incubated with 100 pmoles biotin-labeled DBE dsDNA, 65 pmoles biotin-labeled DAO dsDNA, or 100 pmoles biotin-labeled random dsDNA in binding buffer (10 mM Tris pH7.5, 50 mM KCl, 1 mM DTT, 2.5% glycerol, 5 mM $MgCl_2$, 50 ng/ μL poly (dI·dC), 0.05% Non-idet P-40) at 4°C for 2 h. After incubation, 25 μl of 30% Streptavidin-sepharose (GE Healthcare) slurry equilibrated with TNE_{50} buffer (10 mM Tris, 50mM NaCl, 1mM EDTA, pH7.5) + 0.1% Nonidet P-40 were added to the mixture and incubated for 2 h at 4°C. Beads were collected by centrifugation at 2500*g* and washed twice with TNE_{100} buffer (10 mM Tris, 100 mM NaCl, 1 mM EDTA pH7.5) + 0.1% Nonidet P-40. The bound proteins were separated on a 4-20% SDS-PAGE and Western blotted with DAF16 antiserum.

2.8. Transfections of NIH3T3 cells and reporter assays

Reporter vector p6xDBE-luc, a derivative of the pGL3-basic firefly luciferase vector (Promega) containing 6 concatemerized copies of DAF-16 family DBE (Furuyama

et al., 2000), was a gift from Professor B.M. Burgering (University Medical Centre Utrecht, Urecht, The Netherlands). NIH3T3 cells grown in 96-well plates were transfected with 50 ng of reporter vector, 10 ng of pRL-TK-*Renella* luciferase vector (Promega) and 10 or 40 ng of pCMV4-DAF-16 clone 2J using Metafectene. Empty pcDNA3.1-V5 expression vector was included as a negative control, and to equalize the total amount of DNA transfected to 120 ng in co-transfections. For serum-response analysis, media was changed to 20% (v/v) bovine calf serum (ATCC, Manassas, VA) or serum-free media at 24 h after transfection. The relative luciferase activity (firefly luciferase activity divided by the *Renilla* luciferase activity) was determined 24 h later using the Dual Luciferase Assay System (Promega). To determine the effect of the PI3K inhibition on transcriptional activity, cells were pre-incubated for 2 h with 50 μM of the specific PI3K inhibitor LY294002 (Calbiochem, San Diego, CA) in a final concentration of 0.05% DMSO. After 2 h, the cells were either washed and incubated with 20% serum above, or 20% serum and 50 μM LY294002.

2.8. Sequence Analysis

Dideoxy DNA sequencing using the ABI BigDye Terminator Cycle Sequencing Ready Reaction Kit v3.1 was performed by the Nevada Genomics Center, and the reactions run on an ABI3730 DNA Analyzer. Sequence analysis was performed using BioEdit Sequence Alignment Editor version 5.0.9 (Hall, 1999), and multiple alignments performed using CLUSTAL W at Swiss EMBnet (http://www.ch.embnet.org/software/ClustalW.html). Neighbor-joining trees were constructed using MEGA version 3.1 (Kumar et al., 2004). Homology searches were done using BLASTP (Altschul et al., 1990;

Altschul et al., 1997) at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/; BLOSUM) and WU-BLAST Parasite Genome Database Query at the European Bioinformatics Institute (http://www.ebi.ac.uk/blast2/parasites.html). Percentage identity and similarity was determined using the BLASTP algorithm with the BLOSUM62 matrix. Conserved motifs were identified by searching Scansite at MIT (http://scansite.mit.edu/) using MotifScan software (Obenauer et al., 2003).

3. Results

3.1. Characterization of hookworm DAF-16 molecules

An *A. ceylanicum* expressed sequence tag (EST) sequence with homology to the *daf-16* cDNA from *C. elegans* was used to design oligonucleotide primers to isolate the orthologous cDNAs from the hookworms *A. ceylanicum* and *A. caninum*. The 3′ end of *Ae-daf-16* was isolated from an *A. ceylanicum* L3 cDNA library using a hemi-nested PCR strategy. The second reaction, using a nested gene-specific primer and the T7 promoter sequence present in the vector flanking the cloning site, produced an amplicon of approximately 400 bp. DNA sequencing of this clone revealed that it contained the 3′ end of *Ae-daf-16*, including the poly d(A) tail, and that it overlapped the EST sequence. A similar hemi-nested strategy was used to isolate the 5′ end of *Ae-daf-16* from *A. ceylanicum* L3 first-strand cDNA. In this case, the 5′ primer was the conserved nematode spliced leader (SL) sequence, which has been used to identify several hookworm 5′ ends, together with two nested gene-specific reverse primers. The second reaction produced an

amplicon of approximately 850 bp that contained the SL sequence and overlapped the 5′ end of the *A. ceylanicum* EST sequence. New primers were designed to amplify the entire coding sequence of *Ae*- and *Ac-daf-16* cDNAs in a single piece, which were cloned into pGEM T Easy vector and confirmed by DNA sequencing.

The full-length *Ae-daf-16* cDNA contained the conserved 22 nucleotide nematode spliced leader at the 5′ end (Bektesh and Hirsh, 1988), which was followed by 5′ - untranslated nucleotides and an ATG codon encoding the starting methionine at nucleotide 29. The cDNA encodes an open reading frame of 584 amino acids, ending with a TGA termination codon at nucleotide 1781. This is followed by a 429 nucleotide 3′ untranslated region (UTR) containing a rare AATAAT polyadenylation signal 12 nucleotides upstream of the poly (dA) tail (nucleotides 2204 to 2209). This signal is found in only 1% *C. elegans* of cDNA 3′ ends (Blumenthal and Steward, 1997; Hajarnavis et al., 2004).

Both *A. ceylanicum* and *A. caninum daf-16* cDNA sequences are predicted to encode proteins of 584 amino acids. The hookworm *daf-16* cDNA sequences of the open reading frame are 94% identical, and the deduced amino acid sequences are nearly the same (98% identical, 99% similar). Of the 9 amino acid differences, only two were conserved. The *Ae*-DAF-16 deduced protein has a predicted mass of 62,664 Da and a calculated pI of 6.28, while the *Ac*-DAF-16 protein is similar, with a mass of 62,693 Da and a pI of 6.17. The N-terminus of the hookworm DAF-16 deduced amino acid sequences lack hydrophobic signal peptides (Bendtsen et al., 2004), consistent with its role as a transcription factor.

DAF-16 is a transcription factor characterized by a forkhead or "winged helix" DNA binding domain. This binding domain is unique, and is found in proteins involved in determining developmental fates (Hacker et al., 1992). The *A. ceylanicum* and *A. caninum* forkhead domain deduced amino acid sequences are identical. A search of GenBank (BLASTP algorithm) (Altschul et al., 1997) using the *Ac*-DAF-16 amino acid sequence confirmed the presence of a forkhead or "winged helix" transcription factor domain (architecture cd00059.3; Interpro IPR001766) from amino acids trp_{216} to val_{308}. As predicted, the best matches from the Genbank search were isoforms of *Ce*-DAF-16. There are three major DAF-16 isoforms in the *C. elegans* genome, which have recently undergone changes in their letter designations between the published and Wormbase versions (http://www.wormbase.org, release WS178). Isoforms "b" (formerly "a2") and "c" (formerly "a1") differ by only two amino acids, whereas *Ce*-DAF-16a (formerly "b" isoform) is 530 amino acids (Ogg et al., 1997). The N-terminal 191 amino acids of *Ce*-DAF-16 b and c isoforms and 211 amino acids of *Ce*-DAF-16a are encoded by different exons and are only 32% identical, whereas the 319 C-terminal residues are from the same exons, and therefore nearly identical in all isoforms (Lee et al., 2001). Pairwise alignments indicate that the hookworm DAF-16 molecules are most closely related to *C. elegans* b (46% identity/59% similarity) and c (47/59) isoforms, and more distantly related to the *Strongyloides stercoralis* FKTF isoforms a (26/41) and b (27/43).

As expected based on database searching and alignments, phylogenetic analysis using aligned forkhead domain amino acid sequences grouped the hookworm DAF-16 proteins most closely with *Ce*-DAF-16b, *Ce*-DAF-16c, and *Ss*-FKTF-1a (Figure 2.1A). More distantly related were *Ce*-DAF-16a and *Ss*-FKTF-1b. The relationship is readily

apparent from an alignment of the forkhead domains (Figure 2.1B). The hookworm forkhead domain is 94% identical and 98% similar to *Ce*-DAF-16 b and c isoforms, and 87% identical and 91% similar to *Ss*-FKTF-1a, but only 66% identical and 89% similar to the *Ce*-DAF-16a isoform and 63% identical and 80% similar to *Ss*-FKTF-1b. This level of similarity suggests that the hookworm DAF-16 proteins represent orthologs of *Ce*-DAF-16.

At 584 amino acids, the hookworm DAF-16 proteins are larger than the *C. elegans* isoforms, which range from 508 to 530 amino acids, as well as *Ss*-FKTF-1b (566 amino acids), but not *Ss*-FKTF-1a (741 amino acids). Hookworm DAF-16 proteins have an N-terminal extension of approximately 56 amino acids, and an alanine-rich 17 amino acid insert following the forkhead domain that is absent from all of the other nematode DAF-16 isoforms. Ten of the 17 amino acids are alanines. Like *Ss*-FKTF-1a and *Ce*-DAF-16b and c, the N-terminal domain of the hookworm proteins is serine-rich, containing 38 serine residues out of 216 (17.6%) (Massey et al., 2003). The is a predicted nuclear export signal (NES) conforming to the canonical LXXLXL sequence (la Cour et al., 2004) at Leu_{548} to Leu_{553}. NES sequences are common in proteins that are shuttled between the nucleus and cytoplasm, such as transcription factors (la Cour et al., 2004).

The serine/threonine kinase Akt/PKB phosphorylates *Ce*-DAF-16 in response to ILS signaling (Paradis and Ruvkun, 1998; Brunet et al., 1999). There are four potential Akt phosphorylation sites in *Ce*-DAF-16 that conform to the conserved RXRXXS/T motif (Alessi et al., 1996): Thr_{54}, Ser_{240}, Thr_{242}, and Ser_{314} (isoform c numbering) (Cahill et al., 2001). At least three of them are phosphorylated by Akt (Cahill et al., 2001). MotifScan software at Scansite (http://scansite.mit.edu/) identified three potential Akt

phosphorylation sites in *Ac*-DAF-16, at Ser_{107}, Thr_{312}, and Ser_{381}. There is also a serine (Ser_{311}) adjacent to Thr_{312} that is not in the proper Akt context, and therefore may not be phosphorylated. In the *C. elegans* and *S. stercoralis* molecules, the analogous serine and threonine residues are separated by a single amino acid, and both are located in conserved Akt phosphorylation motifs.

In *C. elegans*, phosphorylation at the Thr_{54} and Ser_{314} sites was shown to be required for binding of a 14-3-3 protein that negatively regulates *Ce*-DAF-16 by promoting its exclusion from the nucleus (Cahill et al., 2001; Henderson and Johnson, 2001). The Ser_{381} site in *Ac*-DAF-16 is probably homologous to Ser_{314} of *Ce*-DAF-16, whereas Ser_{107} is most likely equivalent to the Thr_{54} of *Ce*-DAF-16. Interestingly, all of the *Ac*-DAF-16 Akt phosphorylation sites are located in consensus 14-3-3 binding sites (Muslin et al., 1996), suggesting that the hookworm DAF-16 molecules are regulated by a similar mechanism as *Ce*-DAF-16. It remains to be determined which of these sites are phosphorylated in *Ac*-DAF-16.

3.2 Stage specificity by RT-PCR

PCR was performed on first-strand cDNA from egg, L1/L2, L3, and adult *A. caninum* life cycle stages to determine when the *Ac-daf-16* as expressed during the hookworm life cycle. The *daf-16* primers were designed to span an intron, thereby amplifying a larger band if genomic DNA contaminated the cDNA preparation (Figure 2.2B). Primers that amplify a portion of the 3′-untranslated region of the protein kinase A catalytic subunit (Hawdon et al., 1995a) were used as a positive control. As shown in Figure 2.2A, both the control primers and the *daf-16* primers amplified products of the

appropriate size (~380 bp) from all life cycle stages, indicating that *Ac-daf-16* mRNA is transcribed throughout hookworm development.

3.3. Ac-DAF-16 polyclonal antiserum

A 163 amino acid peptide from the C-terminal portion of *Ac*-DAF-16 was expressed as a polyhistidine fusion protein in *E. coli*, purified, and used to make a polyclonal antiserum in a rabbit. A western blot using this antiserum is shown in Figure 2.3A. The DAF-16 antiserum recognizes a protein of approximately 62 kD in *A. caninum* L_3 soluble extracts. A second, smaller band of approximately 37 kD is recognized weakly, and may represent an as yet undefined cross-reacting isoform similar to *Ce*-DAF-16e (33 kD).

3.4. Ac-DAF-16 expression in HEK293T cells

Active, properly folded recombinant *Ac*-DAF-16 protein was required for functional studies. Attempts to express soluble r*Ac*-DAF-16 in the yeast *Pichia pastoris* were unsuccessful (data not shown), so we used mammalian cell culture for *in vitro* expression. The mammalian expression construct pCMV4-*daf*-16 was verified by sequencing the inserted DNA fragment. The full length coding region of 1782 bp pCMV4-*daf-16* encodes a 584-amino acid fusion protein of *Ac*-DAF-16 with a FLAG tag at the C-terminus. The calculated molecular mass of the fusion protein is 63933.4 Da. Two sequence-verified clones (Daf16.4Ba.pCMV.2J and Daf16.4Ba.pCMV.1D) were used for transient transfection of HEK293T cells. Transfected, empty vector transfected and non-transfected (mock-treated) cell lysates were separated by SDS-PAGE and blotted

to PVDF membranes for Western blots. An immunoreactive band corresponding to the predicted molecular weight (64 kDa) of the expressed fusion protein was detected by both anti-FLAG antibody M2 (Figure 2.3B) and *Ac*-DAF-16 antiserum (Figure. 2.3C) in the transfected cells, but not in the mock-treated cells or the cells transfected with empty vector.

To determine the titer of anti-*Ac*-DAF-16 antiserum, Western blots of cell lysates expressing *Ac*-DAF-16 were probed with increasing dilutions of the polyclonal antiserum. The antiserum recognized *Ac*-DAF-16 as dilutions as low as 1:30,000 (not shown).

3.5. rAc-DAF-16 DBE binding and transcriptional activation

The consensus DBE is located in the promoter region of DAF-16/forkhead-regulated genes. It was identified by a binding/PCR amplification/selection strategy, and is conserved between mammals and *C. elegans* (Furuyama et al., 2000). A closely related sequence, which differed from the DBE at two positions, was found upstream of *Ac-dao-1* gene, which encodes the ortholog of *C. elegans* FKBP-like protein that is up-regulated during hookworm L3 activation (DAO-BE). We hypothesized that *Ac*-DAF-16 would function like *Ce*-DAF-16 and other forkhead transcription factors by binding to the DBE and driving transcription of the downstream gene. To determine if *Ac*-DAF-16 binds to consensus DBE, oligomers containing the DBE (5′-TTGTTTAC-3′), DAO-BE (5′-TTGTTAAT-3′) and a random sequence were synthesized with biotin label at the 5′-end. The biotin-labeled oligomers were incubated with soluble HEK293T cell lysates, followed by streptavidin beads to pull-down biotin-labeled oligonucleotide/ protein complexes. Recombinant *Ac*-DAF-16 bound strongly to the biotinylated DBE oligo, as indicated by

recognition of several bands of approximately 64 kDa by anti-DAF-16 antiserum in the pull-down assay containing DBE and pCMV4-*daf*-16 transfected cell lysate (Figure 2.4, lane 2). The DAF-16 antiserum weakly recognized the same bands in the pull-down assay containing DAO-BE and pCMV4-*daf*-16 transfected cell lysate (Figure 2.4, lane 5). The bands were absent in the pull-down assays containing the random oligomer and *Ac*-DAF16 transfected cell lysates or mock cell lysate (Figure 2.4, lanes 8 and 9), DBE and mock cell lysate (lane 3), or DAO-BE and mock cell lysate (lane 6). These pull-down assay results suggest that *Ac*-DAF-16 recognizes and binds strongly to the conserved DBE, but much more weakly to the DAO-BE.

To determine whether *Ac*-DAF-16 could drive transcription from the DBE promoter element in vivo, pCMV4-*daf*-16 and the luciferase reporter assay p6xDBE-luc were co-transfected into NIH3T3 cells together with a *Renilla* luciferase reporter plasmid. In starved NIH3T3 cells, DAF-16 expression activated transcription of the DBE by fivefold (in cells transfected with 10 ng of pCMV4-*daf*-16) or ninefold (40 ng of pCMV4-*daf*-16) (Figure 2.5). Treatment of transfected cells with 20% serum significantly inhibited transcription from the DBE. This indicates that *Ac*-DAF-16 can bind to and drive expression from the *C. elegans* DBE when they are co-expressed in mammalian cells. Furthermore, endogenous cell signaling pathways are able to negatively regulate transcriptional activity by *Ac*-DAF-16 in response to serum in NIH3T3 cells. However, treatment with the specific PI3K inhibitor LY294002 (Vlahos et al., 1994) had no effect on the serum-stimulated negative regulation of transcription (not shown).

4. Discussion

Developmentally arrested hookworm infective L3 encounter a host-specific signal during invasion of their definitive host that re-activates suspended developmental programs. These programs culminate in maturation to the parasitic adult stage. Timely and proper response to this signal is critical for the successful establishment of a parasitic relationship. Failure to initiate development in the correct host would represent a missed infection opportunity, and resumption of development in a non-permissive host would result in either direct death of the parasite, or "reproductive death" due to isolation from potential mates. Therefore, proper execution of the events of this “transition to parasitism” is critical.

The molecular signaling events associated with the transition to parasitism in hookworms are only now beginning to be deciphered, primarily because of insights about nematode development gained from studies of *C. elegans*. Exit from the developmentally arrested dauer stage in *C. elegans* has been compared with the transition to parasitism that occurs when free-living infective stages enter their definitive host (Rogers and Sommerville, 1963; Hawdon and Schad, 1991a; Hotez et al., 1993; Hawdon and Hotez, 1996). In *C. elegans*, recovery from dauer requires ILS, but not TGF-β, signaling (Tissenbaum et al., 2000). Using pharmacological agents and an in vitro feeding assay, we have shown previously that hookworm L3 use an analogous ILS pathway during larval activation (Tissenbaum et al., 2000; Hawdon and Datu, 2003; Brand and Hawdon, 2004). Given this conserved signaling pathway, we hypothesized that a hookworm ortholog of the *C. elegans* transcription factor DAF-16 was the terminus of ILS signaling during

activation. To investigate this hypothesis, we have cloned a DAF-16 ortholog from the hookworms *A. caninum* and *A. ceylanicum*, and expressed recombinant *Ac*-DAF-16 in mammalian cells to investigate its function in larval activation.

An *A. ceylanicum* expressed sequence tag (pk84h05) with similarity to *Ce-daf-16* was used to design gene specific oligonucleotide primers for PCR. The 5' and 3' ends of *Ae-daf-16* cDNA were amplified in separate reactions, and a second set of primers design from those sequences to complete the entire sequence. Oligonucleotide primers designed from the complete *Ae-daf-16* sequence were used to amplify the full-length coding sequence of *daf-16* from *A. caninum* cDNA. The deduced amino acid sequence was used to search the non-redundant GenBank database, and found to be a member of the forkhead family of transcription factors, characterized by a forkhead or "winged helix" DNA binding domain. Originally identified in *Drosophila*, this binding domain is found in proteins involved in determining developmental fates (Hacker et al., 1992). In *C. elegans*, DAF-16 controls longevity and resistance to stress, as well as dauer development (Lin et al., 1997; Ogg et al., 1997; Henderson and Johnson, 2001). Phylogenetic analysis of the binding domains indicates that the hookworm DAF-16 molecules are most closely related to isoforms b and c of *C.elegans*, as well as FKTF-1a of the parasitic nematode *Strongyloides stercoralis* (Massey et al., 2003). Interestingly, *Ss*-FKTF-1b, but not *Ss*-FKTF-1a, partially rescued the dauer defective phenotype of *daf-2; daf-16* double mutants when expressed in *C. elegans* (Massey et al., 2006). The *C. elegans* DAF-16 isoforms appear functionally interchangeable (Lee et al., 2001), so differences in biological effects are likely due to differential tissue expression. However, Estevez et al (Estevez et al., 2006) suggested that the *C. elegans* isoforms a and b, which differ in their DNA binding

domains, bind to different parts of the tryptophan hydroxylase *(tph-1*) promoter and have opposite effects on *tph-1* expression. We are currently seeking additional isoforms in hookworms.

There are also several sequence differences between the isoforms outside the DNA binding domain that may alter as yet undetected regulatory interactions. For example, the hookworm DAF-16 molecules contain a serine-rich N-terminal domain, found also in *Ss*-FKTF-1a, *Ce*-DAF16b and c, but not in *Ce*-DAF-16a *Ss*-FKTF-1b, or the mammalian orthologs (Massey et al., 2003). Also, the hookworm DAF-16s contain a unique alanine-rich sequence in the C-terminal domain missing from the other orthologs. The functional significance of these sequences is unknown.

In *C. elegans*, DAF-16 regulates aging and development. When environmental conditions are inhospitable for development, ILS is inactive, and DAF-16 enters the nucleus and binds to DNA, resulting in expression of genes required for the dauer stage. When conditions improve, the serine/ threonine kinase Akt/PKB phosphorylates *Ce*-DAF-16 in response to ILS though the insulin receptor DAF-2 (Paradis and Ruvkun, 1998; Brunet et al., 1999). Phosphorylation creates binding sites for 14-3-3 proteins, which escort DAF-16 from the nucleus and sequester it in the cytoplasm (Cahill et al., 2001; Li et al., 2007). The hookworm DAF-16 orthologs contain three serine or threonine residues in the proper context for AKT phosphorylation (Ser_{107}, Thr_{312}, and Ser_{381}), all of which are located in consensus14-3-3 binding sites (Muslin et al., 1996). This also suggests a role for 14-3-3 proteins in the regulation of hookworm DAF-16 (Cahill et al., 2001; Li et al., 2007). *Ac*-DAF-16 is expressed in all developmental stages, similar to *C. elegans* DAF-16 and *S. stercoralis* FKTF (Ogg et al., 1997; Lee et al., 2001; Massey et al., 2003),

consistent with their regulation by phosphorylation and cellular relocation, rather than at the level of transcription.

The near identity of the hookworm DAF-16 DNA binding domain to that of *Ce*-DAF-16b suggested that *Ac*-DAF-16 might bind to the same DNA sequence. *Ce*-DAF-16 and mammalian orthologs bind, via the forkhead domain, to a specific consensus DAF-16 family binding element, or DBE, with the sequence 5′-TTGTTTAC-3′ (Furuyama et al., 2000). The core sequence of the DBE is TGTT, and a change of any base in TGTT to C eliminated the binding affinity of DAF-16 homologues. Using streptavidin bead pull-down assay, we demonstrated that recombinant *Ac*-DAF-16 expressed in mammalian cells strongly bound to the conserved DBE. The multiple bands seen on the Western blot may represent differentially phosphorylated forms of *Ac*-DAF-16. Recombinant *Ac*-DAF-16 also bound weakly to a DBE-like sequence, TTGTTAAT, found 72 bp upstream of the start codon of the hookworm gene *Ac-dao-1*. This gene encodes the likely ortholog of *C. elegans* FKBP-like protein, and is up-regulated during hookworm L3 activation (not shown). This putative DAO-BE differs by 2 nucleotides in the flanking region of DBE sequence, which probably accounts for its weak binding to *Ac*-DAF-16. We have yet to test whether this sequence is sufficient to drive expression of the *dao-1* gene, or whether the gene is regulated by *Ac*-DAF-16. There are no other DBE sequences in the approximately 170 bp of upstream *Ac-dao-1* sequence we identified, but it is possible that a DBE is located further upstream.

Given its ability to bind to the DBE, we next tested whether *Ac*-DAF-16 could drive transcription of a reporter gene from this binding element. A reporter plasmid (p6xDBE) containing 6 consecutive DBE sequences cloned upstream of the firefly

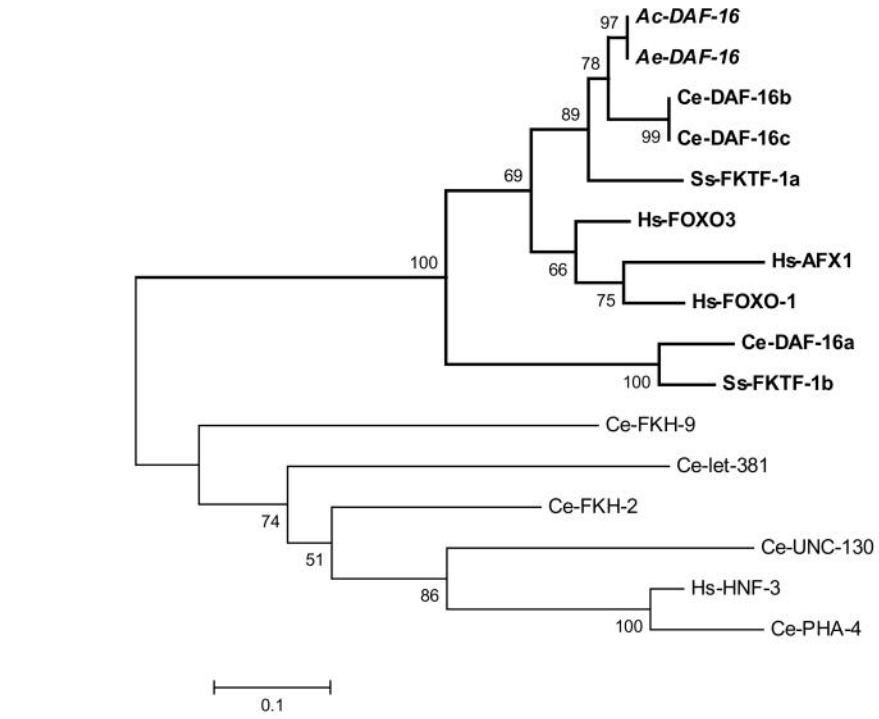
luciferase gene (Furuyama et al., 2000) was co-transfected with pCMV4-*daf*-16:FLAG into NIH3T3 cells. Under starved conditions, increasing amounts of *Ac*-DAF-16 increased transcription of the reporter gene, as determined by luciferase assay. Addition of serum inhibited transcription to control levels. While these results were expected, we were unable to block inhibition of transcription with the specific PI3K inhibitor LY294002 (Vlahos et al., 1994). This suggests that while endogenous serum-activated signaling pathways were able to negatively regulate *Ac*-DAF-16 transcriptional activity, it is unlikely to require ILS in this cell line. AKT kinase, which phosphorylates DAF-16, can be activated by pathways other than ILS in response to serum growth factors (Van Der Heide et al., 2004). For example, the serum and glucocorticoid regulated kinase (SGK) recognizes the same sequence motif as Akt (Alessi, 2001), and also phosphorylates mammalian FoxO factors (Van Der Heide et al., 2004) . Investigations of *Ac*-DAF-16 function in other cell lines are ongoing.

In summary, we have identified an ortholog of *C. elegans* DAF-16 from the hookworms *A. caninum* and *A. ceylanicum*. Recombinant *Ac*-DAF-16 binds to a conserved DNA binding element, and can drive transcription from this promoter sequence in mammalian cells. Transcription is negatively regulated by serum in this system, indicating that *Ac*-DAF-16 functions similarly to *Ce*-DAF-16. Given the similarity with the *C. elegans* dauer, genes that are controlled by DAF-16 are probably important for the maintenance of the developmentally arrested "dauer" state in hookworms, and consequently represent an interesting area of investigation. Identification and subsequent manipulation of these genes may provide novel avenues for intervention in the hookworm

life cycle. The interaction between *Ac*-DAF-16 and DBE will provide a valuable tool in the search for target genes regulated by DAF-16.

Note: Nucleotide sequence data reported in this paper are available in the GenBank™, EMBL and DDBJ databases under the accession number **EU635467** and **EU635468.**

A.


Ac-DAF-16
Ae-DAF-16
Ce-DAF-16b
Ce-DAF-16c
Ss-FKTF-1a
Hs-FOXO3
Hs-AFX1
Hs-FOXO-1
Ce-DAF-16a
Ss-FKTF-1b
Ce-FKH-9
Ce-let-381
Ce-FKH-2
Ce-UNC-130
Hs-HNF-3
Ce-PHA-4
97
78
89
99
69
100
66
75
100
74
51
86
100
0.1

B.

Ac-DAF-16 1 NAWGNLSYADLITQAIMQSPEKRLTLSQVYEWMVQNVPYFRDKGDSNSSAGWKNSIRHNL
Ce-DAF-16b 1 NAWGNMSYAELITTAIMASPEKRLTLAQVYEWMVQNVPYFRDKGDSNSSAGWKNSIRHNL
Ss-FKTF-1a 1 NAWGSHSYADLITQAIQSSPEQRLTLSQVYEWMVTNVPFFRDKGDSNSSAGWKNSIRHNL
Ce-DAF-16a 1 NPWGEESYSDIIAKALESAPDGRLKLNEIYQWFSDNIPYFGERSSPEEAAGWKNSIRHNL
Ss-FKTF-1b 1 NPWGEESYSDLIAKALECAPDKRMKLNEIYQWFSENIPYFNDRSSQEEAAGWKNSIRHNL
Ac-DAF-16 61 SLHSRFMRIQNEGAGKSSWWVINPDAKPGR
Ce-DAF-16b 61 SLHSRFMRIQNEGAGKSSWWVINPDAKPGR
Ss-FKTF-1a 61 SLHNRFMRIQNEGAGKSSWWVINPDAKNGR
Ce-DAF-16a 61 SLHSRFMRIQNEGAGKSSWWVINPDAKPGR
Ss-FKTF-1b 61 SLHNRFMRIQNEGAGKSSWWVINPDAKNGR

Figure 2.1

Figure 2.1 The phylogenetic relationship between nematode DAF-16 and selected forkhead protein family members. A. Neighbor joining tree of forkhead domains. Forkhead domains were aligned using CLUSTAL W software on the Swiss EMBnet server (http://www.ch.embnet.org). Amino acid distances were calculated using the Poisson correction model in the MEGA program version 3.1 (Kumar et al., 2004). Bootstrap values (1000 replications) are shown at each node. *Ac, A. caninum; Ae, A. ceylanicum; Ce, Caenorhabditis elegans; Ss, Strongyloides stercoralis; Hs, Homo sapiens*. Accession numbers: *Ce*-DAF-16a, AAK82917; *Ce*-DAF-16b, AAK82918; *Ce*-DAF-16c, AAK82919; *Ss*-FKTF-1a, AAQ23177; *Ss*-FKTF-1b, AAQ23178; *Hs*-AFX1, P98177; *Hs*-FOXO-1, NP_002006; *Ce*-FKH-9, AAA81139; *Ce*-FKH-2, AAA82436; *Hs*-HNF-3, AAD51979; *Ce*-UNC-130, CAA88935; *Hs*-FOXO3, AAH68552; *Ce*-let-381, AAB37792; *Ce*-PHA-4, CAB07378. B. Alignment of representative nematode forkhead domains. Shading indicates residues that are identical (black) or similar (gray) to *Ac*-DAF-16. Protein sequences were aligned using CLUSTAL W software and displayed using BOXSHADE software located on the Swiss EMBnet server. The *A. ceylanicum* and *C. elegans* isoform c were omitted from the alignment because the forkhead domains are identical to *A. caninum* and *C. elegans* b isoforms, respectively.

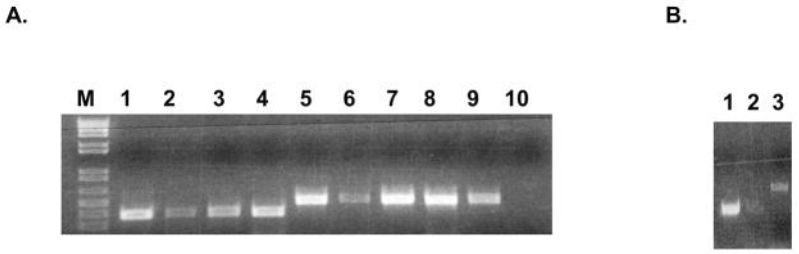


Figure 2.2 Expression of *Ac-daf-16* mRNA during the hookworm life cycle. A. First strand cDNA from *Ancylostoma caninum* life cycle stages was amplified with the gene specific primers DAF16-F3 and DAF16-R3 (lanes 1-4) or control primers PKA5'-4 and PKA3'-1 (lanes 5-10) and separated on a 1.5% agarose gel. Lanes 1 and 5 = egg cDNA; lanes 2 and 6 = L1/L2 cDNA; lanes 3 and 7 = L3 cDNA; lanes 4 and 8 = adult cDNA; lane 9, genomic DNA; lane 10, no DNA control. B. Primers DAF16-F3 and DAF16-R3 amplify a larger amplicon from genomic DNA. Lane 1, *Ac-daf-16* plasmid DNA; lane 2, *A. caninum* L3 cDNA; lane 3, *A. caninum* genomic DNA.

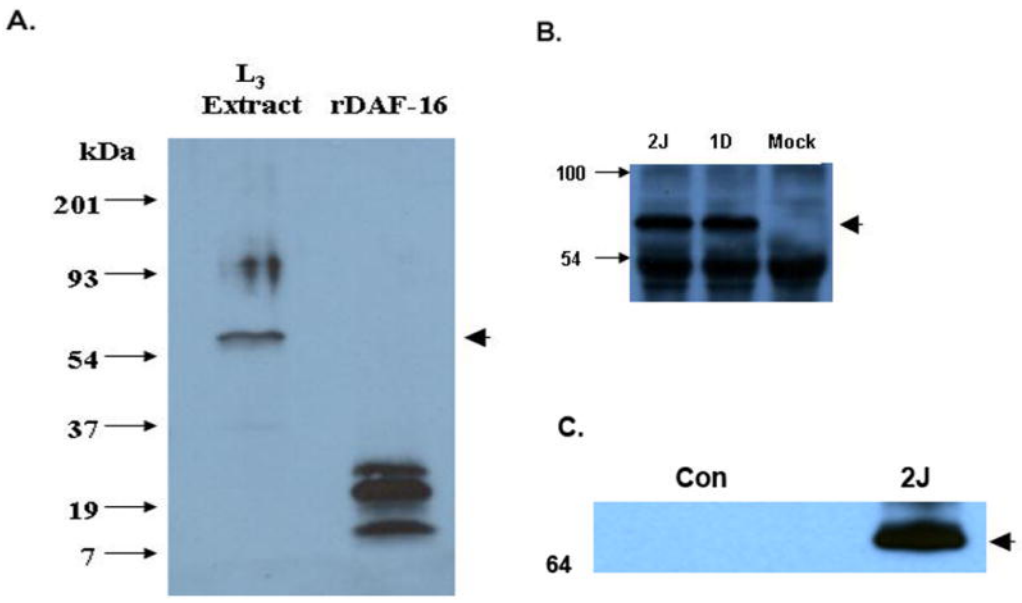


Figure 2.3 Daf-16 expression in mammalian cells and antiserum characterization. The arrowheads indicate the position of *Ac*-DAF-16. A. Western blot of *A. caninum* L3 extract and recombinant DAF-16 C-terminus (rDAF-$16_{365-528}$) expressed in *E. coli*. Proteins were detected with *Ac*-DAF-16 polyclonal antiserum. B. Expression of *Ac*-DAF-16 in HEK293T cells. Lanes 2J and 1D contain lysates of HEK293T cells transfected with clones Daf16.4Ba.pCMV.2J and Daf16.4Ba.pCMV.1D. The mock lane contains lysates from non-transfected cells. The blot was probed with anti-FLAG (M-2) antibody. C. Western blot of HEK293T cells expressing recombinant *Ac*-DAF-16. The blot was probed with a 1:20,000 dilution of *Ac*-DAF16 antiserum. M, mock transfected; T, transfected with Daf16.4Ba.pCMV.2J.

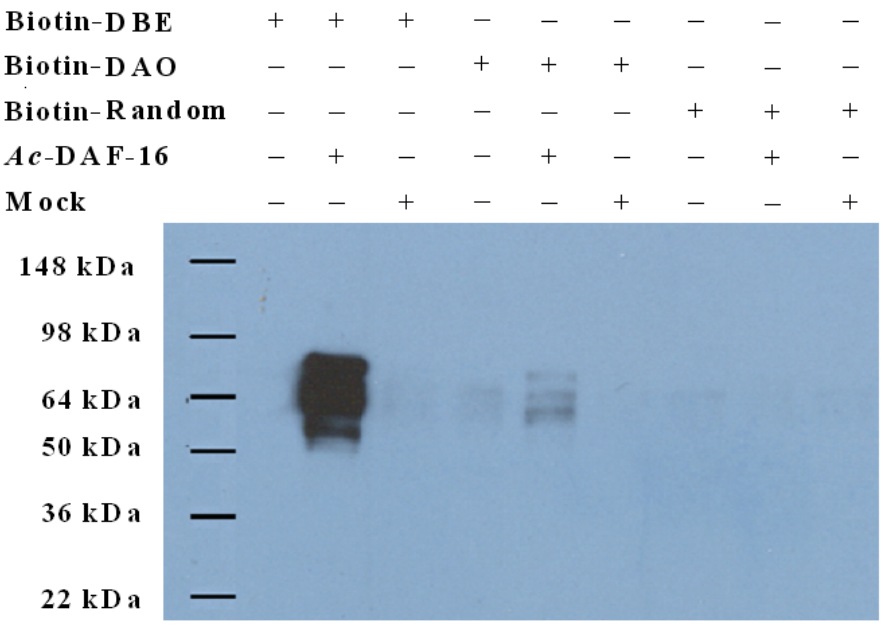


Figure 2.4 Streptavidin bead pull-down to detect DBE binding activity of recombinant *Ac*-DAF-16. Biotin labeled double strand DNA oligonucleotides were incubated with cell lysates from pCMV4-Daf16 transfected HEK293T cells or untransfected (mock) cells, and DAF-16-oligo complexes pulled down with streptavidin conjugated Sepharose beads. Complexes were detected by Western blotting with DAF-16 polyclonal antiserum. DBE, DAF-16 binding element; DAO, DBE-like element found upstream of the *Ac-dao-1* start site; random, random sequence oligonucleotide.

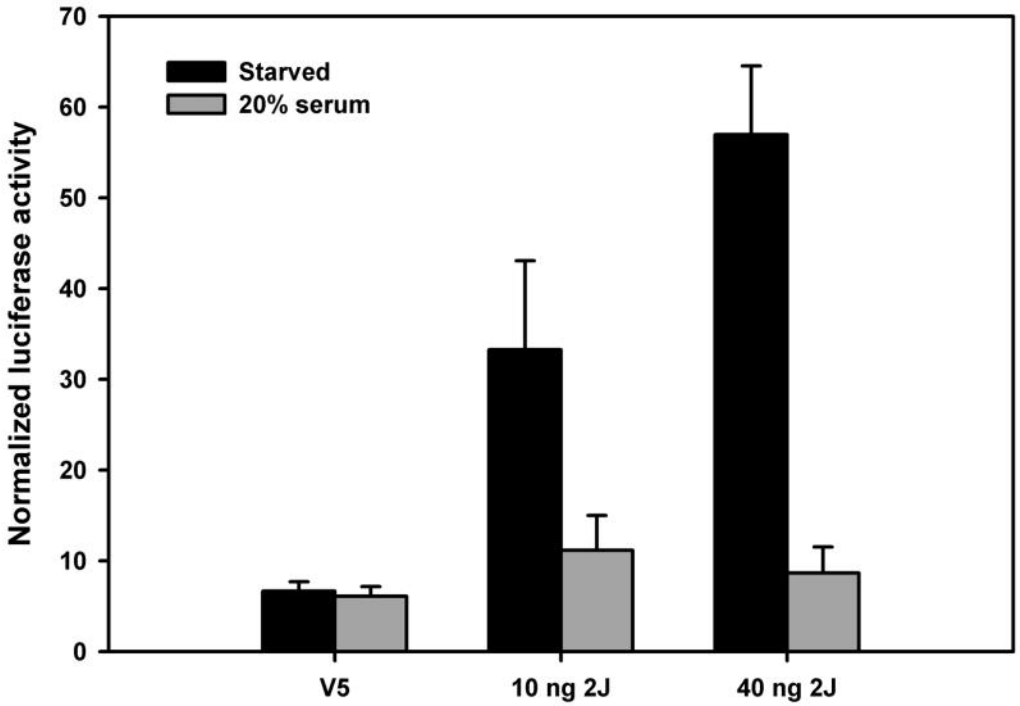


Figure 2.5 Recombinant *Ancylostoma caninum* DAF-16 drives expression from the DBE. Plasmid pCMV4-DAF-16 -2J encoding full length FLAG tagged Ac-DAF-16 was cotransfected (10 or 40 ng) with 50 ng of reporter vector p6xDBE-luc and 10 ng of pRL-TK-*Renella* luciferase vector into NIH 3T3 cells. After 24 hrs, cells were incubated in either medium alone (black bars) or with medium plus 20% bovine calf serum (gray bars). The relative luciferase activity (firefly luciferase activity divided by the *Renilla* luciferase activity) was determined 24 hrs later.

Chapter 3. Identification of hookworm DAF-16/FOXO response elements and direct gene targets

Xin Gao[1], Zhengyuan Wang[2], John Martin[2], Sahar Abubucker[2], Xu Zhang [2], Makedonka Mitreva[2], and John M. Hawdon[1*]

[1] Department of Microbiology and Tropical Medicine, The George Washington University Medical Center, Washington, D.C.

[2] The Genome Center, Department of Genetics, Washington University School of Medicine, St. Louis, Missouri.

Abstract

The infective stage of the parasitic nematode hookworm is developmentally arrested in the environment and needs to infect a specific host to complete its life cycle. The canine hookworm (*Ancylostoma caninum*) is an excellent model for investigating human hookworm infections. The transcription factor of *A. caninum*, *Ac*-DAF-16, which has a characteristic forkhead or "winged helix" DNA binding domain (DBD), has been implicated in the resumption of hookworm development in the host. However, the precise roles of *Ac*-DAF-16 in hookworm parasitism and its downstream targets are unknown. In the present study, we combined molecular techniques and bioinformatics to identify a group of *Ac*-DAF-16 binding sites and target genes.

In present study, the DNA binding domain of *Ac*-DAF-16 was used to select genomic fragments by *in vitro* genomic selection. Twenty four bound genomic fragments were analyzed for the presence of the DAF-16 family binding element (DBE) and possible alternative *Ac*-DAF-16 bind motifs. The 22 genes linked to these genomic fragments were identified using bioinformatics tools and defined as candidate direct gene targets of *Ac*-DAF-16. Their developmental stage-specific expression patterns were examined. Also, a new putative DAF-16 binding element was identified.

Our results show that *Ac*-DAF-16 is involved in diverse biological processes throughout hookworm development. Further investigation of these target genes will provide insights into the molecular basis by which *Ac*-DAF-16 regulates its downstream gene network in hookworm infection.

Key words: DAF-16, DBD, *In Vitro* Genomic Selection

1. Introduction

Many parasitic nematodes, including hookworms, infect the definitive host as developmentally arrested third-stage larvae (L3). The L3 is analogous to the dauer stage of the free-living nematode *Caenorhabditis elegans* in many biological aspects (Hawdon & Shad, 1991; Hotez et al., 1993; Rogers & Sommerville, 1963). The FOXO-family forkhead transcription factor DAF-16 mediates dauer formation of *C. elegans* in response to cues indicating poor environmental conditions. When conditions improve, DAF-16 is negatively regulated by an insulin-like signaling (ILS) pathway that culminates in transport of phosphorylated DAF-16 out of the nucleus (Paradis & Ruvkun, 1998; Paradis et al., 1999; Ogg & Ruvkun, 1998; Ogg et al., 1997; Morris et al., 1996; Kimura et al., 1997; Gottlieb et al., 1994). The primary protein structure of DAF-16 contains a conserved forkhead or "winged helix" DNA binding domain (DBD) with three major α-helices and two large wing-like loops (Clark et al., 1993; Hacker et al., 1992; Weigel et al., 1989). Orthologs of DAF-16 have been recently characterized in the parasitic nematodes *Ancylostoma caninum* (*Ac*-DAF-16), *Strongyloides stercoralis* (*Ss*-DAF-16), and *Haemonchus contortus* (*Hc*-DAF-16) (Massey et al., 2003; Massey et al., 2006; Gao et al., 2009; Hu et al., 2009). Heterologous rescue of *C. elegans daf-16* mutants (Massey et al., 2006; Hu et al., 2009) and reporter assays in mammalian cells (Gao et al., 2009; Kiss et al., 2009) indicate that parasitic nematode DAF-16 orthologs play similar regulatory roles during development, providing further support for the use of dauer exit

as a model to investigate the molecular events of infection and successful establishment of a parasitic relationship with the host (Hawdon & Schad, 1991).

Murine DAF-16/FOXO was shown to bind an 8-bp consensus DAF-16 family member binding element (DBE) *in vitro* (Furuyama et al., 2000). Several approaches have since been used to identify DAF-16 target genes in *C. elegans* , with the results suggesting that DAF-16 is recruited to a large number of promoters to modulate the expression of genes involved in development, metabolism, stress responses, and longevity (Oh et al., 2006; Lee et al., 2003; Murphy et al., 2003; McElwee et al., 2003).

Our lab has been focusing on the infectious process of hookworms, one of the most common infectious diseases in tropical and subtropical countries, causing anemia and malnutrition in almost a billion people (Bethony et al., 2006). The canine species *A. caninum* is a commonly used model for investigation of human hookworm infections. The DAF-16 ortholog from *A. caninum* (*Ac*-DAF-16) was shown to be transcriptionally active and capable of interacting with a hookworm 14-3-3 protein, suggesting a critical role in gene expression associated with hookworm L3 development and the transition to parasitism (Gao et al., 2009; Kiss et al., 2009). Given the functional conservation between dauer recovery and hookworm infection, there is considerable interest in the transcriptional outputs of DAF-16 in hookworm and their function in parasitic development. Dissecting the hookworm DAF-16 downstream effector network will have important implications in the development of new intervention strategies for hookworm and other nematode infections.

The present study utilizes *in vitro* genomic selection, a technique built on the concept of systematic evolution of ligands by exponential enrichment (SELEX), where

natural genomic sequences are used as a source for selection and amplification (Tuerk et al., 1990). A combination of *in vitro* genomic selection and subsequent cloning has been developed as a powerful method to identify naturally occurring DNA-binding sites in a genomic context and provide a foundation for investigation of the *in vivo* targets of DNA-binding proteins (Shostak et al., 2006; Jakobsen et al., 2004; Kojima, et al., 2006). We employed the *in vitro* genomic selection strategy using *Ac*-DAF-16 DBD to screen digested hookworm genomic DNA, and identified high-affinity binding sites in the hookworm genome and potential *Ac*-DAF-16 gene targets. Finally, the expression profile of the *Ac*-DAF-16 related transcripts was determined by examining cDNAs from four developmental stages of *A. caninum*.

2. Materials and Methods

2.1 Ac-DAF-16 DBD cloning, expression and purification

To clone the *Ac*-DAF-16 DBD, a fragment containing the C-terminally FLAG-tagged DBD (corresponding to amino acids 214-314) was amplified from a cDNA clone of *Ac*-DAF-16 isoform b, pCMV4-daf16 (Gao et al., 2009). The specific forward primer (DAF16-DD-E-FLAG-FN: 5' – GATACCATGGGCAA TGCGTGGGGTAATCTC - 3', containing restriction site *Nco*I, underlined) and reverse primer (DAF16-DD-E-FLAG-RH: .5' – GATCAAGCTT*ATCGTCG TCATCCTTGTAGTC* CAAGGTGGAGGCTCG AAC - 3', containing restriction site *Hind*III, and the FLAG Tag, italics), were incubated with the template in a PCR. The cycling conditions were 2 min at 95°C; followed by 35 cycles of 1 min at 95°C, 1 min at 55°C, 1 min at 72°C and a final extension for 6 min at

72°C. The purified amplicon was digested with *Nco*I and *Hind*III, ligated into expression vector pET28a (Novagen) cut with the same restriction enzymes, and transformed into *Escherichia coli* DH5α competent cells. The expression construct contained both an in-frame FLAG tag and an in-frame hexahistidine (His) tag at C-terminus of the *Ac*-DAF-16 DBD, as confirmed by DNA sequencing (Nevada Genomics Center, Reno, NV).

The resulting plasmid DNA (pET28a-*Ac*-DAF-16 DBD/ FLAG/His) was transformed into *E. coli* Rosetta (DE3) competent cells (Strategene), and expression was induced by addition of 1mM isopropyl-β-D-thio-galacto-pyranoside (IPTG) to log-phase bacterial culture at 37°C for 4 hrs. An aliquot of *E. coli* Rosetta (DE3) cells were removed from the culture prior to induction to serve as a pre-induction control, and second aliquot from a culture grown under the same conditions except in the absence of IPTG served as an un-induced control.

Induced bacterial cells and control bacterial cells were collected by centrifugation at 5000 rpm for 15 mins. After cell lysis by sonication in the presence of protease inhibitors (Pierce, Thermo Scientific, Rockford, IL), the expressed recombinant *Ac*-DAF-16 DBD (r*Ac*-DAF-16 DBD) was affinity-purified by Ni-NTA resin (Qiagen, Valencia, CA). Purified r*Ac*-DAF-16 DBD peptide was fractionated by SDS-PAGE eletrophoresis through 4-20% Tris-glycine pre-cast Novex gradient gels (Invitrogen, Carlsbad, CA), and examined by staining with Coomassie Brilliant Blue R250 (Sigma, St. Louis, MO) and Western Blot with anti-His (C-term) antibody (Invitrogen). The concentration of the expressed peptide was determined by the Micro BCA Protein Assay (Thermo Scientific).

2.2 Binding Assay for Ac-DAF-16 DBD

The DNA binding ability of r*Ac*-DAF-16 DBD was tested by streptavidin bead pull-down assay using the double stranded DNA molecules as described previously (Gao et al., 2009). Briefly, 5′ end biotin-labeled forward oligonucleotides and their unlabeled complements for DBE or predicted motif elements were synthesized (IDT, Coralville, IA). Forward and reverse oligonucleotides were annealed to form double-stranded (ds) DNA and 100 pmoles were used in the pull-down assays. To label selected genomic fragments, 5′ end biotin-labeled T7 primer was synthesized and used in PCR reaction. The amplicons were purified and 2 pmoles were used in the pull-down assays. Biotin-labeled dsDNA was incubated with 200 ng of *Ac*-DAF-16 DBD in binding buffer (10 mM Tris pH7.5, 50 mM KCl, 1 mM DTT, 2.5% glycerol, 5 mM $MgCl_2$, 50 ng/μL poly (dI.dC), 0.05% Nonidet P-40) at 4 °C for 2 h. After incubation, 25 μL of 30% Streptavidin–Sepharose (GE Healthcare) slurry equilibrated with TNE_{50} buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.5) + 0.1% Nonidet P-40 were added to the mixture and incubated for 2 h at 4 °C. Beads were collected by centrifugation at 2500*g* and washed twice with TNE_{100} buffer (10 mM Tris, 100 mM NaCl, 1 mM EDTA pH 7.5) + 0.1% Nonidet P-40. The bound proteins were separated on a 4–20% SDS–PAGE and Western blotted with anti-FLAG antibody (Sigma).

2.3 Hookworm genomic DNA preparation

The Baltimore strain of *A. caninum* (US National Parasite Collection #100655.00) was maintained in beagles as previously described (Schad, 1982). The George Washington University Medical Center Institutional Animal Care and Use Committee approved this study (protocol #A147). Infective L3 larvae were recovered from

coproculture by a modified Baermann technique and stored in buffer BU (50 mM Na_2HPO_4, 22 mM KH_2PO_4,70 mM NaCl, pH6.8) (Hawdon & Schad, 1991b) at room temperature until used.

Eighty thousand hookworm L3 were frozen in liquid nitrogen and ground to a fine powder using a pre-chilled mortar and pestle. Following physical disruption of the worms, hookworm genomic DNA was isolated using the Wizard SV Genomic DNA Purification System (Promega) according to the manufacture's instructions. Briefly, digestion solution master mix containing RNase A (75 μg/ mL) and proteinase K (1.5 mg/ mL) was added to the homogenate and the sample was incubated at 55°C overnight. The sample was then mixed with lysis buffer, applied to the minicolumn assembly, and eluted with buffer TE (10 mM Tris.HCl, 1 mM EDTA, pH8.0). Fifteen micrograms of *A. caninum* genomic DNA were digested with 48 units of *Bfu*CI restriction endonuclease (NEB) overnight and stored at -20°C. The quality of DNA was examined by 0.8% agarose gel electrophoresis and the concentration was determined using a NanoDrop ND-1000 spectrophotometer.

2.4 In vitro Genomic Selection

Thirty microliters of anti-FLAG M2 affinity matrix (Sigma) were rinsed with TBS buffer (10 mM Tris HCl, pH 7.4150 mM NaCl, 0.1 mM EDTA) three times. Two micrograms of purified r*Ac*-DAF-16 DBD /FLAG/His were immobilized on the prepared matrix by incubation for 2 hours at 4°C in 150 μL of TBS buffer. The matrix was then washed with TBS buffer containing 1 M NaCl, followed by two washes with the binding buffer (10 mM Tris at pH7.5, 2.5% glycerol, 10 mM MgCl2, 50 mM KCl, 1 mM

dithiothreitol (DTT), 0.05% NP-40). All washes were performed at 4°C for 5 min on a nutating mixer. Immobilized *Ac*-DAF-16 DBD was then incubated with 5 μg of BfuCI digested *A. caninum* genomic DNA, in 150 μL of binding buffer containing 50ng/μL poly (dI•dC) for 30 min at room temperature on a nutating mixer. To control for non-specific binding and subsequent PCR enrichment of spurious genomic fragments, a neutral control selection using only the anti-FLAG M2 affinity matrix was performed under identical conditions.

Protein-DNA matrix complexes were washed with 150 μL binding buffer for 5 min on a nutating mixer at 4°C, followed by a wash with binding buffer containing 250 mM KCl and a third wash containing 500 mM KCl. Washing conditions were optimized with the consensus DBE, and 500 mM KCl was shown to be sufficient to remove most of the unspecific binding. *Ac*-DAF-16 DBD-bound DNA was eluted from the matrix with 250 μL of the binding buffer containing 1 M KCl (10 min with 150 μL followed by 15 min with 100 μL) on a nutating mixer. Two hundred and fifty microliters of TE (10 mM Tris HCl, pH 8.0, 1 mM EDTA) were then added to the eluate. The eluted DNA was phenol/chloroform-extracted and ethanol-precipitated in the presence of 5 μg of glycogen. The DNA pellet was re-dissolved in 12 μL H_2O. The following oligonucleotides, which contain the *Not*I restriction site, were synthesized: forward *Not*I*Bfu*CI (5′-AAAAGGGGCGGCCGC-3′) and reverse *Not*I*Bfu*CI (5′-GATCGCGGCCGCC CCTTTT-3′), and annealed to form double-stranded (ds) *Not*I*Bfu*CI oligomer DNA for use in a ligation reaction. One pmole of ds *Not*I*Bfu*CI oligomer was ligated to 10 μL of the re-dissolved *Ac*-DAF-16 DBD bound DNA with T4 DNA ligase in 20 μL total volume. Two microliters of the ligation product were used as

template in PCR reactions with *Not*I*Bfu*CI PCR primer (5′-AAAAGGGGCGGCCGCG ATC-3′). The cycling conditions were 4 min at 95°C, followed by 29 cycles of 1 min at 95°C, 1 min at 57°C, 2 min at 72°C and a final extension for 6 min at 72°C.

Amplified PCR products were purified with NucleoSpin® Extract II kit (Clontech). Two micrograms of the DNA were used for subsequent rounds of binding experiments, as described above. After each round of binding and elution, PCR was performed for a total of 4 rounds. Purified PCR products from the last round were subcloned either into the *Not*I site of pBluescript KS+ vector (Stratagene) or directly into pGEM -T Easy vector (Promega) for sequencing and analysis.

2.5 Motif analysis of hookworm DAF-16 bound genomic fragment

Constructs containing *Ac*-DAF-16 bound fragments from the *in vitro* genomic selection cloned in pBluescript KS+ / pGEM-T were sequenced with vector specific primer T7. The sequence chromatograms were screened using sequence display software Chromas 2.33 (Technelysium Ltd.) to identify high quality regions, vector linker regions, and the low complexity regions. All subsequent analyses were based on the high-quality sequences with vector linker sequences removed.

Nucleotide sequences were searched against the draft *A. caninum* genome scaffold sequences (Abubucker et al., 2008) (Mitreva, unpublished) using BLAST to confirm the identity of the hookworm genomic fragments. If the overall percent identity of an alignment was above 95%, the genome scaffold was assumed to be a hit. The bound fragments that failed to hit any *A. caninum* genome scaffolds were further searched against nucleotide databases at GenBank (http://www.ncbi.nlm.nih.gov/) to detect

possible contamination from other genomes. The scaffold hits were next analyzed for their coding potential by mapping the cDNAs to the genome sequence (see below).

Given the current knowledge of reported FOXO transcription factor binding sites, *Ac*-DAF-16 bound *A. caninum* genome fragments were first inspected for existence of the consensus DBE, 5′-TTG/ATTTAC-3′ (and the corresponding reverse compliment 5′-GTAAAC/TAA -3′) (Furuyama et al., 2000), or its reverse counterpart, 5′-CATTTA/GTT-3′ (and the corresponding reverse compliment 5′-AAT/CAAATG-3′). The motif discovery algorithm Gibbs Motif Sampler (Neuwald et al., 1995; Liu et al., 1995) was applied to sequences lacking canonical binding sites to seek over-represented motifs as additional candidate *Ac*-DAF-16 response elements. The default eukaryotic parameters were used to run the Gibbs Motif Sampler program with a slight modification made to the target motif widths, with 6, 8, 10, 12, and 14 bp chosen based on the width range of *C. elegans* transcription factor binding sites in Open REGulatory ANNOtation database (ORegAnno) (Griffith et al., 2008). Two other motif discovery algorithms, MEME (Bailey et al., 2006) and Mdscan (Liu et al., 2002), were also attempted for de novo motif discovery; however, unlike the robust Gibbs Motif Sampler, either their performance in detecting the consensus DBE with a set of positive control sequences was poor or the parameter setting was not flexible. Therefore, only Gibbs Motif Sampler was used for the current data set. Returned motifs were ranked based on the occurrence rates in the bound genomic fragments. Sequence logos for identified motifs were generated using WebLogo (Crooks et al., 2004).

2.6 Identification of potential Ac-DAF-16 regulated genes

Ancylostoma caninum genome scaffold sequences containing *Ac*-DAF-16 bound genomic fragments, stage specific *A. caninum* expressed sequence tag (EST) databases (Martin et al, 2009), and the recent *A. caninum* cADNA databases (Wang et al., 2010) were obtained from Nematode.net (Mitreva et al., 2005). The available data sets cover 93% of the *A. caninum* transcriptome.

For each genome scaffold hit, the genomic template including the *Ac*-DAF-16 bound region and up to 3 kb extensions at both ends were extracted. Those templates were searched against EST databases / transcriptome databases using BLAST to identify corresponding transcripts. The cut-off for the overall percent identity of an alignment was 93% and a length of at least a 100 bp. Gene location was predicted with greater confidence if more ESTs were aligned. Gaps between two adjacent exon segments in an EST alignment were treated as possible introns and were confirmed by checking intron boundary sequences (GT/AG rule) (Keller & Noon, 1984). The transcript contigs were then derived from those ESTs.

The identified transcripts were compared with existing protein sequences at GenBank for functional annotation, using a maximum E-value of 1×10^{-10} and a minimum of 50% similarity as cut-off. Their expression specificity across different developmental stages (L3, aL3, F and M) were characterized using a statistical approach defined by Audic (Wang et al., 2010; Audic & Claverie, 1997). Briefly, the cDNAs were organized into transcript contigs, and ESTs were grouped to the corresponding cDNA libraries derived from different developmental stages for each transcript contig. The frequencies of library specific cDNAs for each contig were recorded and analyzed using a modified Fisher's exact test with a significance of $p < 1$ e-05. This allowed definition of

the stage specificity of transcript expression with significance. Transcripts were also mapped to the three organizing principles of the Gene Ontology (GO) based on sequence similarity displayed using AmiGo (http://amigo.geneontology.org), and are available at Nematode.net (http://nematode.net).

3. Results

3.1 Ac-DAF-16 DNA binding domain (DBD) expression and purification

Amino acid sequence alignment of different FOXO proteins revealed that the DBD is approximately 100 amino acid-long, with a conserved N-terminal region, and a divergent, but arginine/lysine-rich C-terminal region (Figure 3.1). Several crystal structures for FOXO transcription factor DBDs have been solved, and reveal that the structural basis for FOXO protein recognition of DNA comes from direct base-specific contacts as well as phosphate contacts between DNA molecules and critical C-terminal arginine/lysine amino acid residues in the FOXO DBD (Brent et al., 2008; Tsai et al., 2007). Based on the information from those crystal structures and sequence alignment between different DAF16/FOXO molecules, the DBD of *Ac*-DAF-16 was defined to start at amino acid Asn214 and end at amino acid Asp314.

Properly functioning recombinant hookworm DAF-16 DBD peptide was required for the *in vitro* genomic selection technique. A fragment of 303 bp corresponding to *Ac*-DAF-16 DBD (aa 214-314), and a 249 bp fragment corresponding to a truncated *Ac*-DAF-16 DBD (aa 220-302) lacking the arginine/lysine-rich C-terminal region (*ΔAc*-DAF-16 DBD) were cloned and expressed (Figure 3.2A). The calculated molecular

masses for the coding regions of these two constructs are 14047.4 Da (pET28a-*Ac*-DAF-16 DBD) and 12510.6 Da (pET28a-Δ*Ac*-DAF-16 DBD). As expected, bands corresponding to the predicted molecular weights (14 kDa for r*Ac*-DAF-16 DBD, and 12 kDa for r*Δ Ac*-DAF-16 DBD) were detected by Coomassie Blue-staining (Figure 3.2B). Immunoblots using an anti-His tag (C-term) antibody indicated that the bands were present only in *E. coli* cultures that had been transformed with the expression constructs and induced with IPTG (Figure 3.2C, Lanes 3 and 4).

Ac-DAF-16 was previously shown to bind to the consensus DBE sequence and initiate reporter gene transcription (Gao et al, 2009). Our pull-down assay results (Figure 3.3) indicated that r*Ac*-DAF-16 DBD, but not r*Δ Ac*-DAF-16 DBD, recognizes and binds strongly to the conserved DBE, indicating that the arginine/lysine -rich section at the C-terminus of DBD is critical for its binding activity.

3.2 In vitro genomic selection

The *in vitro* genomic selection enabled unbiased identification of transcription factor binding sites in the absence of influence from chromatin and other cofactors (Turek et al., 1990). Immobilized r*Ac*-DAF-16 DBD was first prepared by binding to anti-FLAG M2 affinity matrix, which was confirmed by silver staining (Figure 3.4 A) and Western blot (Figure 3.4 B), and used to screen *Bfu*CI digested *A. caninum* total genomic DNA (Figure 3. 4C). Discrete bands appeared progressively over the four rounds of binding and PCR amplification (Figure.3.4D, lanes 2 to 5), suggesting preferential amplification of particular genomic fragments. Cloning the amplified DNA fragments from the fourth round resulted in a total of 311 clones (117 pBluescript KS+

constructs and 194 pGEM-T Easy constructs), and high-quality sequences were obtained for 274 of them. The length ranged from 100 bp to 300 bp. Sequence analysis showed that 198 sequences contained low-complexity microsatellite regions, which were characterized by the presence of the repetitive trinucleotide "GTT" or its reverse complement "AAC" with a repeat number of 5 to 15. The remaining 76 sequences represented 25 distinct genomic fragments, and 24 of them were successfully mapped to *A. caninum* genomic sequences. A single fragment remained unidentified after extensive sequence similarity searching against nucleotide databases at Genebank, suggesting that it might by found in a section of the *A. caninum* genome that has not been sequenced. Five of the 25 fragments were overrepresented, and the remaining 20 fragments were recovered once or twice (Table 1). The sequences of the distinct genomic fragments were submitted to the Genome Survey Sequence database (dbGSS) at the NCBI, and the accession numbers reported in Table 3.

A control selection was performed to control for non-specific binding to the antibody-bead matrix. A total of 211 clones were picked, and 190 high quality sequences were obtained. Sequence analysis showed that anti-FLAG M2 affinity matrix bound to a different set of hookworm genome fragments (Table 4). Among these sequences, only one was shared with the DAF-16 DBD-selected fragments, indicating that selection using the DBD generated specific genomic fragments associated with DAF-16 binding elements.

3.3 Motif analysis of rAc-DAF-16 DBD bound hookworm genome fragments

The consensus DBE is an 8-bp oligonucleotide, with the sequence 5′-TTGTTTAC-3′ (Furuyama et al., 2000). The core sequence of DBE is TGTT, and a single base-pair replacement in this sequence might significantly weaken its interaction with forkhead proteins. Therefore, only a single mismatch outside of the DBE core sequence was allowed when the 24 selectively bound fragments were inspected for DBEs. Using this criterion, 13 fragments contained the DBE or its reverse counterpart (Table 1), and two contained two copies of the DBE arranged as direct repeats separated by variable length of nucleotides. The remaining 11 genomic fragments were analyzed further, and a new over-represented sequence, 5′-GAC/GAA/TG-3′, was found, occurring 17 times in 9 of those fragments (Figure 3.5A, Table 1). Genomic fragments selected by anti-FLAG M2 were also analyzed for over-represented motifs. One element (5′-AGGAAGAG- 3′) was found in 36% of the control fragments, but only 5% (5/103) of the fragments contained a DBE-like element (Table 4), indicating that the DBE and over-represented GAC/GAA/TG motif were bound specifically by r*Ac*-DAF-16 DBD.

The newly identified motif was tested for its ability to bind to *Ac*-DAF-16 DBD. Oligos containing the four most abundant motifs (5′-GACAAG-3′, 5′-GACATG-3′, 5′-GAGAAG-3′, 5′-GAGATG-3′) were tested in a streptavidin bead pull-down assay. The oligos failed to pull down r*Ac*-DAF-16 DBD (data not shown). One possible explanation for this result was that additional flanking sequence might be required for efficient binding. To test this possibility, biotinylated T7 primer was used to amplify representative genomic fragments, which were tested for their ability to bind the DAF-16 DBD. As shown in Figure 3.5B, r*Ac*-DAF-16 DBD strongly bound to positively selected genomic fragments that contained a DBE (fragments 4.6 and 2.23) or the newly identified

motif (fragments 3.23 and 3.28), but not to the control anti-FLAG M2 affinity matrix selected (control) genomic fragments. This suggests that sequence flanking the element is required for in vitro binding, and that the newly identified motif specifically binds to the *Ac*-DAF-16 DBD when located in the proper context.

3.4 Genes or gene clusters linked to the recovered rAc-DAF-16 DBD bound fragments

Transcription factor binding sites are typically linked to their direct target genes. To identify the genes or gene clusters linked to the r*Ac*-DAF-16 DBD bound fragments, approximately 6 kb of the *A. caninum* genome scaffold sequences (Abubucker et al., 2008; Mitreva, unpublished) containing the fragments at the center were searched against stage-specific hookworm cDNA databases. Nineteen of the 24 genome scaffolds retrieved about 5000 cDNAs from different developmental stages, one scaffold matched an rRNA region, and the other four scaffolds either had poor sequence alignments or failed to retrieve any cDNAs (Table 2, Table 5).

Clustering the retrieved cDNAs yielded 22 transcript contigs. One r*Ac*-DAF-16 bound fragment was allowed to be linked to one or two transcript contigs. Sequence alignment between those transcripts and genome scaffolds revealed that r*Ac*-DAF-16 bound fragments resided in coding regions, introns, or 3′ untranslated regions (3′-UTR) (Table 5). However, not all the relative locations, especially the upstream ones, could be detected by this method due to incomplete annotation of the hookworm genome. Transcript sequences were searched against protein databases in Genbank and Gene Ontology Consortium for prediction of their function based on homology. The results classified r*Ac*-DAF-16 linked transcripts into two groups. The first contained transcripts

homologous to known or hypothetical proteins, and displayed a variety of putative biological processes such as intra- and extracellular signaling (PKA, NLP-2, SRT-42), metabolism (CPT-II, ABC transporter class F, phosphate transporter), development (cuticlin), and transcription / translation (SNR-3); the second contained the transcripts without any identified homologs across available protein databases and therefore represent putative *A. caninum* specific genes.

The recent studies of Wang et al (Wang et al., 2010) generated 1.5 million *A. caninum* cDNAs from four developmental stages of *A. caninum*: infective L3 larva (L3), serum stimulated (activated) L3 larva (aL3), adult male (M), and adult female (F) , covering approximately 93% of the *A. caninum* transcriptome. This dataset allowed construction of a stage-specific digital expression profile for many transcripts (Wang et al., 2010). Examining the available expression pattern profiles indicated that five r*Ac*-DAF-16 linked transcripts (PKA, contig00807, contig25170, contig12656, and contig08715) were significantly up-regulated and three (contig40879, contig04080, and phosphate transporter) were significantly down-regulated by serum stimulation (Table 2, Table S3), which is hypothesized to mimic the very early events of hookworm infections (Hawdon & Schad, 1991). Three r*Ac*-DAF-16 linked transcripts (contig00807, contig25170, and NLP-2) are larval specific, and therefore turned off during development from L3 to adult. A major difference between parasitic hookworms and free-living *C. elegans* is that hookworms are dioecious, and the differential digital expression profile indicated that nine r*Ac*-DAF-16 linked transcripts are gender-specific (PKA, contig12656, contig40879, cuticlin1-17, CPT, contig04080, phosphate transporter, contig08715, and contig44862). Therefore, the DAF-16 linked transcripts identified here

are involved in multiple biological processes during at least 4 hookworm developmental stages.

4. Discussion

During infection, hookworm L3 resume developmental programs that had been arrested during the environmental stage, resulting in the successful establishment of a parasitic relationship with their host. Most of the molecular events associated with this "transition to parasitism" are still unknown. Dauer recovery of the free-living nematode *C. elegans* has long been used as a model for investigating the mechanisms of this transition to parasitism due to the biological similarities between the dauer stage and infective L3 of hookworms (Rogers and Sommerville, 1963; Hawdon and Schad, 1991; Hawdon and Hotez, 1996). Specifically, the conserved insulin signaling pathway, which is required for *C. elegans* to exit the dauer stage in response to the improving environmental conditions, is also involved in the hookworm infection process (Tissenbaum and Ruvkun, 1998; Tissenbaum et al, 2000; Hawdon and Dattu, 2003). Studies using inhibitors have shown that the ILS pathway is involved in hookworm larval activation, and the hookworm ortholog of the central transcription factor in this signaling pathway, DAF-16, has been identified and characterized (Brand and Hawdon, 2004; Gao et al, 2009). Furthermore, *Ac*-DAF-16 was transcriptionally active and interacted with hookworm 14-3-3 protein in a phosphorylation-status dependent way in cultured mammalian cell (Gao et al., 2009; Kiss et al, 2009).

DAF-16 is a FOXO transcription factor that is negatively regulated by ILS (Paradis and Ruvkun, 1998). It functions in numerous biological processes, including metabolism, life span, stress responses, and dauer formation and exit, by mediating downstream gene expression in response to environmental and nutritional conditions (Barbieri et al, 2003). Many of the downstream targets have been identified in *C. elegans*, and provide insights into the mechanism by which DAF-16 mediates multiple phenotypes outputs (Oh et al., 2006; Lee et al., 2003; Murphy et al., 2003; McElwee et al., 2003). While many of these mechanisms are conserved in hookworms, the life cycles of parasitic nematodes differ significantly from that of *C. elegans*, and at least some of the DAF-16 outputs are likely to mediate processes specific to nematode parasitism. Therefore, fundamental mechanistic questions about hookworm infection can be addressed by the identification of DAF-16 binding sites and direct gene targets, and how those downstream effectors are coordinated in hookworm development.

To begin identifying *Ac*-DAF-16 target genes, we used an *in vitro* genomic selection technique to enrich for genomic fragments that were bound by the recombinant *Ac*-DAF-16 DBD. *In vitro* genomic selection with transcription factors differs from other affinity-based strategies because it uses the native genomic background, and therefore direct transcription factor targets are identified without influences from complex cellular environments (Whittle et al., 2009). The cyclical strategy selects binding sites with relatively high affinity and reduces indirect or nonspecific binding. In the present study, a group of genomic fragments with high affinity for r*Ac*-DAF-16 DBD were enriched, as indicated by higher-than-expected occurrence of some genomic fragments. Sequence analysis indicated that more than half of those fragments contained a reported consensus

DBE or DBE-like element, and these elements were present in four of the five most represented DBD bound genomic fragments. The DBE or DBE-like elements occurred as single or multiple copies, and the element orientations and nucleotide spacers between the elements varied. As *C. elegans* DAF-16 and its mammalian homologs bind to this consensus DBE, the presence of the DBE in *Ac*-DAF-16 selected hookworm genome fragments suggests a conserved function for hookworm DAF-16. However, nearly half of the fragments we isolated did not contain a canonical DBE or DBE-like sequence, suggesting that the DAF-16 DBD bound to a previously unknown binding element. Further analysis of these fragments identified an over-represented 6-bp element, 5′-GAC/GAA/TG -3′. The *Ac*-DAF-16 DBD bound to amplicons containing this element, but not individual oligos, suggesting that the new element requires surrounding sequence for DBD binding. Our results do not rule out the presence of additional DAF-16 binding elements in the hookworm genome. It is not unusual that a single transcription factor has variable bind sites (Badis et al., 2009). In any case, we have identified a new, previously unreported DAF-16 DBD binding element in hookworms GAC/GAA/TG element, supporting a role for *Ac*-DAF-16 in multiple, perhaps novel hookworm-specific, biological processes.

The availability of the draft genome of *A. caninum* and comprehensive expression data enabled a detailed analysis of the DBD-bound genes. Twenty four of 25 r*Ac*-DAF-16 DBD selectively bound fragments were confirmed as *A. caninum*, and their proximal genomic regions were analyzed for coding regions. A total of 22 transcripts within a 6 kb range surrounding r*Ac*-DAF-16 DBD bound fragments were identified as *Ac*-DAF-16 primary target genes. The use of a 6 kb search range was based on *C. elegans*

transcription factor binding site analyses (Lee et al., 2003; Tissenbaum et al., 1998). However, not all the genomic sequence hits in the present study returned transcript contigs within this range. Therefore, an extended search might be necessary to identify them, as *A. caninum* has a larger genome than *C. elegans* (Abubucker et al., 2008) and the *A. caninum* genome project is still underway.

Functions of the proteins encoded by the transcripts were predicted based on their homology to known or hypothetical proteins. This analysis suggested that DAF-16 is involved in a variety of biological processes. For example, PKA is a conserved serine/threonine kinase, activated by second messenger cAMP, and converts various extracellular signals into intracellular processes (Taskén &Aandahl, 2004). Serpentine receptors are G protein-coupled transmembrane receptors that play important roles in *C. elegans* chemoreception (Troemel et al., 1995). The NPL protein is critical for synaptic transmission between neurons (Nathoo et al., 2001), and membrane transporters are involved in transport of a wide variety of substrates across extra- and intracellular membranes (Zhao et al., 2006; Webb et al., 1994). Cuticlins comprise the insoluble, high-order material in nematode cuticle and determine the developmental morphology and mobility of the worms (Sapio et al., 2005). Small nuclear ribonucleoproteins (SNR) are a part of RNA post-transcriptional modification machinery (Staley & Woolford, Jr, 2009). Among these genes, the serpentine receptor and ABC transporter were also reported in previous studies of DAF-16 targets in *C. elegans*, indicating that some pathways downstream of DAF-16 were conserved in free living and parasitic nematodes (Lee et al., 2003; Murphy et al., 2003; Oh et al., 2006). Additionally, some r*Ac*-DAF-16 DBD linked gene transcripts failed to match any homologs by exhaustive search of the available

databases, and were defined as *A. caninum* specific. These are of particular interest, as they are absent from *C. elegans* and consequently might be involved in parasitism. Furthermore, several of these molecules could be envisioned functioning during the transition to parasitism based on homology and expression pattern.

Combining the transcriptomic and genome sequences revealed the relative location of *Ac*-DAF-16 binding sites to their linked genes in the *A. caninum* genome. However, this method is biased towards identification of *Ac*-DAF-16 binding sites located between exons due to incompleteness of hookworm genome annotation and the inability to identify 5′ ends and promoter sequences. Nonetheless, all 10 fragments that could be definitively linked to a gene had binding elements in introns or downstream sequences, suggesting that they may be regulated differently from genes with DBEs in the promoter. In *C. elegans*, DAF-16 target genes containing DBEs located downstream of the start codon were more likely to be negatively regulated by DAF-16 (Liu et al., 2004). Characterizing intergenic *Ac*-DAF-16 binding sites will depend on further information about the genomic structure for the corresponding genes. The present study surprisingly indicates that *Ac*-DAF-16 binding sites, unlike *C. elegans* DAF-16 binding sites, reside at variable locations relative to the gene transcripts (Oh et al., 2006), suggesting that *Ac*-DAF-16 might regulate expression of genes with diverse functions and exert its action through different mechanisms.

In the absence of functional information for most of the identified r*Ac*-DAF-16 linked genes, evidence of differential expression is the most important source for prioritizing future investigations. Using the extensive cDNA dataset available for *A. caninum*, the expression profile for the gene transcripts have been quantitatively analyzed

by comparing the frequency of EST occurrence in the different cDNA libraries. Examination of the expression patterns for the identified transcript contigs in the present study suggests that *Ac*-DAF-16 regulates gene expression in all hookworm developmental stages studied, including exit from developmentally arrested infective L3 stage, maturation to adults, and sexual differentiation in adults.

Using the affinity-based *in vitro* genomic selection procedure, we have shown for the first time that *Ac*-DAF-16 directly binds to response elements in the hookworm genome. The relative location of *Ac*-DAF-16 bound elements to the linked genes is variable, with an apparent bias towards downstream locations. The *Ac*-DAF-16 direct target candidates that were identified include both conserved and *A. caninum* specific genes, and will be subject to future functional investigations. With more comprehensive screening such as chromatin immunoprecipitation and improved *A. caninum* genome and transcriptome data, more *Ac*-DAF-16 downstream targets will be detected. Subsequent manipulation of these genes may lead to novel avenues for intervention in the hookworm life cycle.

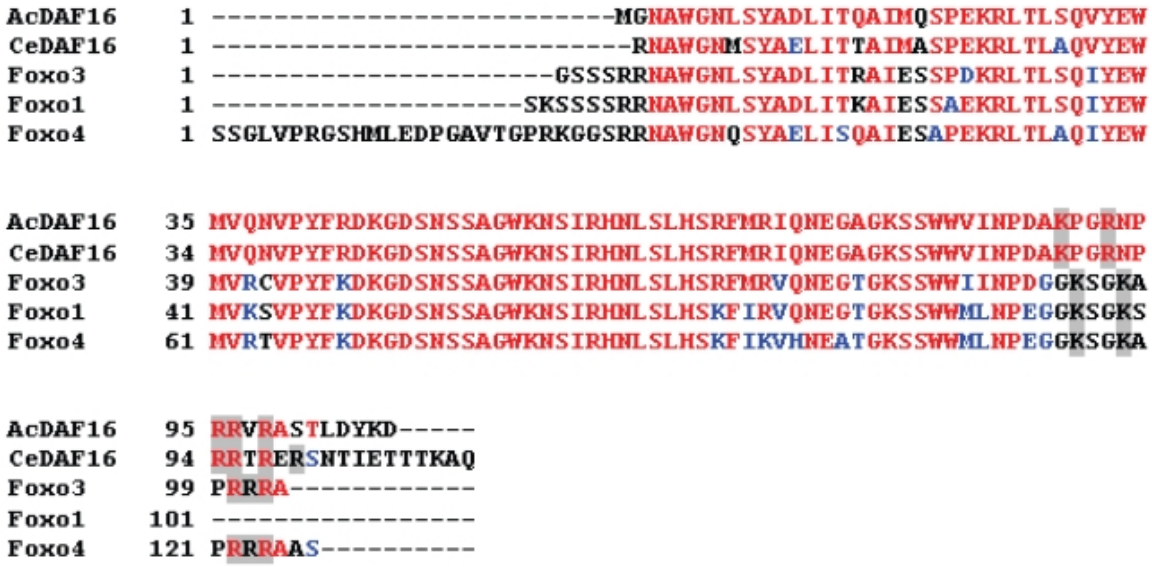


Figure 3.1 Amino acid sequences of DNA binding domains from selected FOXO transcription factors were aligned using CLUSTAL W software on the San Diego Supercomputer Center Biology Workbench server (http://workbench.SDSC.edu) and displayed using BOXSHADE3.21 software located on Swiss EMBnet server (http://www.ch.embnet.org). Identical amino acids are in red type, and conserved amino acids in blue. C-terminal arginine and lysine residues are shaded. The DBDs are from the following species: AcDAF16 (*Ancylostoma caninum* accession ACD85816); CeDAF16 (*Caenorhabditis elegans*, AAB84390); Foxo3, (*Mus musculus,* AAH19532); Foxo1 (*Homo sapiens,* AAH70065); Foxo4 (*Homo sapiens,* AAI06762).

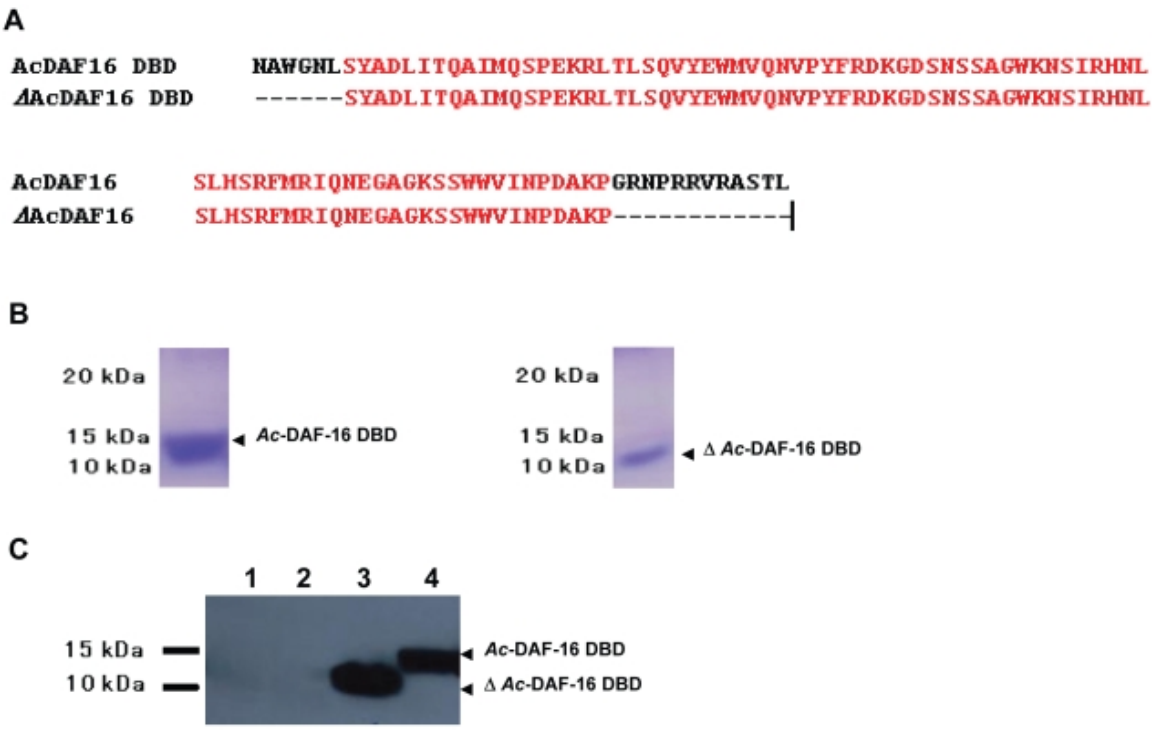


Figure 3.2 Expression of *Ac*-DAF-16 DBD (aa 214-314) and *Δ Ac*-DAF-16 DBD (aa 220-292) in *E. coli* Rosetta (DE3) cells. (A) Amino acid sequences of *Ac*-DAF-16 DBD and *Δ Ac*-DAF-16 DBD. Identical amino acids are in red. (B) Coomassie staining of purified r*Ac*-DAF-16 DBD and rΔ*Ac*-DAF-16 DBD. (C) Western blot of r*Ac*-DAF-16 DBD and r*ΔAc*-DAF-16 DBD probed with anti-his (C-term) antibody. Lane 1, non-transformed Rosetta (DE3) cells; Lane 2, pET28a-*Ac*-DAF-16 DBD transformed Rosetta (DE3) cells in the absence of IPTG; Lane 3, pET28a-*ΔAc*-DAF-16 DBD transformed Rosetta (DE3) cells induced with IPTG; Lane 4, pET28a-*Ac*-DAF-16 DBD transformed Rosetta (DE3) cells induced with IPTG. The arrowheads indicate the position of *Ac*-DAF-16 DBD or *Δ Ac*-DAF-16 DBD.

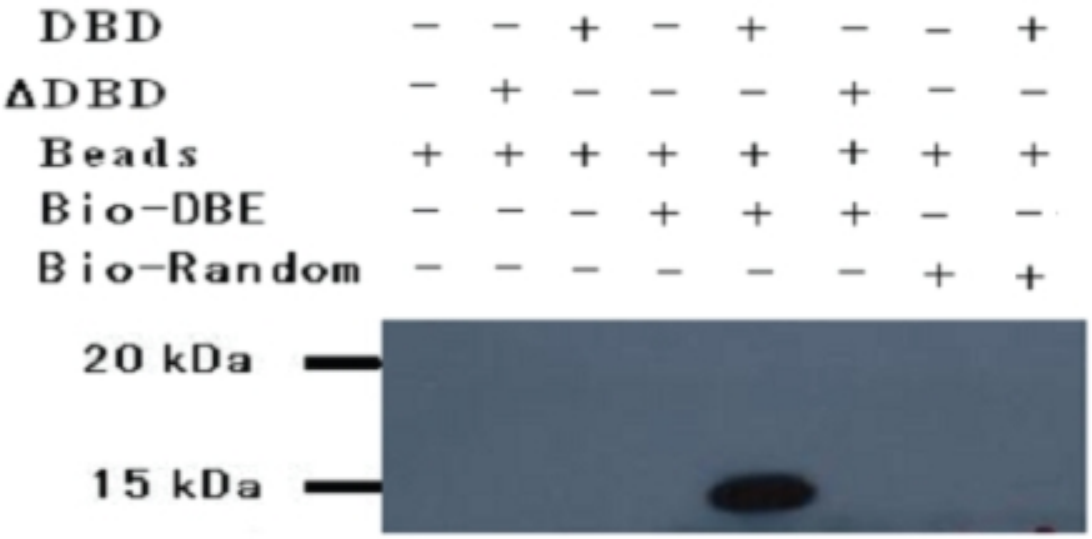


Figure 3.3 Streptavidin bead pull-down to detect DBE binding activity of recombinant *Ac*-DAF-16 DBD and Δ*Ac*-DAF-16 DBD. Biotinylated dsDBE (Bio-DBE) was incubated with purified *Ac*-DAF-16 DBD or Δ*Ac*-DAF-16 DBD, and peptide/oligonucleotides complexes were pulled down with streptavidin conjugated Sepharose beads. Precipitated DBD/oligo complexes were separated by SDS-PAGE and blotted to PVDF membranes for Western blotting using an anti-His (C-term) antibody. Bio-random represents a biotinylated oligomer of random sequence.

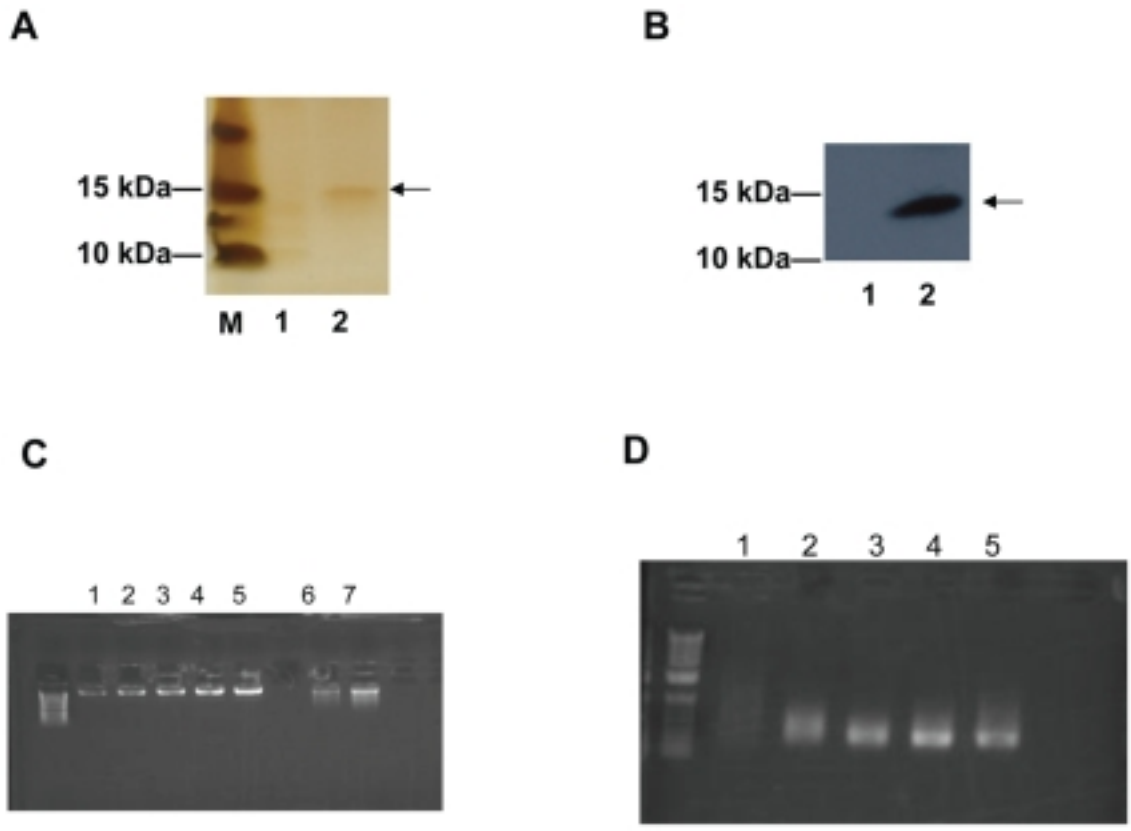


Figure 3.4 *In vitro* genomic selection of *A. caninum* DNA fragments containing *Ac*-DAF-16 binding sites. (A-B) Immobilized r*Ac*-DAF-16 DBD on Anti-FLAG M2 matrix confirmed by silver staining (A) and Western Blot with anti-FLAG antibody (B). The arrowheads indicate the position of *Ac*-DAF-16 DBD. M, protein standard; lane 1, Anti-FLAG M2 matrix; lane 2, anti-FLAG M2 matrix incubated with 2 ug of *Ac*-DAF-16 DBD and washed with 500 mM KCl. (C) *A. caninum* genomic DNA preparation. An 0.8% agarose gel was used to examine the DNA quality, and increasing amounts of a λ DNA standard were loaded to estimate DNA concentration. Lane 1-5, 30 ng λ DNA standard, 60 ng λ DNA standard, 90 ng λ DNA standard, 120 ng λ DNA standard, 150 ng λ DNA standard, respectively; lane 6, 2.5 μL of *A. caninum* genomic DNA sample; Lane7, 5 μL of *A. caninum* genomic DNA sample. (D) PCR amplification of the genomic fragments after each selection round. Lane 1, *A. caninum* genomic DNA sample cut with *Bfu*CI; lane 2, 1st round purified PCR product; lane 3, 2nd round purified PCR product; lane 4, 3rd round purified PCR product; lane 5, 4th round purified PCR product.

A.

B.

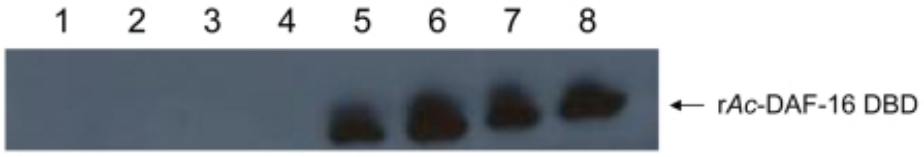


Figure 3.5 Identification of a novel DAF-16 DBD binding element. (A) Sequence logo of the putative DBD binding motif discovered using Gibbs Motif Sampler in bound fragments lacking a canonical DBE. (B) Streptavidin bead pull-down to detect the binding activity between positively selected genomic fragments and r*Ac*-DAF-16 DBD. Biotin labeled PCR products were incubated with r*Ac*-DAF-16 DBD. The protein/ biotinylated PCR amplicon complexes were separated by SDS-PAGE and blotted to PVDF membrances for Western Blotting using anti-FLAG antibody. Lane1 - 4, Genome fragments from control selection; Lane 5, Fragment 3.23 (GACAAG motif); Lane 6, Fragment 3.28 (GACAAG motif); Lane 7, Fragment 4.6 (DBE); Lane 8, Fragment 2.23 (DBE).

Table 1. **Proposed binding motifs found in fragments that bound to DAF-16 DBD *in vitro*.**

Genomc selection clone #	Motif present	Occurrence Rate
Fragment 1.11	TTGTTTAC $(N)_n$ TTGTTTAC	18
Fragment 1.25	CATTTGTT [a]	10
Fragment 1.3	GAAAACAA [b] $(N)_n$GTAAACAT [b]	1
Fragment 1.4	TTGTTTAT [b]	1
Fragment 2. 10	*GAGAAG*	1
Fragment 2.14	GTAAACAT [b]	1
Fragment 2.18	*GACAAG*	2
Fragment 2.19	GTAAACAA $(N)_n$ GTAAATAA	1
Fragment 2.23	AACAAATA [a,b]	12
Fragment G2.28	AACAAATA [a,b]	1
Fragment G2.38	*GACATG* $(N)_n$ GAGAAG	1
Fragment 3.23	*GACATG* $(N)_n$ GACATG	1
Fragment 3.28	*GGCAAG* [c]	1
Fragment 3.30		7
Fragment G4.41	CTGTTTAC [b] $(N)_n$CTGTTTAC [b]	1
Fragment 4.26	ATAAACAA $(N)_n$GTAAATAA	7
Fragment 4.6	GATTTGTT [a,b]	2
Fragment 5.20	*GGCAAG* [c]	1
Fragment G5.21	*GACAAG* $(N)_n$ GAGAAG	1
Fragment G5.7	*TACAAG*[c] $(N)_n$ GGCAAG [c]	1
Fragment 6.33		1
Fragment 6.34	TTGTTTAC	1
Fragment 6.48		1
Fragment 8.1	TATTTGTT $(N)_n$CATTTGTT [a,b]	1

Predicted new r*Ac*-DAF-16 binding sites were indicated as italic.
[a]Reverse DBE.
[b]One mismatch compared with consensus DBE.
[c]One mismatch compared with predicted GAC/GAA/TG binding site.

Table 2. Expression pattern of genes that bound to DAF-16 DBD in vitro.

Genomic Fragment	Contig ID[1]	Homology/description	Adult Expression Pattern[2]	Larval Expression Pattern[3]
1.3	11016	cAMP-dependent protein kinase (PKA)	down in F	up in aL3
3.28	00807			larval specific; up in aL3
1.4	52455		nc	nc
4.6	25170	Hypothetical protein R05C11.4		larval specific; up in aL3
6.34	03743	ABC transporter, class F family (abcf-2)	up in F	up in aL3
1.25	12656		M specific	up in aL3
2.18	20406	Neuro-peptide like protein family (NLP-2)		larval specific
2.18	40879	Hypothetical protein T01B6.1	F specific	down in aL3
2.19	09925	cuticlin1–17	up in F	up in aL3
3.30	53549	Carnitine palmitoyl transferase (CPT)	up in F	nc
6.68	04080		M specific	down in aL3
G2.38	05677	Phosphate transporter	F specific	down in aL3
G2.38	08715		Male specific	up in aL3
G5.7	44862		Adult specific; down in F	

[1]From Wang et al, 2010 .
[2]Change in relative transcript reads between adult female (F) and male (M) libraries. Up, up regulated; down, down regulated.
[3]Change in relative transcript reads between infective L3 and activated L3 (aL3) libraries.

Table 3. GenBank Accession numbers and dbGSS identification numbers of *Ancylostoma caninum* genomic fragments isolated by genomic selection.

dbGSS ID No.	User ID No.	GenBank Accession No.
28275220	Fragment-1-11	GS928112
28275221	Fragment-1-25	GS928113
28275222	Fragment-1-3	GS928114
28275223	Fragment-1-4	GS928115
28275224	Fragment-2-10	GS928116
28275225	Fragment-2-14	GS928117
28275226	Fragment-2-18	GS928118
28275227	Fragment-2-19	GS928119
28275228	Fragment-2-23	GS928120
28275229	Fragment-3-23	GS928121
28275230	Fragment-3-28	GS928122
28275231	Fragment-3-30	GS928123
28275232	Fragment-4-26	GS928124
28275233	Fragment-4-6	GS928125
28275234	Fragment-5-20	GS928126
28275235	Fragment-G-2-28	GS928127
28275236	Fragment-G-2-38	GS928128
28275237	Fragment-G-4-41	GS928129
28275238	Fragment-G-5-21	GS928130
28275239	Fragment-G-5-7	GS928131
28275240	Fragment-G-6-33	GS928132
28275241	Fragment-G-6-34	GS928133
28275242	Fragment-G-6-48	GS928134
28275243	Fragment-G-8-1	GS928135
28275244	Fragment-G-8-38	GS928136

Table 4. Frequencies of DBE containing genome fragments recovered from immobilized r*Ac*-DAF-16 DBD genomic selection and control genomic selection

	DAF-16 DBD selection	Control selection
Total constructs sequenced	311	211
High quality Sequences	274	190
Unique genome fragments	24 [a]	103
DBE containing genome fragments[b] (%)	13 (54%)	5 (5%)
Predicted new motif (% of sequences)	GACAAG (46%)	AGGAAGAG (36%)

[a] Not including sequences containing microsatellites.
[b] Includes DBEs that differ by a single base from the canonical DBE (TTG/ATTTAC)

Table 5. Identity and stage-specific relative transcript levels of *Ac*-DAF16-DBD- selected genes

			Relative transcript levels[2]			
Contig[1]	**Homology/Description**	**DBE location**	**F**	**M**	**L3**	**L3a**
contig11016	cAMP-dependent protein kinase (PKA)	Intron	↓	nc	nc	↑
contig00807			a	a	↓	↑
contig52455			nc	nc	nc	nc
contig04093						
contig28536						
contig25170	hypothetical protein R05C11.4	Coding region	a	a	nc	↑
contig03743	ABC transporter, class F family member (abcf-2)	3′ UTR	↑	↓	↓	nc
contig32336	Serpentine Receptor, class T family member (SRT-42)					
contig12656		Coding region	a	nc	nc	↑
contig20406	Neuro-peptide like protein family(NLP-2)	Downstream	a	a	↑	↑
contig40879	hypothetical protein T01B6.1		↑	a	nc	↓
contig09925	cuticlin1-17	Intron	↑	↓	a	↓
contig53549	carnitine palmitoyl transferase (CPT)	Intron	↑	↓	nc	nc
contig04080			a	nc	↑	nc
contig50874						
contig50821						
contig19818						
contig05677	phosphate transporter	Intron	nc	a	↑	nc
contig08715			a	↑	↓	nc
contig44862			nc	↑	a	a
contig41747	Small Nuclear Ribonucleoprotein family member (snr-3)	3′ UTR				
RRNA		Intergenic spacer				

[1]From Wang et al, 2010.
[2]Relative to expected transcript read number for the stage-specific library
a: transcript was absent from the stage-specific cDNA library, but not due to limited sequencing
nc: no significant change between observed transcript read number and the expected read number.

Chapter 4. Hookworm small nuclear ribonucleoprotein-3: gene structure, expression profile, and implications of 3'-untranslated region (3'-UTR) in the regulation of gene expression

Xin Gao, Kevin Goggin, Camille Dowling, Jason Qian, and John M. Hawdon

Department of Microbiology and Tropical Medicine, The George Washington University Medical Center, Washington, D.C.

Abstract

Forkhead transcription factor *Ac*-DAF-16/FOXO is implicated in many important biological processes during hookworm development. To better understand how *Ac*-DAF-16 achieves such diverse biological roles, it is critical to characterize it target genes. Our previous studies have identified a hookworm small nuclear ribonucleoprotein, SNR-3, as one of its potential downstream targets. Hookworm SNR-3 displayed 77% amino acid identity to the *C. elegans* SNR-3. However, their gene structures are quite different. The *Ac-snr*-3 gene spans at least 5.6 kb in the genome and contains 3 exons and 2 introns. The coding region is solely composed of the first two exons. Structural motifs and invariable amino acids are conserved in hookworm SNR-3. The transcriptional profile of hookworm *snr*-3 during post infection development showed that it was actively transcribed at the late L4 stage, suggesting a function in maturation to adulthood. The 3′-end genomic fragment (Fragment 2.23) of *A. caninum* SNR-3 was able to drive transcription of a reporter gene in the presence of *Ac*-DAF-16, regardless of its relative location (upstream or downstream), indicating that it might act as a *cis*-regulatory element mediating the effects of *Ac*-DAF-16/FOXO in hookworm. The unresponsiveness of this fragment to serum stimulation suggested that co-acting protein factors might be involved in its interaction with *Ac*-DAF-16/FOXO.

Key words: Hookworms, *Ac-snr*-3, *Ac-daf*-16, Forkhead transcription factor, *cis*-regulatory element

1. Introduction

Hookworm disease is a major public health problem in large portions of Southeast Asia, Africa and South America (WHO, 2010). The disease is caused by parasitic hookworms that have a complex life cycle involving free-living stages and host-dwelling stages (Hawdon and Hotez, 1996). The transformation of the hookworm from one developmental stage to another is accompanied by many drastic changes in morphological, physiological, and biochemical properties (Rogers et al., 1982; Hawdon et al., 1991). Faithful execution of these developmental processes requires a precise and carefully orchestrated series of steps dependent on the proper activation/inactivation of a selected group of genes, the so-called parasitic genes (Hawdon & Shad, 1991a; Hotez et al., 1993).

Gene expression begins with transcription initiation. Multiple protein factors, known as transcription factors, bind to specific DNA sequences. This binding event can activate or inhibit the transcription machinery so that gene expression is appropriately regulated (Lee & Young, 2000). Practically nothing is known about regulatory mechanism of gene expression in hookworm. However, this situation is expected to change with the prospect of the rapidly increasing hookworm genomic information (McCarter et al., 2005; Mitreva et al., 2007). The hookworm genome project, which was initiated 4 years ago and is nearly complete (Mitreval, personal communication, 2010), serves as a good starting point for studying transcriptional regulatory elements, and provides opportunities to address how stage specific gene expression patterns are established in hookworms.

We have long been interested in understanding the critical genes and pathways relevant to the developmental process of hookworms. Our previous studies demonstrated that insulin-like signaling (ILS) was required in the transition to parasitism of infective third-stage hookworm larvae (L3) and identified the first hookworm molecule in this pathway, a forkhead transcription factor *Ac*-DAF-16, from *Ancylostoma caninum* (Brand et al., 2004; Gao et al., 2009; Kiss et al., 2008). We showed that *Ac*-DAF-16 was transcriptionally active through a consensus DAF-16 binding element (DBE) and capable of interacting with the molecular chaperone *Ac*-FTT-2 in a phosphorylation-site dependent manner. More recently, we reported the isolation of a small set of hookworm genomic fragments that bound to the *Ac*-DAF-16 DNA binding domain (DBD), and took advantage of the availability of the hookworm genome and transcriptome data to further analyze the digital expression profiles of the linked transcripts (Wang et al., 2010; Gao et al., 2010). The predicted functions of those linked transcripts suggest that hookworm DAF-16, similar to its *C. elegans* and mammalian orthologs, is involved in diverse biological processes. It has been postulated that DAF-16/FOXO achieves such diverse biological roles by regulating different subsets of gene expression in response to different stimuli (Honda & Honda, 1999; Lee et al., 2003; Murphy et al., 2003).

To better decipher how hookworm DAF-16 mediates its diverse biological functions, a more comprehensive characterization of hookworm DAF-16 target genes is required. In our genome-wide screening of *Ac*-DAF-16 binding sites and the subsequent informatic searches for the aligned ESTs surrounding those binding sites, a hookworm ortholog of the *Caenorhabditis elegans* small nuclear ribonucleoprotein-3 (SNR-3/snRNP D1), *Ac*-SNR-3, was identified as a potential *Ac*-DAF-16 downstream target. *Ac-snr-3*

contained a DBE (the DAF-16 family binding element) located at its 3'-untranslated region (3'-UTR) (Gao et al., 2010).

SNR-3 is one of seven conserved Sm proteins that form a heptameric core complex required for the biogenesis and function of the small nuclear ribonucleoproteins (snRNPs) that catalyze mRNA splicing (Will et al., 2001; Kambach et al., 1999). In humans, Sm proteins are also clinically interesting since they are targets of autoantibodies in patients with systemic lupus erythematosus (SLE) (Lerner & Steitz, 1979; Craft et al., 1992). In *C. elegans*, SNR-3 has been implicated in controlling development, as evidenced by its association with extremely pleiotropic loss-of-function phenotypes affecting embryonic viability, larval viability, and nuclear localization of P granules when gene specific RNAi was applied (Barbee et al., 2002; Kamath et al., 2003; Simmer et al., 2003).

The relative location of *Ac*-SNR-3 to the identified DBE and its importance in worm development has led us investigate this gene in more detail. In the present study, we analyzed gene structure of hookworm *snr*-3 and its phylogenetic relationship to the SNR-3/snRNP D1 proteins from other species. The expression profile of hookworm SNR-3 was characterized across post infection developmental stages. The genome fragment downstream of its coding region increased basal luciferase gene expression compared to the controls in the presence of *Ac*-DAF-16. This study represents the first post-infection exploration of gene expression in hookworms, and the first functional analysis of a regulatory element present in the hookworm genome.

2. Materials and Methods

2.1 Sequence analysis of hookworm SNR-3

Hookworm expression sequence tags (ESTs), cDNA sequences, and genomic scaffolds and contigs were obtained from Nematode.net (http://www.nematode.net), Genbank at National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/genbank/), or through personal communication (Mitreva, 2010). Local DNA databases of hookworm genome scaffolds and ESTs were created and queried using BioEdit Sequence Alignment Editor Version 7.0.5.3 (Hill, 1999). The open reading frame and untranslated regions of the cDNA were predicted based on homology searches using BLASTX and BLASTP programs at NCBI. Intron-exon boundaries were located by comparison of ESTs and genomic DNA sequences using the Clustal W program located on Biology Workbench server (http://workbench.sdsc.edu). The rooted neighbor-joining phylogenetic tree was generated using Phylip's software package (located on Biology Workbench server) based on multiple sequence alignment results calculated from Clustal W. Transcription factor bind sites were predicted using Promo located at ALGGEN server (http://alggen.lsi.upc.es/).

2.2 Parasite culture and sorting

Ancylostoma ceylanicum was maintained in male beagles (Lab adapted protocol). Animals were housed and treated in accordance with institutional animal care and use committee guidelines at The George Washington University. Infective *A. ceylanicum* L3s were recovered from coproculture by modified Baemann technique after incubation at

22°C for approximately one week, and stored for periods up to 5 weeks in BU buffer (50 mM Na_2HPO_4/22 mM KH_2PO_4/70 mM NaCl, pH6.8) (Hawdon & Schad, 1991b), at room temperature until used for infection or cDNA preparation.

Four- to five-week-old hamsters were inoculated orally using a feeding needle with 2000 *A. ceylanicum* infective L3s orally for recovery of early parasitic stages, or with 100 hookworm L3 for recovery of later stages (L4 and adult). Hookworms were collected from intestines of animals at 24 h, 48h, 72 h, and 12 days post-infection. Mixed worm populations recovered at each time point were screened under dissecting microscope and divided into different developmental categories based on morphological characteristics. Worms from < 72 hrs post infection fell into category I to III: category I worms are almost identical to infective L3, category II demonstrate early development of the buccal cavity, and category III have a fully developed buccal cavity. Worms from 12 days post infection were grouped into category IV, which were nearly adults, and category V, adult worms. Up to 3 worms from each category were further examined under compound microscope at higher magnification.

2.3 RNA isolation

Worm pellets were rinsed with phosphate-buffered saline (PBS) (pH7.4) (Fisher Scientific, Pittsburgh, PA), suspended in 1 mL TRIzol reagent (Invitrogen, Carlsbad, CA) containing 0.5 mm zirconium oxide beads, and homogenized twice for 5 min at maximum speed in a bullet blender (Next Advance Inc, NY). Total RNA was isolated according to the manufacturer's instructions. Briefly, the homogenate was processed by sequential phase separation and isopropanol precipitation. The RNA pellet was washed

with 70% ethanol and resuspended in RNAse-free water. The resulting RNA was treated with RNase-free DNAse I (NEB, Ipswich, MA) to remove any trace of genomic DNA, followed by column purification using RNeasy Mini Kit (Qiagen, Valencia, CA).

2.4 cDNA synthesis

Due to the limited number of worms obtained from hamster small intestines at each time point, all of the cDNA samples were prepared from the total RNA using the Super SMART^{TM} PCR cDNA synthesis kit (Clontech, Mountain View, CA). Briefly, first-strand cDNA was synthesized and tailed in the presence of Superscript II reverse transcriptase (Invitrogen, Carslbad, CA), 3′ SMART CDS Primer II A that is a modified oligo-dT primer, and SMART II A Oligonucleotide, so that a linker sequence (5'-AAGCAGTGGTATCAACGCAGAGT-3') was added to both 5′- and 3′- ends of the first-strand cDNA. First strand cDNA was purified using NucleoSpin Extract II Columns (Clontech, Mountain View, CA) and further amplified by PCR using primers corresponding to the linker sequence to form double stranded cDNA. The cycling number for amplification was optimized by visual inspection of the exponential phase for each PCR reaction. Double stranded cDNA was also purified using NucleoSpin Extract II Column (Clontech, Mountain View, CA) and stored at –20°C until analyzed.

2.5 Quantitative real time PCR (qRT-PCR)

Specific forward (SNR-3-F: 5'-TCTGGCGAAAACATCCTGTC-3') and reverse (SNR-3-R: 5'-TATCCAGAGATCCGGAACTC -3') primers were designed from an *A. caninum snr*-3 expressed sequence tag (EXWOO7002HNN28) and used to amplify a

fragment of 111 bp. A fragment of 167 bp from the 60S acidic ribosomal protein (60S) was amplified with the primers 60SRP-F (5'-CTGCGTCTGCTGAAGAA-3') and 60SRP-R (5'-GTCTTG TTG CAT TTC GAG CA-3') for use as an internal reference gene (Trivedi and Arasu, 2005). All PCR reactions were performed over 40 cycles in triplicate.

Optimal cycling conditions and primer concentration for both *snr*-3 and 60S genes were determined using end-point PCR. Amplification efficiencies of the *snr*-3 and 60S were determined using standard curve experiments. cDNA templates from different developmental stages were mixed with *snr-3* primer set and RT^2 Real-Time™ SYBR Green qPCR Master Mix (SABiosciences, Frederick, MD), and the reaction run in a Bio-Rad CFX96 detection system (BioRad, Hercules, CA) at the optimal conditions described below. The total reaction volume was 25 μL, including 12.5 μL SYBR Green Master Mix, 100 nM forward and 100 nM reverse primers, and 10 ng cDNA. The thermal cycler program was 1 cycle of 95˚C for 10 min, followed by 40 cycles of 95˚C for 10 s, 62˚C for 10 s, and 72˚C for 30s. The same reaction for 60S was set up in parallel.

The threshold cycles (C_T) were recorded to calculate the ratio of starting concentration or copy number of *snr*-3 to that of 60S in the original cDNA template, and the relative expression level of *snr*-3 for each stage-specific cDNA sample was normalized to that in the untreated infective L3s by efficiency adjusted $\Delta\Delta C_T$, a relative quantification method (Qiagen Technical Bulletin, 2004).

2.6 DNA constructs

The 242 bp amplicon corresponding to the hookworm DAF-16 bound element located at 3'-UTR of hookworm *snr*-3 was amplified from a pGEM-T EASY construct

containing genomic fragment 2.23 (Gao et al., 2010). The specific forward primer (223-FX: 5′- GATCTCTAGAGCAGCTTTATTCAAGGCGTC-3′, containing restriction site *Xba*I, underlined) and reverse primer (223-RF: 5′-GATCCCGGCCGGCGGAACTCGTTCACCAAAC-3′, containing restriction site *Fse*I, underlined), were incubated with the template in a PCR. The cycling conditions were 2 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 55°C, 1 min at 72°C and a final extension for 6 min at 72°C. The purified amplicon was digested with *Xba*I and *Fse*I, ligated downstream of the luciferase gene on reporter vector pGL4.24 (Promega, Madison, WI) that was cut with the same restriction enzymes to generate pGL4.24-223 construct, and transformed into *Escherichia coli* DH5α competent cells. Similarly, the amplicons from the specific forward primer (223-FXh: 5′-GATCCTCGAGGCAGCTTTATTCAAGGCGTC, containing restriction site *Xho*I, underlined) and reverse primer (223-RH: GATCAAGC TTCGGAACTCG-TTCACCAAAC, containing restriction site *Hind*III, underlined), were digested with *Xho*I and *Hind*III, and ligated upstream of the luciferase gene in reporter vector pGL4.24 to generate construct 223-pGL4.24. A similar strategy was employed to clone Fragment 2.23 into a promoter-less luciferase vector, pGL4.12 (Promega, Madison, WI), to generate constructs 223-pGL4.12 with Fragment 2.23 upstream of luciferase, and pGL4.12-223 with Fragment 2.23 downstream of luciferase. The insert sequence of all the reporter constructs was confirmed by DNA sequencing (McLab, South San Francisco, CA).

2.7 Cell line and cell culture conditions

The fibroblast cell line NIH3T3 from American Type Culture Collection (ATCC) (Manassas, VA) was used. The cells were maintained in DMEM medium (ATCC, Manassas, VA) containing 10% (v/v) calf bovine serum (CBS) (ATCC, Manassas, VA), 100 I. U. /mL penicillin (Mediatech, Herdon, VA), 100 ug/mL streptomycin (Mediatech, Herdon, VA), 2 mM % L-glutamate (Mediatech, Herdon, VA), at 37˚C in an atmosphere of 5% CO_2. The cells were passaged weekly.

2.8 Transient transfection and reporter assays

NIH3T3 cells grown to 80-90% confluency in 24-well plates with 0.5 mL medium were transfected using metafectene (Biontex-USA, San Diego, CA) with 500 ng of reporter constructs, 500 ng of pCMV4-DAF-16 construct 2J (Gao et al., 2009), and 25 ng of pGL4.74 (Promega, Madison, WI), which is a *Renella* luciferase vector and serves as an internal control. The control vector transfection received 500 ng of reporter constructs, 500 ng of pCMV-TAG4 without an insert (Stratagene, La Jolla, CA), and 25 ng of pGL4.74. To analyze serum response, medium was replaced with 20% (v/v) CBS or serum-free medium at 24 hrs post transfection. Medium was removed 24 hrs later and the cells rinsed with PBS (pH7.4) once.

The relative luciferase activity (firefly luciferase activity divided by the *Renilla* luciferase activity) was determined using Dual Luciferase Assay Reagent (Promega, Madison, WI) and GloMax 96 Microplate Luminometer (Promega, Madison, WI). Briefly, 100μL of PBS and 100 μL of Dual-Glo^{TM} Luciferase Reagent (reconstituted) were added sequentially to each culture well, mixed and incubated at room temperature for 10 min. Cell lysate (150 μL) was then transferred into a microtiter plate and firefly

luciferase signals recorded. Dual-Glo™ Stop-Glo™ (75 μL) was added to microplate wells for 10 min and *Renilla* luciferase signals were recorded. All the assays were performed in triplicate.

The data was analyzed using the student's t test. Differences were considered statistically significant when $p<0.05$, and the results were expressed as means ± SD

3. Results

3.1 Gene organization of hookworm snr-3

A. caninum ESTs EXWOO7002HNN28, EWO9RDQ02JAW57 and EXWOO7002GY2EZ were retrieved when 6kb surrounding *Ac*-DAF-16 bound genomic fragment 2.23 was used to search an *A. caninum* EST database available at the Genome Center at Washington University (Gao, et al., 2010). The translated amino acid sequences of those ESTs were homologous to *C. elegans snr-3.* The contig assembled from those ESTs was further searched against *A. ceylanicum* EST databases at Nematode.net and Genbank, and an *A. ceylanicum* 692 bp contig (AE03416) was identified. AE03416 contained the full length SNR-3 open reading frame, encoding a protein of 123 amino acids, a 129 bp 5'-untranslated region, and a 154 bp 3'-untranslated region containing a polyA tail signal and polyA tail (Figure 4.1A).

Comparison of *A. caninum* genomic scaffolds, *A. caninum snr-3* ESTs, and *A. ceylanicum snr-3* contig sequences indicated that hookworm *snr-3* gene spans at least 5.6 kb in the genome, and contains 3 exons and 2 introns (Figure 4.1b). The coding region is solely composed of the first two exons. The introns are located between amino acid 62

(Gly) and 63 (Gln), and within the 3'-UTR (6 nt downstream of the stop codon), respectively. The 5'- and 3'- splicing sites of the exon-intron borders are in agreement with the consensus splicing sequences (5' GT/AC 3') (Padgett et al., 1986). Intron I is 3915 bp long and intron II is 1115 bp long. This organization is quite different from that of *C. elegans snr-3* (Fig 1b), as the latter is only around 1 kb and composed of 4 exons and 3 introns (http://www.wormbase. org).

Interestingly, the genomic Fragment 2.23 that bound *Ac*-DAF-16 overlaps with both intron II and exon III, and the predicted DBE resides at the 3'-end of the exon III (Gao et al., 2010).

3.2 Functional domains and evolutionary relationships of snr-3

A BLASTP search of the non-redundant Genbank database using the conceptual translation of *Ac*-SNR-3 revealed high homology with SNR-3/snRNP D1 proteins from organisms as diverse as *C. elegans* (77% identity and 89% positive over 88 amino acids), *Drosophila melanogaster* (73% identity and 90% positive over 86 amino acids), *Arabidopsis thaliana* (65% identity and 83% positive over 95 amino acids), *Mus musculus* (65% identity and 83% positive over 96 amino acids), *Homo sapiens* (63% identity and 80% positive over 96 amino acids), and yeast *Schizosaccharomyces japonicus* (58% identity and 83% positive over 94 amino acids), suggesting a high level of evolutionary conservation.

As shown in Figure 4.2a, all SNR-3/snRNP D1 proteins including *Ac*-SNR-3 share two characteristic Sm protein motifs, Sm1 and Sm2, as well as a featured C-terminal structural domain, a Gly-Arg repeat. Sm motifs are essential for formation of a

closed ring of the Sm complex (Hermann et al., 1995; Kambach et al., 1999). The Gly-Arg repeat is expected to have strong binding affinity for nucleic acids due to the relative abundance of basic residues. Additionally, it represents one of the Sm-D immunoreactive determinants and shows significant sequence similarity with a region in the Epstein-Barr nuclear antigen-1 that is a non-Sm protein (Rokeach et al., 1992; James et al., 1994).

Phylogenetic analysis using aligned SNR-3 sequences across different species grouped the hookworm SNR-3 most closely with *C. elegans* SNR-3 and more distantly with *A. thaliana* snRNP D1 (Figure 4.2b).

3.3 Developmental expression patterns of snr-3 post infection

To determine the expression pattern of *snr*-3 during the hookworm post infection development, cDNAs were synthesized from *A. ceylanicum* worms from all five categories representing the distinct developmental stages recovered after host infection. *A. ceylanicum* in Golden Syrian Hamsters has been a useful model to mimic the sequelae and physical symptoms that occur during human hookworm infection (Ray et al., 1972; Gupta et al., 1987; Garside et al., 1989). Its ability to infect a small rodent host permits relatively inexpensive *in vivo* time-series experimentation. Furthermore, the identity of the translated amino acid sequences of *A. caninum* and *A. ceylanicum* SNR-3s is as high as 96% (Figure 4.2a).

Since it is almost impossible to get a synchronized population of the *A. ceylanicum* worms (or any hookworms) after they enter the host, recovered worms were categorized based on appearance under a dissecting microscope. Examination of sample

worms from each category under higher magnification indicated that the five categories roughly represented L3, late L3, L4, late L4 and adult stages, respectively.

qRT-PCR using pre-amplified double stranded cDNA indicated that transcription of hookworm *snr-3* was significantly inhibited immediately after infection (Figure 4.3), and remained low or undetectable in all stages up to the early L4.. However at 12 days post infection, levels of *snr-3* were up-regulated in the late L4 stage and dropped again when worms developed into adults, suggesting that active splicing involving *snr*-3 occurs during the L4 stage.

3.4 Genomic fragment downstream of hookworm snr-3 coding region was implicated in the regulation of Ac-snr-3 gene expression.

The 242 bp genomic Fragment 2.23 that spans intron II/ exon III of the hookworm *snr-3* gene displayed strong *in vitro* binding to the DBD of *Ac*-DAF-16, and a DBE was identified in its 3′-terminus (Gao et al., 2010). Theses observations suggested that Fragment 2.23 regulates *snr-3* gene expression. To test this hypothesis, genomic Fragment 2.23 was cloned downstream and upstream of the firefly luciferase gene in pGL4.24 containing a minimal promoter, or the promoter-less luciferase vector pGL4.12, respectively (Figure 4.4).

The resulting luciferase constructs, designated as 223-pGL424, pGL424-224, 223-pGL412, and pGL412-224 were analyzed for their effects on reporter expression by transient co-transfection with *Ac*-DAF-16 in NIH3T3 cells. As seen in panel A of Figure 4.5, both constructs 223-pGL424 and pGL424-224 significantly increased luciferase expression in the presence of *Ac*-DAF-16 compared with empty pGL4.24 vector

transfection control ($P<0.001$). This effect was greater when Fragment 2.23 resided upstream of the luciferase gene, as there was two-fold higher relative luciferse activity for construct 223-pGL424 than for construct pGL424-223. More interestingly, in constructs without a promoter, 223-pGL412 and pGL412-224, Fragment 2.23 was sufficient to drive luciferase expression if *Ac*-DAF-16 was co-transfected (panel B of Figure 4.5). The regulatory activity of Fragment 2.23 does not seem to depend on its position (upstream or downstream) relative to luciferase gene. Taken together, the results showed that *Ac*-DAF-16 interacts with Fragment 2.23 *in vivo* and activates gene transcription.

DAF-16 is negatively regulated by the ILS pathway (Brunet et al., 1999; Lee et al., 2001; Lin et al., 2001). The DBE-mediated transcriptional activity of DAF-16 in response to serum stimulation was reported previously using a mammalian cell culture system (Furuyama et al., 2000; Gao et al., 2009). Therefore we were interested in determining whether or not the serum had similar effects on interaction between *Ac*-DAF-16 and Fragment 2.23. No difference in relative luciferase activities was observed for cells cultured in medium containing 20 % serum compared to medium alone (Figure 5.5), suggesting additional, as yet unidentified factor(s) might co-act with DAF-16 in a native genomic background.

4. Discussion

The forkhead transcription factor DAF-16 has emerged as a master regulator for many important biological processes since it was first identified in *C. elegans* (Kimura et al., 1997; Lin et al., 1997; Ogg et al., 1997). Accumulating studies indicate that DAF-16

mediates its functions by regulating the expression of different target genes (Halaschek-Wiener et al., 2005; McElwee et al., 2003; Murphy et al., 2003; Oh et al., 2006; Lee et al., 2003). The presence of DAF-16 orthologs has been reported from several parasitic nematodes in the past decade (Massy, et al., 2003; Gao et al., 2009; Hu et al., 2010). Parasitologists are particularly interested in DAF-16 and how it regulates its downstream gene network, since DAF-16 plays critical roles in dauer formation and exit in *C. elegans*, which has been adopted as a model for studying the infectious process of parasitic nematodes. The present study described the first characterization of a putative hookworm DAF-16 target gene, *Ac-snr-3*.

Ac-snr-3 encodes a core Sm protein, Sm-D1, in snRNP particles. Sequence analysis indicated that the hookworm *snr-3* gene spans at least 5.6 kb in genome and contains 2 introns and 3 exons. Exon III is solely composed of the 3' untranslated region. The deduced amino acid sequence of *Ac*-SNR-3 is most similar to its counterpart from *C. elegans*. However, the gene structures are quite different for these two species even though they belong to the same phylogenetic clade (Blaxter et al., 1998). The *C. elegans snr-3* gene is shorter in overall length and has much shorter introns. The significance of this finding is not clear. Hookworm SNR-3 protein sequence contained all the important hallmarks of the Sm superfamily including the conserved Sm-motifs and invariable amino acids involved in hydrophobic interactions, salt bridge formation, and hydrogen bondings (Kambach et al., 1999).

In contrast to the mammalian and yeast systems, none of the protein components of the splicing machinery have been identified in hookworm. Nematodes are unique in that their expression of protein-coding genes requires *trans*-splicing in addition to *cis*-

splicing. In *C. elegans*, about 70% of mRNAs are *trans*-spliced to one of two 22 nucleotide spliced leaders (SLs) (Blumenthal, 2005). The processing event of *trans*-splicing is closely related to *cis*-splicing (Brumenthal et al., 1988). Hookworm presumably employs a similar gene splicing mechanism, as SL sequences have been identified from *A. caninum* mRNAs (Hawdon et al., 1995).

The expression changes of Sm proteins during worm development are potentially important in processing pre-mRNA into mature mRNAs, thereby affecting the size and complexity of stage-specific proteosomes. For this reason, we generated a transcriptional profile of hookworm *snr-3* using qRT-PCR data from the post-infection stages. Unexpectedly, hookworm *snr*-3 transcription was turned off immediately and remained at a low level for an extended time after the infective L3 entered the host small intestine. Microscopic examination under high magnification showed that the morphology and behavior of the L3s at this stage was almost identical to those of the infective L3s. However within the worm, dramatic biochemical changes obviously have been triggered in preparation for resumption of parasitic development once host signals are received, and *snr*-3 represents an early responding gene that is negatively regulated by those signals. Alternatively, sufficient SNR-3 proteins for post-infection development may have been stored in the L3 for exit from the arrested state and the subsequence development. However, the possibility that different, as yet unidentified, isoforms of hookworm *snr-3* might be transcribed in early post-infection stages cannot be excluded. The transcription of hookworm *snr*-3 peaked at the late L4 stage with greater than two-fold increase compared with the control infective L3s, suggesting an active splicing event might be required for the worm's maturation. Since the data were obtained from whole

worms, the transcriptional increase at this stage in specific tissues or specific cells might be even greater.

The distinct temporal transcription profile of hookworm *snr*-3 after infection indicated that this gene was subject to tight regulation. Specifically, its drastic down-regulation during early infective stages suggested that this gene was positively regulated by DAF-16, assuming that DAF-16 was negatively regulated by ILS during infection. A 242 genomic fragment (Fragment 2.23) containing a reverse DBE (5′-AACAAATA-3′) at the 3′-terminus of the *Ac-snr*-3 gene was found to strongly interact with *Ac*-DAF-16 DBD in binding assays, as evidenced by high occurrence rate in our isolated genomic DNA (Gao et al., 2010). We hypothesized that this genomic fragment acts as a *cis*-regulatory element affecting expression of the *Ac-snr*-3 gene. Reporter constructs containing Fragment 2.23 upstream or downstream of the luciferase gene were co-transfected with *Ac-DAF*-16 into mammalian cells and the relative luciferase activities measured. The results confirmed that Fragment 2.23 could function *in vivo* in the presence or absence of the minimal promoter to drive expression of the reporter gene. Since the transcriptional activity does not depend on its relative position to the reporter gene, it suggests that Fragment 2.23 might simultaneously influence the expression of both upstream and downstream genes in the genomic background. Shared *cis*-regulatory elements are not unusual in strongly correlated gene pairs or gene clusters and have been implicated in coordinating gene expression (Jegga et al., 2002; Shostak et al., 2004; Xu et al., 2006; Armstrong & Chamberlin, 2009). Further characterization of neighboring genes will undoubtedly unravel more information about the function of the regulatory elements within this genomic fragment. It is also worth noting that the mammalian cell culture

system provides a simplified physiological setting for activity studies of gene regulatory elements, and the observed positive regulation mediated by Fragment 2.23 using this system might not be indicative of the true regulatory direction in the physiological setting of the hookworm.

Previous findings showed that serum depletion greatly up-regulated reporter gene expression mediated by the consensus DBE, which was consistent with the conclusions that the critical factor for the determination of transcriptional activity of DAF-16 orthologues was their cellular location (Gao et al., 2009). When similar experiments were performed in the current study, Fragment 2.23 was not responsive to serum stimulation. These results were not surprising since Fragment 2.23 is a much longer DNA sequence compared with the 8-bp consensus DBE, and scanning the sequence for transcription factor binding sites using the Promo program showed that it potentially contains more complicated regulatory information (Figure 4.6). Under such circumstances, Fragment 2.23 might be capable of recruiting additional protein factors that modify the interaction between Fragment 2.23 and *Ac*-DAF-16. For example, a well-documented forkhead protein co-factor is C/EBPβ (CCAAT-enhancer-binding protein beta), and its binding motif is highly over-represented in Fragment 2.23 (O'Brien et al., 1994; O'Brien et al., 1995; Durham et al., 1999; Nasrin et al., 2000). Transcriptional stimulation resulting from interaction between FOXO and C/EBPβ has been demonstrated to be greater than additive (Christian et al., 2002). Another regulatory element found in Fragment 2.23 is HSF-1 (Heat Shock Factor-1) binding element (HSE). HSEs co-exist with DAF-16 binding elements in several *C. elegans* promoters, where together they activate expression of specific genes such as the *shsp* (Smalll Heat-Shock Proteins) (Hsu et al.,

2003). The data obtained from our study also suggest that using a native genomic fragment might help to identify new co-factors interacting with *Ac*-DAF-16 and provide insights into the mechanisms that are relevant to how signaling events involving DAF-16 are integrated with the transcriptional machinery.

Conclusions

In conclusion, we report the gene structure of a hookworm DAF-16 putative target, *Ac-snr*-3, which was identified by genome wide screening for *Ac*-DAF-16 binding elements. *Ac-snr*-3 encodes a core small nuclear ribonucleoprotein and the predicted amino acid sequence shares considerable homology with other Sm-D1 proteins. Hookworm *snr*-3 mRNA levels were significantly down-regulated immediately after host entry and up-regulated at the late L4 stage, indicating a potential role in L4 development. The 3'-terminal genomic fragment of this gene displayed *Ac*-DAF-16-dependent *cis*-regulatory activity. Further investigations of the regulation of hookworm *snr*-3 expression will illuminate the physiological role of *Ac*-DAF-16.

A

NTCCTCCAACCCTCGAAATTTTGCCCTCTCTAAAGGGTCACAAAACCTGGAGCTCCCCCCGCGGTGGCGCCCCTTCTAGA

M K L V R F L R K L

AATTAGGGGATCCCCCGGGTTGCAGAGATTTGGCACGAGGGTTTGAGAGATGAAGTTGGTTCGTTTCTTGAGGAAACTGT

Start codon

CCCATGAGTCAGTCCCCATTGACCTGAAGAATGGGACGCTCGTGCAGGGTGGTATCATTGGTGTCGATGTTGCTATGAAC

M H L R S V K M S L K N K D P I N L D S L S I R G N N

ATGCACCTTCGCTCGGTGAAGATGTCGCTGAAGAATAAGGACCCTATCAATCTCGACTCGCTTTCTATTCGAGGCAACAA

I R Y I I L P D S L P L D T M L I D D E P R K E E T

CATTCGGTACATAATTCTGCCGGATTCCCTCCCTCTCGATACCATGCTCATTGATGACGAGCCTCGCAAGGAGGAAACCA

R A R A G A R G R G R G G P R G R G R G G P R G R G G

GAGCTCGTGCAGGTGCTCGTGGTAGGGGACGTGGCGGCCCTCGTGGAAGGGGCAGGGGTGGGCCACGTGGACGTGGTGGA

P R G R G R *

CCACGCGGTCGAGGTCGCTAGGTTAGGTGTTGCTGGGTTCCGCTCTGGCGAAAACATCCTGTCGGGCTATCGCTCATGTT

Stop codon

TTTTGTCTCTTCGAGTATATCCGAATTTTCAAGGTTGTTATTTGTTTGGTGAACGAGTTCCGGATCTCTGGATAAATAAA

PolyA signal

TCTTTTTTGGCTATAAAAAAAAAAAAAAA

PolyA tail

B

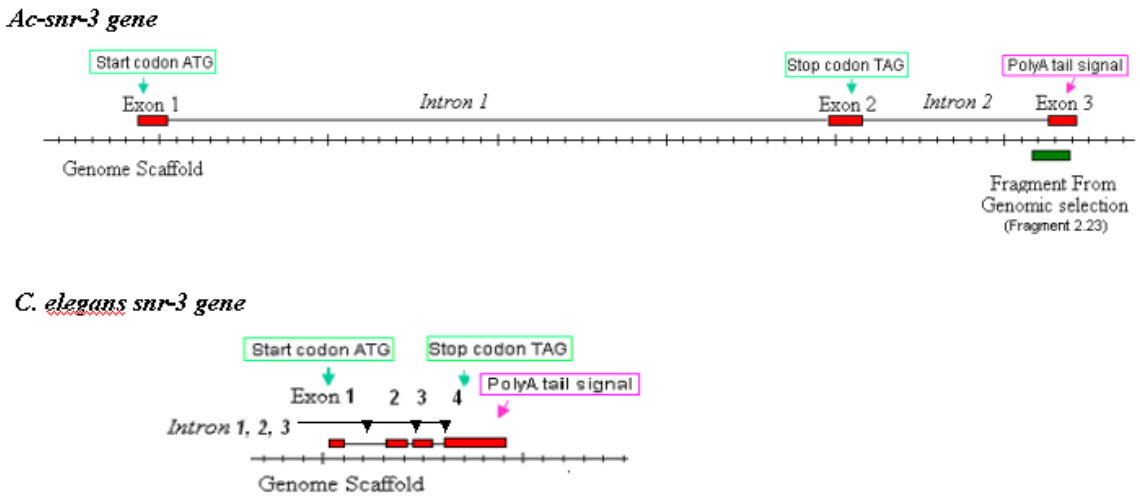


Figure 4.1

Figure 4.1 *Ac*-snr-3 gene. (A) cDNA sequence(AE03416) and translated amino acid sequence of *Ac-snr*-3. (B) Genomic organization of hookworm *Ac-snr*-3 compared with its *C. elegans snr-3* (www.wormbase.org). Cassettes represent exons (red), lines represent introns, and scaled lines represent genome scaffolds. The length of the cassettes and lines is proportional to the length of the corresponding DNA fragments.

A

Sm1
A. ceylanicum SNR-3 1 MKLVRFLRKLSHESVPIDLKNGTLVQGGIIGVDVAMNMHLRSVKMSLKNKDPINLDSLSI
A. caninum SNR-3 1 MKLVRFLMKLIHESVTIELKNGTLVQGGIIGVDVAMNMHLRSVKMSLKNKDPINLDSLSI
C. elegans SNR-3 1 MKLVRFLMKLSHETVNIELKNGTQVSGTIMGVDVAMNTHLRAVSMTVKNKDPVKLDTLSI
Sm2
Gly-Arg repeat
A. ceylanicum SNR-3 61 RGNNIRYIILPDSLPLDTMLIDDEPRKEETRARAGA-RGRG-RGGPRGRGRGGPRGRGGP
A. caninum SNR-3 61 RGNNIRYIILPDSLPLDTMLIDDEPRKKAARARAGA-RGRG-RGGPRGRGRGGPRGRGGP
C. elegans SNR-3 61 RGNNIRYIILPDPLALDTLLIDDEPRKKARAARAGASRGRGGRGGMRGRGGRGGRGRGGP
A. ceylanicum SNR-3 119 RG--RGR
A. caninum SNR-3 119 RG--RGR
C. elegans SNR-3 121 RGGGRGR

B

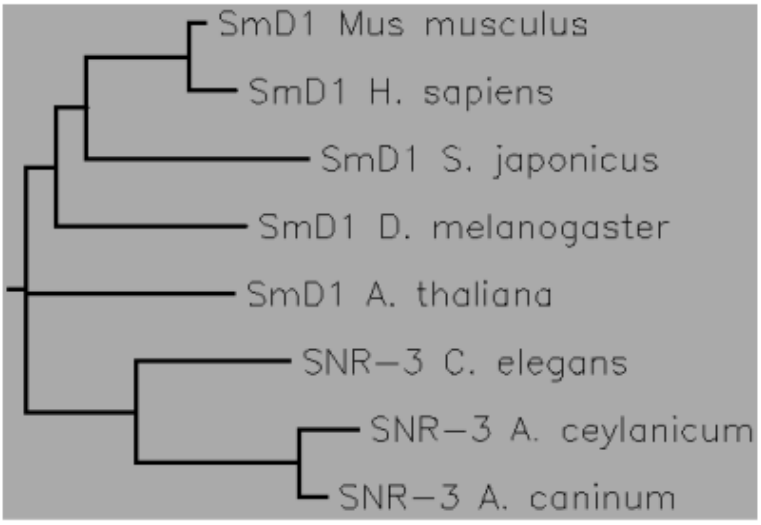

SmD1 Mus musculus
SmD1 H. sapiens
SmD1 S. japonicus
SmD1 D. melanogaster
SmD1 A. thaliana
SNR−3 C. elegans
SNR−3 A. ceylanicum
SNR−3 A. caninum

Figure 4.2

Figure 4.2 Phylogenetic relationship between *Ac*-SNR-3 and selected Sm-D1 proteins. (A) Alignment of hookworm SNR-3s and *C. elegans* SNR-3. Shading indicates residues that are identical (black) or conserved (grey). Protein sequences were aligned using Clustal W software and displayed using BOXSHADE software located on the Swill EMBnet server. (B) Neighbor joining tree of Sm D1s. Amino acid sequences were aligned using Clustal W software and the tree was drawn using Phylip's software. Accession numbers: SmD1 *Schizosaccharomyces japonicus*, XP_002171442; SmD1 *Mus musculus*, EDL22993; Sm D1 *Homo sapiens*, NP_008869; SmD1 *Drosophila melanogaster*, NP_524774; SNR-3 *Caenorhabditis elegans*, NP_495306; SmD1 *Arabidopsis thaliana*, NP_187416

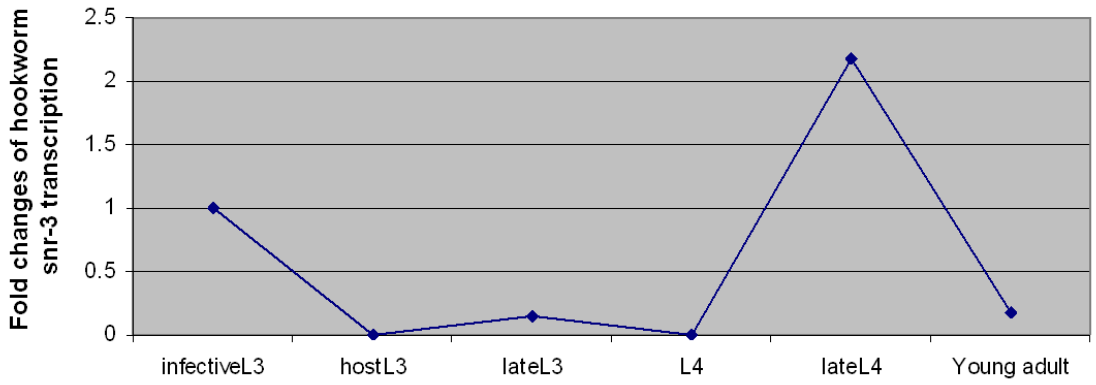


Figure 4.3 Transcription of hookworm *snr-3* during post infection developmental stages. The *x*-axis represents developmental stage, and *y*-axis represents the fold change in *Ac-snr-3* transcription normalized to the hookworm 60s reference gene.

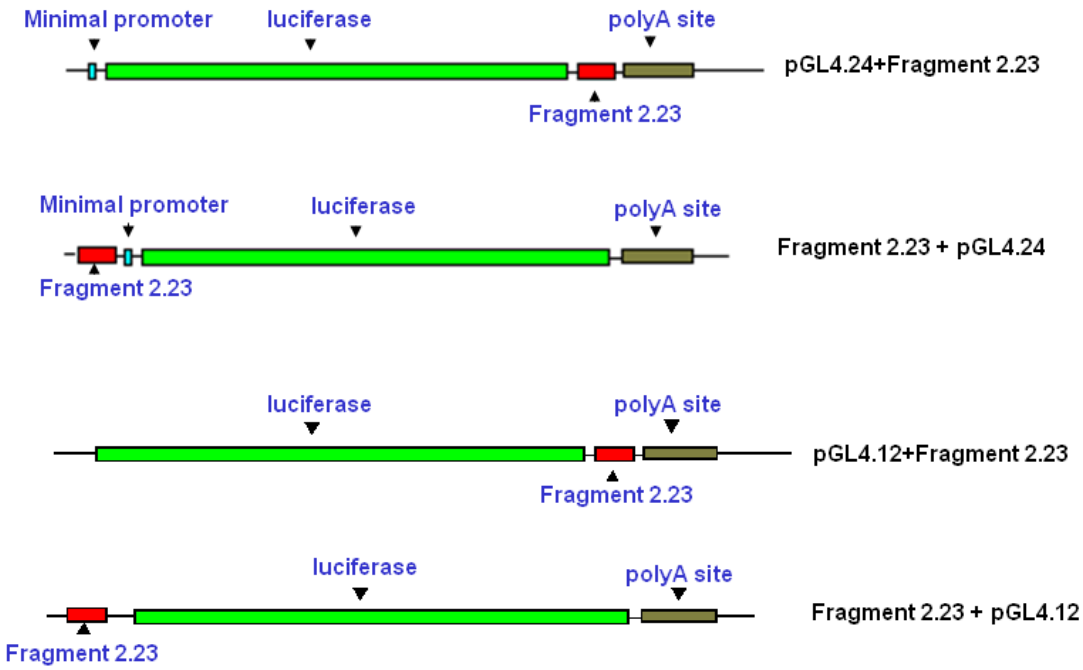


Figure 4.4 Schematic representation of luciferase constructs. Fragment 2.23 (red) was cloned into different positions relative to reporter luciferase gene (green) in vector pGL4.24 and pGL4.12. Blue and grey cassettes represent the minimal promoter (blue) and poly A site of the vector, respectively.

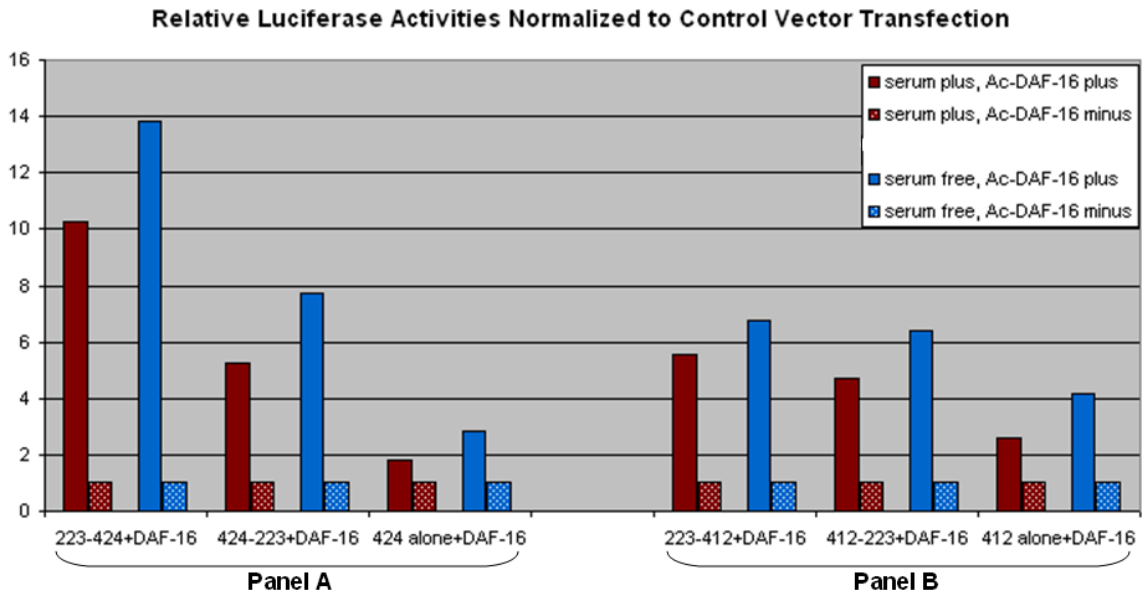


Figure 4.5 *Ac*-DAF-16 drives expression of luciferase from Fragment 2.23. The *x*-axis represents luciferase constructs: Panel A, pGL424 luciferase constructs; Panel B, pGL4.12 luciferase constructs. The *y*-axis represents the relative luciferase activity normalized to control vector transfection.

A

Factors predicted within a dissimilarity margin less or equal than 15 % :

0 MYB2 [T02536]	1 HSF1 [T04394]	2 AR [T00040]	3 c-Myb [T00137]	4 TAF [T00778]	5 RAR-gamma [T00720]	6 RC2 [T00724]	7 MF3 [T00507]	
8 TGGCA-binding protein [T00832]	9 C/EBPalpha [T00108]	10 C/EBPbeta [T00459]	11 v-ErbA [T00886]	12 AP-3 [T01150]	13 C/EBPalpha [T00107]	14 HNF-3beta [T02344]	15 HNF-3 [T02277]	
16 GR-alpha [T00337]	17 NFI/CTF [T00094]	18 HNF-3alpha [T00371]	19 Cutl1 [T02042]	20 HOXA5 [T00377]	21 POU1F1a [T00691]	22 C/EBPalpha [T00104]	23 C/EBPdelta [T00109]	
24 HNF-3beta [T03256]	25 Nkx6-2 [T02050]	26 PF1 [T04784]	27 unc-86 [T01882]	28 HNF-3alpha [T02512]	29 FOXM1b [T02516]	30 FOXM1a [T02517]	31 HNF-3 [T00370]	
32 HNF-1C [T01951]	33 POU2F1 [T00643]	34 HOXD8 [T01426]	35 HOXD8 [T01754]	36 HOXD9 [T01424]	37 HOXD10 [T01425]	38 HOXD9 [T01755]	39 HOXD9 [T01756]	
40 HOXD10 [T01757]	41 HOXD10 [T01758]	42 MATa1 [T00488]	43 RXR-alpha [T01345]	44 GT-1 [T00339]	45 FOXP3 [T04280]	46 GAL4 [T00302]	47 XBP1 [T03707]	
48 Pax-6 [T00682]	49 GA-BF [T00297]	50 Hb [T00395]	51 XBP-1 [T00902]	52 AP-3 (2) [T00039]	53 TCF-2 [T01110]	54 p53 [T00671]	55 Mad [T04378]	
56 CBF2 [T05834]	57 Nrf2:MafK [T05666]	58 GR-beta [T01920]	59 C/EBPbeta [T00017]	60 GR [T00333]	61 GAMYB [T02679]	62 PR B [T00696]	63 PR A [T01661]	
64 Dl [T00196]	65 HNF-3beta [T01049]	66 CP2 [T00152]	67 NF-1 [T01298]	68 VDR [T00885]	69 CREMtau [T01309]	70 CREMtaualpha [T01602]	71 CREMtau1 [T02108]	
72 CREMtau2 [T02109]	73 myogenin [T00528]	74 RFX1 [T01673]	75 MyoD [T00526]	76 DBP [T00183]	77 Alfin1 [T04733]	78 USF2 [T00878]	79 AREB6 [T00625]	
80 E47 [T05421]	81 HNF-3beta [T02513]	82 GATA-1 [T05705]	83 SPF1 [T03975]	84 C/EBPgamma [T02028]	85 ROX1 [T01286]	86 HNF-4alpha1 [T00372]	87 PR B [T00697]	
88 ABF1 [T00056]	89 STE12 [T00772]	90 POU2F1 [T01031]	91 FACB [T02841]	92 PEA3 [T00685]	93 MCB1 [T06035]	94 MCB2 [T06036]	95 C1 (long form) [T01592]	
96 Tll [T00789]	97 DSXF [T00955]	98 DSXM [T00956]	99 Nkx2-1 [T00856]	100 Pax-5 [T00070]	101 TBP [T00794]	102 TFIID [T00820]	103 POU1F1b [T01516]	
104 POU1F1c [T01902]	105 Adf-1 [T00008]	106 TBP [T00798]	107 HSF1 (long) [T01042]	108 HSF1 (short) [T02104]	109 MAC1 [T01265]	110 GAGA factor [T00301]	111 HNF-1A [T00368]	
112 HNF-1B [T01950]	113 DBP [T04875]	114 Crx [T03461]	115 Ftz [T00295]	116 Nkx2-1 [T00857]	117 SF-1 [T04014]	118 LAC9 [T00458]	119 CAT8 [T03227]	
120 TRM1 [T05311]	121 Fra-1 [T01208]	122 PXR-1:RXR-alpha [T05671]	123 c-Ets-2 [T01397]	124 POU2F2 (Oct-2.1) [T00646]	125 MED8 [T03491]	126 C/EBPalpha [T00105]	127 C/EBPbeta [T00581]	
128 C/EBPgamma [T00216]	129 NF-E4 [T00560]	130 PEA3 [T00684]	131 STAT4 [T01577]	132 E2F-1 [T01542]	133 c-Ets-1 [T00112]	134 c-Ets-2 [T00113]	135 f(alpha)-f(epsilon) [T00287]	
136 NF-1 [T00537]	137 NF-1 [T00539]	138 ENKTF-1 [T00255]	139 GATA-2 [T01302]	140 GATA-1 [T00306]	141 YY1 [T04970]			

B

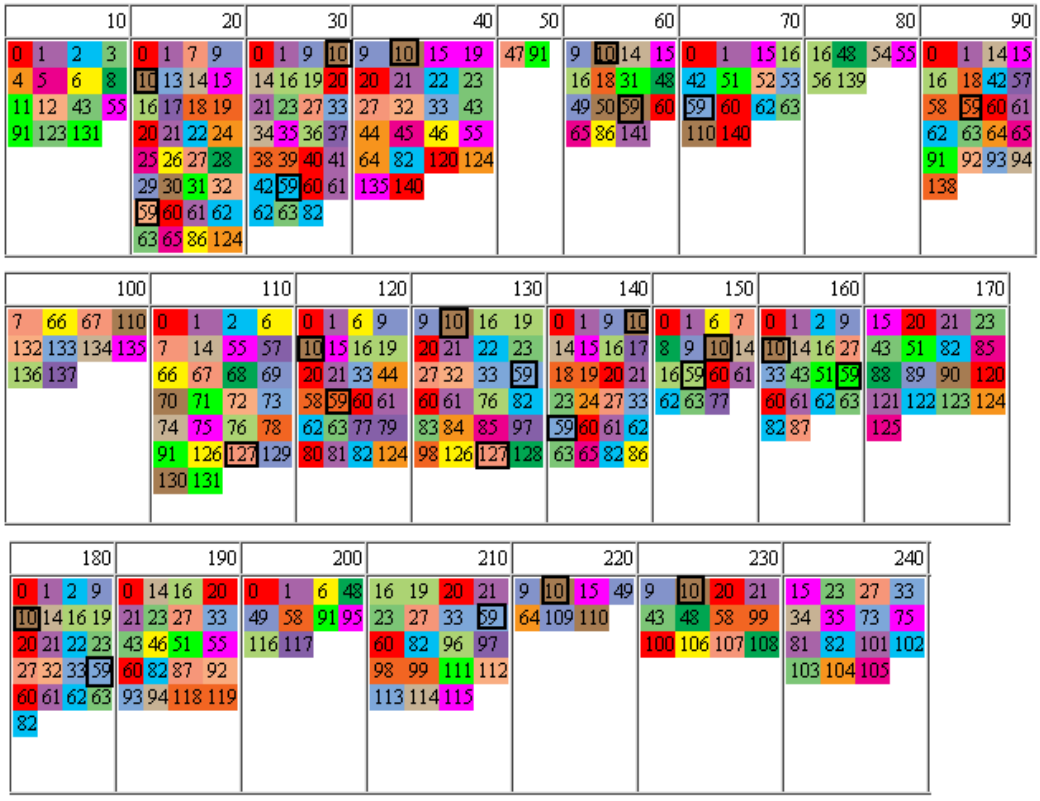


Figure 4.6 Predicted transcription factor binding sites located on Fragment 2.23. (A) List of transcription factors with the binding sites located on Fragment 2.23. Numbers were assigned to each individual transcription factor. C/EBPβ and HSF-1 binding sites are boxed (B) Positions of the predicted transcription factor binding sites labeled as colored numbers. Each panel represents 10 bp. Multiple binding sites of C/EBPβ are boxed

Chapter 5. General Discussion

The studies presented in this dissertation focus on a key ILS pathway transcription factor, *Ac*-DAF-16/FOXO, its response elements, and downstream gene targets in hookworms. Hookworm is a parasitic nematode closely related to the free-living nematode *C. elegans*. The developmentally arrested dauer stage of *C. elegans* is analogous to the infective L3 stage of hookworms (Hawdon & Schad, 1991; Hotez et al., 1993). The hookworm infectious process and subsequent development is of great interest to parasitologists from the perspective of disease control and prevention, however, these areas are still poorly understood.

Free-living *C. elegans* has served as an important model for biological studies of parasitic nematodes including hookworm. The ILS pathway was originally described in *C. elegans* when the components of this pathway were found to regulate worm lifespan and the switch between dauer and reproductive development (Kimura et al., 1997; Dorman et al., 1995; Paradis et al., 1999; Paradis & Ruvkun, 1998; Ogg & Ruvkun, 1998; Hertweck et al., 2004; Lin et al., 1997; Ogg et al., 1997). Later, it became apparent that ILS had important roles in many other biological processes (Braechman et al., 2001; Taguchi & White, 2008; Tissenbaum & Ruvkun, 1998). Inhibitor studies with hookworms using *in vitro* activation assays demonstrated that ILS was also essential for hookworm development (Brand and Hawdon, 2004). The pivotal regulatory molecule in ILS pathway is the forkhead transcription factor DAF-16/FOXO (Lin et al., 1997; Ogg et al., 1997).

The current view about the mechanism of DAF-16/FOXO function is that an insulin- or insulin like growth factor- mediated shuttling system regulates its intracellular localization (nuclear or cytosol), thereby activating or terminating the corresponding transcriptional activities. Central to this shuttling system are a series of post-translational modifications (mainly phosphorylation) of DAF-16/FOXO and its association with scaffold proteins such as protein 14-3-3 (Hertweck et al., 2004; Lin et al., 2001). Screening a synthetic random oligonucleotide pool using recombinant *C. elegans* DAF-16/FOXO and its mammalian homologues identified an 8-bp consensus sequence (DBE), 5'-TTGTTTAC-3', with high binding affinity to all FOXO proteins (Furuyam et al., 2000). This discovery of the DBE, together with other advantages of *C. elegans* such as its complete genome sequence, defined cell lineage and powerful genetics, has been widely employed to investigate the complex regulatory network for *C. elegans* DAF-16/FOXO (Murphy et al., 2003; Lee et al., 2003; Oh et al., 2006).

Although DAF-16/FOXO and its response elements and downstream gene targets have been extensively studied in free-living nematode *C. elegans*, the findings from our studies are significant for the following three reasons. First, hookworm has a far more complicated life cycle involving obligatory developmental arrest and gender differentiation, and a much larger genome (~ 350 Mb) compared with the non-parasitic *C. elegans* (~100 Mb) (Hawdon & Shad, 1991; Hotez et al., 1993; The *C. elegans* Sequencing Consortium, 1998; Abubucker et al., 2008). Therefore, it is important to determine the roles and mechanisms of action of regulators like DAF-16/FOXO within the hookworm physiological background. Second, information from hookworm DAF-16/FOXO and its cognate network of gene targets will allow the relationships within the

DAF-16/FOXO super-family to be assessed and will enrich our understanding of regulatory evolution. Finally, a complete functional description of gene targets of hookworm DAF-16 and how those downstream effectors are coordinated throughout hookworm development might unravel molecular and biochemical events underlying the transition to parasitism and the invading hookworm's interaction with the definitive host, thereby helping to identify new targets for drug and vaccine development

EST data from the early stages of the hookworm genome project provided evidence that DAF-16/FOXO orthologs were present in hookworm (Mitreva et al., 2005). Therefore, the first aim of this research (Chapter 2) was to clone the *Ac-daf*-16 gene and express it as a functional protein, which would allow further mechanistic investigations. Specific primers were designed from the EST sequence to amplify the *daf-16* open reading frame from both *A. ceylanicum* and *A. caninum* cDNAs. The deduced amino acid sequence of hookworm DAF-16/FOXO indicated conservation of FOXO functional domains including the featured forkhead or "winged helix" DNA binding domain, Akt/PKB phosphorylation sites, and 14-3-3 binding sites. Specifically, it's DNA binding domain shares high identity with that of *C. elegans* DAF-16b (98%). A unique aspect of *Ac*-DAF-16/FOXO is the alanine-rich stretch of unknown function residing downstream of the forkhead domain. Reverse-transcription PCR showed that *Ac*-DAF-16 was expressed in all hookworm developmental stages.

For in-depth functional studies of *Ac*-DAF-16, a polycolonal antiserum targeting a 163 amino acid peptide from its C-terminus was raised in New Zealand white rabbits that detected *Ac*-DAF-16 in soluble extracts of *A. caninum* L3. Because the hookworm cell lines and genetic mutants were not available, we employed mammalian cell lines to test

the binding affinity and transcriptional activity of *Ac*-DAF-16. We reasoned that the ILS components within mammalian cells would interact with *Ac*-DAF-16 due to the considerable similarity between ILS and FOXO proteins across different species. The open reading frame of *Ac*-DAF-16 was cloned into mammalian expression vector pCMV-TAG4, and the expression of the FLAG-tagged fusion protein was successfully achieved in HEK293 cells, as demonstrated by Western blot using anti-FLAG antibody and anti-*Ac*-DAF-16 antiserum. A streptavidin bead pull-down assay indicated that the expressed recombinant *Ac*-DAF-16 from the mammalian cell culture specifically recognized and strongly bound to the consensus DBE *in vitro*. When a luciferase reporter construct containing an upstream 6× DBE was co-transfected with this *Ac*-DAF-16 expression construct into NIH3T3 cells, serum depletion significantly stimulated luciferase activity, whereas serum supplementation did not. These results suggested that *Ac*-DAF-16 interacted with DBE *in vivo* and also supported the conclusion that *Ac*-DAF-16 was negatively regulated by ILS. Furthermore, hookworm 14-3-3 protein (*Ac*-FTT-2) was also cloned and expressed in our lab, and shown to interact with *Ac*-DAF-16 in a phosphorylation-site dependent way (Kiss et al., 2009). Apparently, the molecular basis behind *Ac*-DAF-16 fits in the generalized picture for FOXO protein family.

In order to further elucidate the role of *Ac*-DAF-16 in hookworm development and infection, we next sought to isolate its genomic binding sites and identify direct target genes (Chapter 3). *In vitro* genomic selection using recombinant *Ac*-DAF-16 DBD (*rAc*-DAF-16-DBD) was employed to enrich for a small set of the native hookworm genomic fragments with high affinity to *Ac*-DAF-16. Since the selection processes were carried out *in vitro*, it allowed the *Ac*-DAF-16 bound elements to be identified without

influences from complex intracellular environments. One hundred one amino acids of the *Ac*-DAF-16 DBD was expressed as a soluble FLAG-tagged peptide in *E. coli*, and its binding activity was confirmed by *in vitro* DBE pull-down assays. Hookworm genomic fragments prepared by restriction enzyme digestion were mixed with purified r*Ac*-DAF-16 DBD and the bound DNA-peptide complex precipitated with anti-FLAG agarose beads. Genomic fragments from the 4th round of selection were cloned and sequenced. A four-round control selection / amplification was also established using the digested hookworm genomic DNA alone. A total of 274 high quality sequences from *Ac*-DAF-16 DBD selection and 190 from the control selection were analyzed and compared. The results indicated that *Ac*-DAF-16 DBD bound to a set of specific hookworm genomic fragments, as evidenced by the high occurrence rate of some recovered sequences.

Visual inspection determined that a large number of the *Ac*-DAF-16 DBD bound sequences contained low-complexity microsatellite regions characterized by the trinucleotide repeat "GTT", the biological significance of which was not clear. The remaining sequences were queried against the draft assembly of the hookworm genome, and 24 unique sequences were successfully mapped. Five *Ac*-DAF-16 DBD bound elements were over-represented; and among them, four contained the consensus DBE or DBE-like elements, which have been identified in the promoter regions of DAF-16 gene targets in *C. elegans* and *D. melanogaster* (Lee et al., 2003), indicating a conserved function of *Ac*-DAF-16. These findings demonstrated that biologically functional genomic fragments could be isolated using a purified DBD. Nearly half of the isolated fragments contained no DBE or DBE-like elements, suggesting that unidentified *Ac*-DAF-16 binding elements might exist in the hookworm genome. To test this, the

remaining fragments were subject to further motif searches. A 6 bp over-represented motif, 5′-GAC/GAA/ TG -3′, was identified. However, when examined using streptavidin bead pull-down assays, the amplicon containing this new element, but not the 6 bp element alone, was recognized by *Ac*-DAF-16 DBD, suggesting that the surrounding region of this motif might contribute to its interaction with the *Ac*-DAF-16 DBD. It is worth noting that the dataset used for motif analysis was relatively small, and therefore the significance of this motif is subject to further evaluation.

Using the available *A. caninum* transcriptome data, a detailed analysis of the transcripts linked to the *Ac*-DAF-16 binding sites was performed. A total of 22 transcript contigs were returned within a 6 kb region surrounding those 24 *Ac*-DAF-16 DBD bound fragments. Not all the bound elements were linked to the transcripts within this 6 kb ranges, therefore an extended search might be necessary. Functions of the linked transcripts were predicted based on homology and examination of the digital expression profile. As expected, *Ac*-DAF-16 was implicated in diverse biological processes throughout the hookworm developmentm, as is its *C. elegans* counterpart, even though the gene list generated in our study did not overlap with those from *C. elegans* studies. Notably, some identified transcripts did not match any identified homologs across all available protein databases, suggesting they are unique *A. caninum* genes.

The location of the *Ac*-DAF-16 DBD bound elements relative to the linked transcripts were determined by sequence alignment of the identified transcripts with the genome scaffolds. We surprisingly found that *Ac*-DAF-16 binding sites, unlike *C. elegans* DAF-16 binding sites (Oh et al., 2005), resided at variable locations relative to the gene transcripts, indicating that *Ac*-DAF-16 might regulate gene expression through

alternative mechanisms. However, it is worth noting that the locations of many *Ac*-DAF-16 binding sites relative to its transcripts were not determined due to incomplete hookworm genome annotation and the missing information about the 5′ end of hookworm transcripts in general. In addition, the *in vitro* genomic selection in this study used only *Ac*-DAF-16 DBD as a molecular bait, and the full length molecule might have a different binding sites preference.

Finally, we selected hookworm genomic Fragment 2.23 and its linked gene for more comprehensive investigation (Chapter 4). Fragment 2.23 was recovered 12 times in our selection cycles, ranking second in all the isolated genomic fragments, suggesting an extraordinary high binding affinity to *Ac*-DAF-16 DBD. Also, the complete gene sequence was available in the early genome assembly, which allowed us to completely characterize its structure. Fragment 2.23 contains a reverse DBE and the linked transcript encodes a small nuclear ribonucleoprotein D1, *Ac*-SNR-3, which is a core Sm protein in RNA splicing machinery (Chapter 3). The expression of Sm proteins potentially affects the intracellular mRNA pool, and the corresponding proteome features.

In contrast to the mammalian and yeast system, hookworm and other nematodes are particularly interesting because the maturation of their pre-RNAs requires both *cis*-splicing and *trans*-splicing. Presumably, the mechanisms of *cis*-splicing and *trans*-splicing are similar (Brumenthal et al., 1988; Brumenthal, 2005). Until now, none of the protein components of hookworm splicing machinery have been reported although spliced leader sequences have been identified on hookworm mRNAs (Hawdon et al., 1995).

EST sequences from *A. caninum* combined with those from *A. ceylanicum* were assembled to determine the structure of the full-length *Ac-snr*-3 gene. The *Ac-snr*-3 gene spans at least 5.6 kb in the genome, containg 3 exons and 2 introns. The coding region is composed of exons I and II only. The previously identified genomic Fragment 2.23 overlapped with the junction of intron II and exon III, and the reverse DBE resides in 3′-UTR of *Ac-snr*-3. The amino acid sequence encoded by *Ac-snr*-3 contained all the featured motifs for Sm-D1 proteins, such as conserved Sm1 motif, Sm 2 motif, and an arginine-glycine stretch at its 3′-terminus.

The expression profile of *Ac-snr*-3 across post-infection development was examined. The results revealed that *Ac-snr-3* transcription was significantly down-regulated relative to the control during most of the post-infection period, except for the late L4 stage. Perhaps enough *Ac*-SNR-3 protein is stored before initiation of the infectious process to meet the immediate post-infection requirements. However, when worms prepare to develop into adults, a high level of splicing activity might be required. Since the 3′-terminus of *Ac-snr*-3 contains a genomic fragment (Fragment 2.23) that was previously identified to interact with *Ac*-DAF-16, we next examined if this fragment was involved in regulation of *Ac-snr*-3 gene transcription. Reporter gene assays confirmed that Fragment 2.23 was trancriptionally active in the presence of *Ac*-DAF-16, and the transcriptional activity was independent of its location relative to the reporter gene, although higher activity was seen when it resided upstream of the reporter gene. In contrast to data obtained in Chapter 2, cells expressing Fragment 2.23 were not responsive to serum. Further sequence analysis showed that, in addition to the 8 bp DBE, Fragment 2.23 contains much more potential regulatory information, suggesting it might

nucleate more complex transcriptional machinery together with other protein factors. Accordingly, the gene transcription mediated by interactions between Fragment 2.23 and *Ac*-DAF-16 could be affected.

In conclusion, the data generated from these studies yielded the first information about a key component in hookworm ILS, *Ac*-DAF-16, as well as its native genomic binding sites and a subset of its direct gene targets. These findings provide valuable insights into the critical molecular events relative to hookworm infection, parasitism and development.

Since the activity assays for *Ac*-DAF-16 and its binding elements in our studies were all carried out in mammalian cell culture system, the actual responses in hookworm might be quite different since worm and mammalian cells could have different sets of cell surface receptors and intracellular co-factors. Furthermore, removal of the *cis*-regulatory element from its native gene context might results in non-physiological responses in cell culture-based reporter system. At present, the closest heterologous system for functional assessment of hookworm molecules is transgenic lines of *C. elegans*. Our lab has created transgenic *C. elegans* worms expressing *Ac*-DAF-16, and phenotypic characterization is ongoing. Additional transgenic approaches can also be used to assess function of the identified regulatory elements. For instance, expression of the GFP constructs containing *Ac*-DAF-16 bound elements in transgenic *C. elegans* worms might reveal the temporal and spatial expression pattern of its linked gene.

The relatively small inventory of *Ac*-DAF-16 gene targets identified from the present studies made it difficult to compare the hookworm DAF-16 downstream network with the one from *C. elegans* or any other species. This could be because different

methods were employed to identify the *C. elegans* DAF-16 downstream gene network. A potential improvement could be accomplished by reducing the stringency of selection conditions and screening more colonies in the *in vitro* genomic selection procedure. Alternatively, multiple strategies should be attempted to identify hookworm DAF-16 downstream gene targets so that a comprehensive comparison could be established.

Much remains to be understood about my results. For example, the "GTT" repetitive microsatellite sequences represent a large portion of our recovered *Ac*-DAF-16 bound elements. The characterization of those elements will undoubtedly be expedited by ever increasing pace of the hookworm genome project. Identification of co-acting protein factors for hookworm DAF-16 would be another interesting subject. Native genomic fragments have been isolated, and it stands to reason that those fragments could be used to isolate unidentified proteins that interact with hookworm DAF-16. A complete description of the transcriptional machinery involving hookworm DAF-16 will help to unravel its mechanism.

The methodology in my studies can be extended to other cellular signaling transduction pathways in hookworm. Parasites simultaneously receive multiple stimuli at any give time. How these stimuli are translated and integrated at the molecular level affects decision making processes within the parasites. Transcription factors act at the critical interface of converting environmental stimuli into the specific gene expression profile. Future characterization of hookworm transcription factors and large-scale investigation of hookworm signaling transduction pathways and their downstream gene networks will provide insights into their roles in hookworm development and host interaction, and lead to novel intervention strategies in the hookworm life cycle.

References

Abubucker, S., Martin, J., Yin, Y., Fulton, L., Yang, S.P., Hallsworth-Pepin, K., Johnston, J.S., Hawdon, J., McCarter, J.P., Wilson, R.K., Mitreva, M. (2008) The canine hookworm genome: analysis and classification of *Ancylostoma caninum* survey sequences. Mol Biochem Parasitol. 157, 187-192

Accili, D., Arden, K.C. (2004) FoxOs at the crossroads of cellular metabolism, differentiation, and transformation. Cell. 117, 421-426

Ailion, M., Inoue, T., Weaver, C.I., Holdcraft, R.W., Thomas, J.H. (1999) Neurosecretory control of aging in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A. 96, 7394-7397.

Aitken, A., Baxter, H., Dubois, T., Clokie, S., Mackie, S., Mitchell, K., Peden, A., Zemlickova, E. (2002) Specificity of 14-3-3 isoform dimer interactions and phosphorylation. Biochem Soc Trans. 30, 351-360

Albert, P.S., Riddle, D.L. (1988) Mutants of *Caenorhabditis elegans* that form dauer-like larvae. Dev Biol. 126, 270-293

Albonico, M., Stoltzfus, R.J., Savioli, L., Tielsch, J.M., Chwaya, H.M., Ercole, E., Cancrini, G. (1998) Epidemiological evidence for a differential effect of hookworm species, *Ancylostoma duodenale* or *Necator americanus*, on iron status of children. Int J Epidemiol. 27, 530-537

Alessi, D.R. (2001) Discovery of PDK1, one of the missing links in insulin signal transduction. Colworth Medal Lecture Biochem Soc Trans. 29, 1–14

Alessi, D.R., Caudwell, F.B., Andjelkovic, M., Hemmings, B.A., Cohen, P. (1996) Molecular basis for the substrate specificity of protein kinase B; comparison with MAPKAP kinase-1 and p70 S6 kinase. FEBS Letters. 399, 333–338.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic Local Alignment Search Tool. J Mol Biol. 215, 403–410

Altschul, S.F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucl Acids Res. 25, 3389–3402

Anand, S., Hanson, K. (1997) Disability-adjusted life years: a critical review. J Health Econ. 16, 685-702

Aoki, M., Jiang, H., Vogt, P. K. (2004) Proteasomal degradation of the FoxO1

transcriptional regulator in cells transformed by the P3k and Akt oncoproteins. Proc Natl Acad Sci U S A. 101, 13613–13617

Arasu, P. (1998) Anthelmintic efficacy against tissue-arrested larvae of *Ancylostoma caninum* in murine hosts. J Parasitol. 84, 1263-1267

Barbieri, M., Bonafè, M., Franceschi, C., Paolisso, G. (2003) Insulin/IGF-I-signaling pathway: an evolutionarily conserved mechanism of longevity from yeast to humans. Am J Physiol Endocrinol Metab. 285, E1064-1071

Barsyte, D., Lovejoy, D.A., Lithgow, G.J. (2001) Longevity and heavy metal resistance in daf-2 and age-1 long-lived mutants of Caenorhabditis elegans. FASEB J. 15, 627-634

Beall, M.J., Pearce, E.J. (2002) Transforming growth factor-beta and insulin-like signalling pathways in parasitic helminths. Int J Parasitol. 32, 399-404

Bektesh, S., Van Doren, K., Hirsh, D. (1988) Presence of the *Caenorhabditis elegans* spliced leader on different mRNAs and in different genera of nematodes. Genes Devel. 2, 1277–1283

Bektesh, S.L., Hirsh, D.I. (1988) *C. elegans* mRNAs acquire a spliced leader through a *trans*-splicing mechanism. Nuc Acids Res. 16, 5692

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 340, 783–795

Berdichevsky, A., Viswanathan, M., Horvitz, H.R., Guarente, L. (2006) *C. elegans* SIR-2.1 interacts with 14-3-3 proteins to activate DAF-16 and extend life span. Cell. 125, 1165-1177

Bethony, J., Brooker, S., Albonico, M., Geiger, S.M., Loukas, A., Diemert, D., Hotez, P.J. (2006) Soiltransmitted helminth infections: ascariasis, trichuriasis, and hookworm. Lancet. 367, 1521–1532

Bethony, J., Chen, J., Lin, S., Xiao, S., Zhan, B., Li, S., Xue, H., Xing, F., Humphries, D., Yan, W., Chen, G., Foster, V., Hawdon, J.M., Hotez, P.J. (2002) Emerging patterns of hookworm infection: influence of aging on the intensity of Necator infection in Hainan Province, People's Republic of China. Clin Infect Dis. 35, 1336-1344

Bhattacharya, S., Khurana, S., Bhatti, H.S., Singhi, S., Malla, N. (2010) Polyparasitism: Fasciolopsis buski, Ascaris lumbricoides and discussion Hookworm coinfection in a child. Trop Gastroenterol. 31,126-127

Bird, D.M., Riddle, D.L. (1994) A genetic nomenclature for parasitic nematodes. J

Nematol. 26, 138–143

Blackwell, V., Vega-Lopez, F. (2001) Cutaneous larva migrans: clinical features and management of 44 cases presenting in the returning traveler. Br J Dermatol. 145, 434-437

Blumenthal, T., Steward, K. (1997) RNA processing and gene structure. In: Riddle, D.L., Blumenthal, T., Meyer, B.J., Priess, J.R., editors. C. elegans II, Cold Spring Harbor Monograph Series. Vol. 33. Cold Spring Harbor Laboratory Press; Plainview, New York, USA pp. 117–145

Boguski, M.S., Tolstoshev, C.M., Bassett, D.E.J. (1994) Gene discovery in dbEST. Science. 265,1993-1994

Bowman, D.D., Montgomery, S.P., Zajac, A.M., Eberhard, M.L., Kazacos, K.R. (2010) Hookworms of dogs and cats as agents of cutaneous larva migrans. Trends Parasitol. 26, 162-167

Braeckman, B.P., Houthoofd, K., Vanfleteren, J.R. (2001) Insulin-like signaling, metabolism, stress resistance and aging in Caenorhabditis elegans. Mech Ageing Dev. 122, 673-693

Brand, A., Hawdon, J.M. (2004) Phosphoinositide-3-OH-kinase inhibitor LY294002 prevents activation of Ancylostoma caninum and Ancylostoma ceylanicum third-stage infective larvae. Int J Parasitol. 34, 909-914

Brooker, S., Bundy, D.A.P. (2008) Soil-transmitted Helminths (Geohelminths). In: Cook, G.C., Zumla, A.I., eds. Manson's Tropical Diseases. (22nd ed) Saunders Ltd, pp. 1515-1548

Brooker, S., Clements, A.C., Hotez, P.J., Hay, S.I., Tatem, A.J., Bundy, D.A., Snow, R.W. (2006) The co-distribution of Plasmodium falciparum and hookworm among African schoolchildren. Malar J. 5, 99

Brownawell, A.M., Kops, G.J., Macara, I.G., Burgering, B.M. (2001) Inhibition of nuclear import by protein kinase B (Akt) regulates the subcellular distribution and activity of the forkhead transcription factor AFX. Mol Cell Biol. 21, 3534-3546

Brunet, A., Bonni, A., Zigmond, M.J., Lin, M.Z., Juo, P., Hu, L.S., Anderson, M.J., Arden, K.C., Blenis, J., Greenberg, M.E. (1999) Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. Cell. 96, 857-868

Brunet, A., Park, J., Tran, H., Hu, L.S., Hemmings, B.A., Greenberg, M.E. (2001) Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a). Mol Cell Biol. 21, 952-965

Brunet, A., Sweeney, L.B., Sturgill, J.F., Chua, K.F., Greer, P.L., Lin, Y.,Tran, H., Ross, S.E., Mostoslavsky, R., Cohen, H.Y., Hu, L.S., Cheng, H.L., Jedrychowski, M.P., Gygi, S.P., Sinclair, D.A., Alt, F.W., Greenberg, M.E. (2004) Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. Science. 303, 2011–2015

Bundy, D.A.P., Chan, M.S., Savioli, L. (1995) Hookworm infection in pregnancy. Trans Royal Soc Trop Med Hyg. 89, 521-522

Cahill, C.M., Tzivion, G., Nasrin, N., Ogg, S., Dore, J., Ruvkun, G., Alexander-Bridges, M. (2001) Phosphatidylinositol 3-kinase signaling inhibits DAF-16 DNA binding and function via 14-3-3-dependent and 14-3-3-independent pathways. J Biol Chem. 276, 13402-13410

Cantacessi, C., Mitreva, M., Jex, A.R., Young, N.D., Campbell, B.E., Hall, R.S., Doyle, M.A., Ralph, S.A., Rabelo, E.M., Ranganathan, S., Sternberg, P.W., Loukas, A., Gasser, R.B. (2001) Massively parallel sequencing and analysis of the Necator americanus transcriptome. PLoS Negl Trop Dis. 4, e684

Carlsson, P., and Mahlapuu, M. (2002) Forkhead transcription factors:key players in development and metabolism. Dev Biol. 250,1–23

Cassada, R.C., Russell, R.L.(1975) The dauer larva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans. Dev Biol. 46, 326-342

Clark, K.L., Halay, E. D., Lai, E., Burley, S. K. (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature. 346, 412-420

Christian, M., Zhang, X., Schneider-Merck, T., Unterman, T.G., Gellersen, B., White, J.O., Brosens, J.J. (2002) Cyclic AMP-induced forkhead transcription factor, FKHR, cooperates with CCAAT/enhancer-binding protein beta in differentiating human endometrial stromal cells. J Biol Chem. 277, 20825-20832

Clark, K.L., Halay, E.D., Lai, E., Burley, S.K. (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature. 364, 412-420

Coulson, A., Huynh, C., Kozono, Y., Shownkeen, R. (1995) The physical map of the Caenorhabditis elegans genome. Methods Cell Biol. 48, 533–550

Cox, F.E.G. (2002) History of human parasitology. Clin Microbiol Rev.15, 595–612

Craft, J. (1992) Antibodies to snRNPs in systemic lupus erythemato-sus. Rheum Dis Clin North Am. 18, 311–335

Datu, B. J., Gasser, R. B., Nagaraj, S. H., Ong, E. K., O'Donoghue, P., McInnes, R.,

Ranganathan, S., Loukas, A. (2008) Transcriptional Changes in the Hookworm, Ancylostoma caninum, during the Transition from a Free-Living to a Parasitic Larva. PLoS Negl Trop Dis. 2, e130.

Daub, J., Loukas, A., Pritchard, D.I., Blaxter, M. (2000) A survey of genes expressed in adults of the human hookworm, Necator americanus. Parasitology. 120, 171-184

Dorman, J.B. (1995) The age-1 and daf-2 genes function in a common pathway to control the lifespan of *Caenorhabditis elegans*. Genetics. 141, 1399-1406

Draghici, S., Khatri, P., Martins, R. P., Ostermeier, G. C., Krawetz, S. A. (2003) Global functional profiling of gene expression. Genomics. 81, 98-104

Dreyfuss, M.L., Stoltzfus, R.J., Shrestha, J.B., Pradhan, E.K., LeClerq, S.C., Khatry, S.K., Shrestha, S.R., Katz, J., Albonico, M., West, K.P., Jr. (2000) Hookworms, malaria and vitamin A deficiency contribute to anemia and iron deficiency among pregnant women in the plains of Nepal. J Nutr. 130, 2527-2536

Durham, S.K., Suwanichkul, A., Scheimann, A.O., Yee, D., Jackson, J.G., Barr, F.G., Powell, D.R. (1999) FKHR binds the insulin response element in the insulin-like growth factor binding protein-1 promoter. Endocrinology. 140, 3140-3146

Epe, C. (2009) Intestinal nematodes: biology and control. Vet Clin North Am Small Anim Pract. 39, 1091-1107

Estevez, A.O., Cowie, R.H., Gardner, K.L., Estevez, M. (2006) Both insulin and calcium channel signaling are required for developmental regulation of serotonin synthesis in the chemosensory ADF neurons of *Caenorhabditis elegans*. Dev Biol. 298, 32–44

Fleming, F.M., Brooker, S., Geiger, S.M., Caldas, I.R., Correa-Oliveira, R., Hotez, P.J., Bethony, J.M. (2006) Synergistic associations between hookworm and other helminth species in a rural community in Brazil. Trop Med Int Health. 11, 56–64

Flynn, R.J., Baker, D.G. (2007) Flynn's parasites of laboratory animals (2nd edition), Wiley-Blackwell, pp. 623

Fry, M., Jenkins, D.C. (1984) Nematoda: aerobic respiratory pathways of adult parasitic species.Exp Parasitol. 57, 86-92

Furuyama, T., Nakazawa, T., Nakano, I., Mori, N. (2000) Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. Biochem J. 349, 629-634

Garside, P., Behnke, J.M. (1989) *Ancylostoma ceylanicum* in the hamster: observations

on the host-parasite relationship during primary infection. Parasitology. 98, 283-289

Gottlieb, S., Ruvkun, G. (1994) *daf-2*, *daf-16* and *daf-23*: genetically interacting genes controlling Dauer formation in *Caenorhabditis elegans*. Genetics. 137, 107-120

Goyal, N., Gupta, S., Katiyar, J.C., Srivastava, V.M. (1991) NADH oxidase and fumarate reductase of *Ancylostoma ceylanicum*. Int J Parasitol. 21, 673-6

Greer, E. L., Brunet, A. (2005) FOXO transcription factor at the interface between longevity and tumor suppression. Oncogene. 24, 7410-7425

Guo, S., Rena, G., Cichy, S., He, X., Cohen, P., Unterman, T. (1999) Phosphorylation of serine 256 by protein kinase B disrupts transactivation by FKHR and mediates in on insulin-like growth factor-binding protein-1 promoter activity through a conserved insulin response sequence. J Biol Chem. 274, 17184-17192

Gupta, S., Srivastava, J. K., Katiyar, J. C. (1987) *Ancylostoma ceylanicum*: migratory behaviour in golden hamsters after oral and parenteral infection. Ann Trop Med Parasitol. 81, 421-428

Hacker, U., Grossniklaus, U., Gehring, W., Jackle, H. (1992) Developmentally regulated *Drosophila* gene family encoding the fork head domain. Proc Natl Acad Sci U S A. 89, 8754–8758

Hajarnavis, A., Korf, I., Durbin, R. (2004) A probabilistic model of 3′ end formation in *Caenorhabditis elegans*. Nucleic Acids Res. 32, 3392–3399

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 41, 95–98

Hawdon, J.M., Datu, B. (2003) The second messenger cyclic GMP mediates activation in *Ancylostoma caninum* infective larvae. Int J Parasitol. 33, 787-793

Hawdon, J.M., Datu, B., Crowell, M. (2003) Molecular cloning of a novel multidomain Kunitz-type proteinase inhibitor from the hookworm *Ancylostoma caninum*. J Parasitol. 89, 402–407

Hawdon, J.M., Narasimhan, S., Hotez, P.J. (1999) *Ancylostoma* secreted protein 2 : cloning and characterization of a second member of a family of nematode secreted proteins from *Ancylostoma caninum*. Mol Biochem Parasitol. 99, 149–165

Hawdon, J.M., Jones, B.F., Hoffman, D.R., Hotez, P.J. (1996) Cloning and

characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. J Biol Chem. 271, 6672–6678

Hawdon, J.M., Hotez, P.J. (1996) Hookworm: developmental biology of the infectious process. Curr Opin Gen Dev. 6, 618-623

Hawdon, J.M., Jones, B.F., Hotez, P.J. (1995a) Cloning and characterization of a cDNA encoding the catalytic subunit of a cAMP-dependent protein kinase from *Ancylostoma caninum* third-stage infective larvae. Mol Biochem Parasitol. 69,127–130

Hawdon, J.M., Jones, B.F., Perregaux, M.A., Hotez, P.J. (1995b) *Ancylostoma caninum*: metalloprotease release coincides with activation of infective larvae *in vitro*. Exp Parasitol. 80, 205–211

Hawdon, J.M., Schad, G.A. (1991a) Developmental Adaptations in Nematodes. In: Toft, C.A., Aeschlimann, A.,Bolis, L. (Eds.), Parasitism: Coexistence or Conflict? Oxford University Press, Oxford, pp. 274-298

Hawdon, J.M., Schad, G.A. (1991b) Long-term storage of hookworm infective larvae in buffered saline solution maintains larval responsiveness to host signals. J Helm Soc Wash. 58, 140–142

Hawdon, J.M., Schad, G.A. (1990) Serum stimulated feeding in vitro by third stage infective larvae of the canine hookworm Ancylostoma caninum. J Parasitol. 76, 394-398

Henderson, S.T., Johnson, T.E. (2001) daf-16 integrates developmental and environmental inputs to mediate aging in the nematode *Caenorhabditis elegans*. Curr Biol. 11, 1975–1980

Herr, W. (1998) The herpes simplex virus VP16-induced complex: mechanisms of combinatorial transcriptional regulation. Cold Spring Harbor Symp Quant Biol 63, 599–607

Hertweck, M., Göbel, C., Baumeister, R.C. (2004) *C. elegans* SGK-1 is the critical component in the Akt/PKB kinase complex to control stress response and life s pan. Dev Cell. 6, 577-588

Honda, Y., Honda, S. (2002) Oxidative stress and life span determination in the nematode *Caenorhabditis elegans*. Ann N Y Acad Sci. 959, 466-474

Honda Y, Honda S. (1999) The *daf-2* gene network for longevity regulates oxidative stress resistance and Mn-superoxide dismutase gene expression in *Caenorhabditis elegans*. FASEB J. 13, 1385-1393

Hong, H.K., Noveroske, J.K., Headon, D.J., Liu, T., Sy, M.S., Justice,M.J., Chakravarti, A. (2001) The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice. Genesis. 29, 163– 171

Homey, B., Dieu-Nosjean, M. C., Wiesenborn, A., Massacrier, C., Pin, J. J., Oldham, E., Catron, D., Buchanan, M. E., Muller, A., Malefyt, R. D., Deng, G., Orozco, R., Ruzicka, T., Lehmann, P., Lebecque, S., Caux, C., Zlotnik, A. (2000) Up-regulation of macrophage inflammatory protein-3 alpha/CCL20 and CC chemokine receptor 6 in psoriasis. J Immunol. 164, 6621-6632

Horton, J. (2002) Albendazole: a broad spectrum anthelminthic for treatment of individuals and populations. Curr Opin Infect Dis. 15, 599-608

Hsu, A.L., Murphy, C.T., Kenyon, C. (2003) Regulation of aging and age-related disease by DAF-16 and heat-shock factor. Science. 300, 1142-1145

Hotez, P.J., Brindley, P.J., Bethony, J. M., King, C.H., Pearce, E.J., Jacobson, J. (2008) Helminth infections: the great neglected tropical diseases. J Clin Invest. 118, 1311-1321

Hotez, P.J., Bethony, J., Bottazzi, M.E., Brooker, S., Buss, P. (2005) Hookworm: "the great infection of mankind". PLoS Med, e67

Hotez, P.J., Brooker, S., Bethony, J.M., Bottazzi, M.E., Loukas, A., Xiao, S. (2004) Hookworm infection. N Engl J Med. 351, 799-807

Hotez, P.J., Hawdon, J.M., Schad, G.A. (1993) Hookworm larval infectivity, arrest, and amphiparatenesis: the *Caenorhabditis elegans daf-c* paradigm. Parasitol Today. 9, 23-26

Hotez, P.J., Pritchard, D.I. (1995) Hookworm infection. Sci Am. 272, 68-74

Hu, M., Chilton, N.B., Gasser, R.B. (2002) The mitochondrial genomes of the human hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea). Int J Parasitol. 32, 145-158

Hu, M., Chilton, N.B., Abs El-Osta, Y.G., Gasser, R.B. (2003) Comparative analysis of mitochondrial genome data for *Necator americanus* from two endemic regions reveals substantial genetic variation. Int J Parasitol. 33, 955-963

Hu, M.C., Lee, D.F., Xia, W., Golfman, L.S., Ou-Yang, F., Yang, J.Y., Zou,Y., Bao, S., Hanada, N., Saso, H., Kobayashi, R., Hung, M.C. (2004) IkappaB kinase promotes tumorigenesis through inhibition of forkhead FOXO3a. Cell. 117, 225–237

Hu, P.J. (2007) Dauer, WormBook, ed. The C. elegans Research Community, WormBook

Hu, P.J., Xu, J., Ruvkun, G. (2006) Two membrane-associated tyrosine phosphatase homologs potentiate *C. elegans* AKT-1/PKB signaling. PLoS Genet. 2, e99

Huang, H., Regan, K.M., Wang, F., Wang, D., Smith, D.I., van Deursen, J.M., Tindall, D.J. (2005) Skp2 inhibits FOXO1 in tumor suppression through ubiquitin-mediated degradation. Proc Natl Acad Sci U S A. 102, 1649–1654

Jakobsen, M.R., Damgaard, C.K., Andersen, E.S., Podhajska, A., Kjems, J. (2004) A genomic selection strategy to identify accessible and dimerization blocking targets in the 5'-UTR of HIV-1 RNA. Nucleic Acids Res. 32, e67

James, J.A., Mamula,M.J. and Harley, J.B. (1994) Sequential autoan-tigenic determinants of the small nuclear ribonucleoprotein Sm D shared by human lupus autoantibodies and MRL lpr /lpr antibodies. Clin Exp Immunol. 98, 419–426

Jex, A.R., Waeschenbach, A., Hu, M., van Wyk, J.A., Beveridge, I., Littlewood, D.T, Gasser, R.B. (2009) The mitochondrial genomes of *Ancylostoma caninum* and *Bunostomum phlebotomum*-two hookworms of animal health and zoonotic importance. BMC Genomics.10, 79

Jia, K., Chen, D., Riddle, D.L. (2004) The TOR pathway interacts with the insulin signaling pathway to regulate *C. elegans* larval development, metabolism and life span. Development. 131, 3897–3906

Kaestner, K.H., Knöchel, W., Martinez, D.E. (2000) Unified nomenclature for the winged helix/forkhead transcription factors. Genes Dev. 14,142-146

Kalkofen, U.P. (1987) Hookworms of dogs and cats. Vet Clin North Am Small Anim Pract.17, 1341-1354

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Ziperlen, P., Ahringer, J. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature. 421, 231-237

Kapitonov, V.V., Jurka, J. (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. 3, e181

Kaufmann, J. (1996) Parasitic infections of domestic animals: a diagnostic manual, Birkhäuser Basel, pp162-163

Kawasaki, M., Hisamoto, N., Iino, Y., Yamamoto, M., Ninomiya-Tsuji, J., Matsumoto,

K. (1999) A *Caenorhabditis elegans* JNK signal transduction pathway regulates coordinated movement via type-D GABAergic motor neurons. EMBO J.18, 3604-3615

Kipreos, E.T. (2005) Ubiquitin-mediated pathways in *C. elegans*, WormBook, ed. The *C. elegans* Research Community, WormBook

Klass, M., Hirsh, D. (1976) Non-ageing developmental variant of *Caenorhabditis elegans*. Nature. 260, 523-525

Kimura, K.D., Tissenbaum, H.A., Liu, Y., Ruvkun, G. (1997) *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. Science. 277, 942-946

Kitamura, Y.I., Kitamura, T., Kruse, J.P., Raum, J.C., Stein, R., Gu, W., Accili, D. (2005) FoxO1 protects against pancreatic beta cell failure through NeuroD and MafA induction. Cell Metab.2, 153-163

Kumar, S., Tamura, K., Nei, M. (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5, 150–163

Kojima, T., Yamane, T., Nakano, H. (2006) *In vitro* selection of DNA binding sites for transcription factor, PhaR, from *Paracoccus denitrificans* using genetic library on microbeads and flow cytometry. J Biosci Bioeng.101, 440-444

Kops, G.J., de Ruiter, N.D., De Vries-Smits, A.M., Powell, D.R., Bos, J.L., Burgering, B.M. (1999) Direct control of the Forkhead transcription factor AFX by protein kinase B. Nature. 398, 630-634

Kubo, A., Nishitani, V., Minamino, N., Kikumoto, K., Kurioka, H., Nishino, T., Iwano, M., Shiiki, H., Kangawa, K., Dohi, K. (2000) Adrenomedullin gene transcription is decreased in peripheral blood mononuclear cells of patients with IgA nephropathy. Nephron. 85, 201-206

Kwon, E.S., Narasimhan, S.D., Yen, K., Tissenbaum, H.A. (2010) A new DAF-16 isoform regulates longevity. Nature. 466, 498-502

la Cour, T., Kiemer, L., Molgaard, A., Gupta, R., Skriver, K., Brunak, S. (2004) Analysis and prediction of leucine-rich nuclear export signals. Protein Eng Des Sel. 17, 527–536

Laha, T., Loukas, A., Wattanasatitarpa, S., Somprakhon, J., Kewgrai, N., Sithithaworn, P., Kaewkes, S., Mitreva, M., Brindley, P.J. (2007) The bandit, a new DNA transposon from a hookworm-possible horizontal genetic transfer between host and parasite. PLoS Negl Trop Dis. 1, e35

Laha, T., Kewgrai, N., Loukas, A., Brindley, P.J. (2006) The dingo non-long terminal repeat retrotransposons from the genome of the hookworm, *Ancylostoma caninum*. Exp Parasitol. 113, 142-153

Landmann, J.K. and Prociv, P. (2003) Experimental human infection with the dog Hookworm, *Ancylostoma caninum*. Med J Austral. 178, 69–71

Lee, S.S., Kennedy, S., Tolonen, A.C., Ruvkun, G. (2003) DAF-16 target genes that control *C. elegans* life-span and metabolism. Science. 300, 644-647

Lee, R.Y., Hench, J., Ruvkun, G. (2001) Regulation of *C. elegans* DAF-16 and its human ortholog FKHRL1 by the *daf-2* insulin-like signaling pathway. Curr Biol. 11, 1950-1957

Lee, T.I., Young, R.A. (2000) Transcription of eukaryotic protein-coding genes Annu Rev Genet. 34, 77-137

Lehtinen, M.K., Yuan, Z., Boag, P.R., Yang, Y., Villén, J., Becker, E.B., DiBacco, S., de la Iglesia, N., Gygi, S., Blackwell, T.K., Bonni, A. (2006) A conserved MST-FOXO signaling pathway mediates oxidative-stress responses and extends life span. Cell. 125, 987-1001

Lerner, M.R., Steitz, J.A. (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. Proc Natl Acad Sci USA. 76, 5495–5499

Li, W., Gao, B., Lee, S.M., Bennett, K., Fang, D. (2007a) RLE-1, an E3 ubiquitin ligase, regulates *C. elegans* aging by catalyzing DAF-16 polyubiquitination. Dev Cell. 12, 235-246

Li, J., Tewari, M., Vidal, M., Lee, S.S. (2007b) The 14-3-3 protein FTT-2 regulates DAF-16 in *Caenorhabditis elegans*. Dev Biol. 301, 82-91

Li, J., Ebata, A., Dong, Y., Rizki, G., Iwata, T., Lee, S.S. (2008) *Caenorhabditis elegans* HCF-1 functions in longevity maintenance as a DAF-16 regulator. PLoS Biol. 6, e233

Lin, K., Hsin, H., Libina, N., Kenyon, C. (2001) Regulation of the *Caenorhabditis elegans* longevity protein DAF-16 by insulin/IGF-1 and germline signaling. Nat Genet. 28, 139-145

Lin, K., Dorman, J.B., Rodan, A., Kenyon, C. (1997) *daf-16*: An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. Science. 278, 1319-1322.

Lipner, E.M., Gopi, P.G., Subramani, R., Kolappan, C., Sadacharam, K., Kumaran, P., Prevots, D.R., Narayanan, P.R., Nutman, T.B., Kumaraswami, V. (2006) Coincident filarial, intestinal helminth, and mycobacterial infection: helminths fail to influence tuberculin reactivity, but BCG influences hookworm prevalence. Am J Trop Med Hyg. 74, 841-847

Liu, T., Zimmerman, K.K., Patterson, G.I. (2004) Regulation of signaling genes by TGFbeta during entry into dauer diapause in *C. elegans*. BMC Dev Biol. 4, 11

Liu, D., Bienkowska, J., Petosa, C., Collier, R.J., Fu, H., Liddington, R. (1995) Crystal structure of the zeta isoform of the 14-3-3 protein. Nature. 376, 191-194

Liu, Y., Hengartner, M.O., Herr, W. (1999) Selected elements of herpes simplex virus accessory factor HCF are highly conserved in *Caenorhabditis elegans*. Mol Cell Biol. 19, 909-915

Livak, K. J., Schmittgen T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2–ΔΔCt method. Methods. 25, 402–408

Lozoff, B., Jimenez, E., Wolf, A.W. (1991) Long-term developmental outcome of infants with iron deficiency. New Engl J Med. 235, 687-694

Maehama, T., Dixon, J.E. (1998) The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. J Biol Chem. 273, 13375–13378

Massey, H.C. Jr., Bhopale, M.K., Li, X., Castelletto, M., Lok, J.B. (2005) The fork head transcription factor FKTF-1b from *Strongyloides stercoralis* restores DAF-16 developmental function to mutant *Caenorhabditis elegans*. Int J Parasitol. 36, 347-352

Massey, H.C. Jr., Nishi, M., Chaudhary, K., Pakpour, N., Lok, J.B. (2003) Structure and developmental expression of *Strongyloides stercoralis fktf-1*, a proposed ortholog of *daf-16* in *Caenorhabditis elegans*. Int J Parasitol. 33, 1537-1544

Maxwell, C., Hussain, R., Nutman, T.B., Poindexter, R.W., Little, M.D., Schad, G.A., Ottesen, E.A. (1987) The clinical and immunologic responses of normal human volunteers to low dose hookworm (*Necator americanus*) infection. Am J Trop Med Hyg. 37, 126-134

McCarter, J.P., Bird, D.M., Mitrev, M. (2005) Nematode gene sequences: update for december 2005. J Nematol. 37, 417-421

McElwee, J., Bubb, K., Thomas, J.H. (2003) Transcriptional outputs of the *Caenorhabditis elegans* forkhead protein DAF-16. Aging Cell. 2, 111-121. Erratum in: Aging Cell. (2003) 2, 341

McElwee, J.J., Schuster, E., Blanc, E., Thomas, J.H., Gems, D. Shared transcriptional signature in *Caenorhabditis elegans* Dauer larvae and long-lived *daf-2* mutants implicates detoxification system in longevity assurance. J Biol Chem. 279, 44533-44543

Mitreva, M., Wendl, M.C., Martin, J., Wylie, T., Yin, Y, Larson, A., Parkinson, J, Robert H Waterston, R.H., and James P McCarter, J.P. (2006) Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. Genome Biol. 7, R75

Mitreva, M., McCarter, J.P., Arasu, P., Hawdon, J., Martin, J., Dante, M., Wylie, T., Xu, J., Stajich, J.E., Kapulkin, W., Clifton, S.W., Waterston, R.H., Wilson, R.K. (2005) Investigating hookworm genomes by comparative analysis of two *Ancylostoma* species. BMC Genomics. 6, 58

Morris, J.Z., Tissenbaum, H.A., Ruvkun, G. (1996) A phosphatidylinositol-3-OH kinase family member regulating longevity and diapause in *Caenorhabditis elegans*. Nature. 382, 536-539

Moser, J.M., Freitas, T., Arasu, P., Gibson, G. (2005) Gene expression profiles associated with the transition to parasitism in *Ancylostoma caninum* larvae. Mol Biochem Parasitol. 143, 39-48

Murphy, C. T. (2006) The search for DAF-16/FOXO transcriptional targets: approaches and discoveries. Exp Gerontol. 41, 910-921

Murphy, C.T., McCarroll, S.A., Bargmann, C.I., Fraser, A., Kamath, R.S., Ahringer, J., Li, H., Kenyon, C. (2003) Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. Nature. 424, 277-283

Murray, C.J. (1994) Quantifying the burden of disease: the technical basis for disability-adjusted life years. Bull World Health Organ. 72, 429-445

Muslin, A.J., Tanner, J.W., Allen, P.M., Shaw, A.S. (1996) Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine. Cell. 84, 889-897

Muñoz, M.J., Riddle, D.L. (2003) Positive selection of *Caenorhabditis elegans* mutants with increased stress resistance and longevity. Genetics. 163, 171-180

Nasrin, N., Ogg, S., Cahill, C.M., Biggs, W., Nui, S., Dore, J., Calvo, D., Shi, Y., Ruvkun, G., Alexander-Bridges, M.C. (2000) DAF-16 recruits the CREB-binding protein coactivator complex to the insulin-like growth factor binding protein 1 promoter in HepG2 cells. Proc Natl Acad Sci U S A. 97, 10412-10417

Nolan, T.J., Bhopale, V., Megyeri, Z., Schad, G.A. (1994) Cryopreservation of human hookworms. J Parasitol. 80, 648–650

Obenauer, J.C., Cantley, L.C., Yaffe, M.B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res. 31, 3635–3641

O'Brien, R.M., Noisin, E.L., Suwanichkul, A., Yamasaki, T., Lucas, P.C., Wang, J.C., Powell, D.R., Granner, D.K. (1995) Hepatic nuclear factor 3- and hormone-regulated expression of the phosphoenolpyruvate carboxykinase and insulin-like growth factor-binding protein 1 genes. Mol Cell Biol. 5, 1747-1758

O'Brien, R.M., Lucas, P.C., Yamasaki, T., Noisin, E.L., Granner, D.K. (1994) Potential convergence of insulin and cAMP signal transduction systems at the phosphoenolpyruvate carboxykinase (PEPCK) gene promoter through CCAAT/enhancer binding protein (C/EBP). J Biol Chem. 269, 30419-30428

Obsil, T., Ghirlando, R., Anderson, D. E., Hickman, A. B., Dyda, F. (2003) Two 14-3-3 binding motifs are required for stable association of Forkhead transcription factor FOXO4 with 14-3-3 proteins and inhibition of DNA binding. Biochemistry. 42, 15264-15272.

Ogg, S., Paradis, S., Gottlieb, S., Patterson, G.I., Lee, L., Tissenbaum, H.A., Ruvkun, G. (1997) The fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans.* Nature. 389, 994-999

Ogg, S., Ruvkun, G. (1998) The *C. elegans* PTEN homolog, DAF-18, acts in the insulin receptor-like metabolic signaling pathway. Mol Cell. 2, 887-893

Oh, S.W., Mukhopadhyay, A., Dixit, B. L., Raha, T., Green, M.R., Tissenbaum, H.A. (2005a) Identification of direct DAF-16 targets controlling longevity, metabolism and diapause by chromatin immunoprecipitation. Nat Genet. 38, 251-257

Oh, S.W., Mukhopadhyay, A., Svrzikapa, N., Jiang, F., Davis, R.J., Tissenbaum, H.A. (2005b) JNK regulates lifespan in *Caenorhabditis elegans* by modulating nuclear translocation of forkhead transcription factor/DAF-16. Proc Natl Acad Sci U S A. 102, 4494-4499

Onwuliri, COE (1985) Energy metabolism in the developing larval stages of *Ancylostoma tubaeforme* and *Haemonchus contortus*: glycolytic and tncarboxylic acid cycle enzymes. Parasitology. 90, 169-177

Ookuma, S., Fukuda, M., Nishida, E. (2003) Identification of a DAF-16 transcriptional target gene, *scl-1*, that regulates longevity and stress resistance in *Caenorhabditis elegans.* Curr Biol. 13, 427-431

Padgett, R.A., Grabowski, P.J., Konarska,M.M., Seiler, S. and Sharp, P.A. (1986) Splicing of messenger RNA precursors. Annu Rev Biochem. 55, 1119–1150

Paradis, S., Ailion, M., Toker, A., Thomas, J.H., and Ruvkun, G. (1999) A PDK1 homolog is necessary and sufficient to transduce AGE-1 PI3 kinase signals that regulate diapause in *Caenorhabditis elegans*. Genes Dev. 13, 1438–1452

Paradis, S., Ruvkun, G. (1998) *Caenorhabditis elegans* Akt/PKB transduces insulin receptor-like signals from AGE-1 PI3 kinase to the DAF-16 transcription factor. Genes Dev. 12, 2488–2498

Peduzzi, R., Piffaretti, J.C. (1983) Ancylostoma duodenale and the Saint Gothard anaemia. Br Med J (Clin Res Ed). 287, 1942-1945

Pierrou, S., Hellqvist, M., Samuelsson, L., Enerback, S. and Carlsson, P. (1994) Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. EMBO J. 13, 5002-5012

Ponchel, F., Toomes, C., Bransfield, K., Leong, F. T., Douglas, S. H., Field, S. L., Bell, S. M., Combaret, V., Puisieux, A., Mighell, A. J., Robinson, P. A., Inglehearn, C. F., Isaacs, J. D., Markham, A. F. (2003) Real-time PCR based on SYBR-Green I fluorescence: an alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. BMC Biotechnol. 3, 18

Puigserver, P., Rhee, J., Donovan, J., Walkey, C.J., Yoon, J.C., Oriente, F., Kitamura, Y., Altomonte, J., Dong, H., Accili, D., Spiegelman, B.M. (2003) Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1alpha interaction. Nature. 423, 550-555

Ranjit, N., Jones, M.K., Stenzel, D.J., Gasser, R.B., Loukas, A. (2006) A survey of the intestinal transcriptomes of the hookworms, *Necator americanus* and *Ancylostoma caninum*, using tissues isolated by laser microdissection microscopy. Int J Parasitol. 36, 701-710

Ray, D. K., Bhopale, K. K. (1972) Complete development of *ancylostoma ceylanicum* (Looss, 1911) in golden hamsters, mesocricetus auratus. Experientia. 28, 359-361
.

Rena, G., Woods, Y.L., Prescott, A.R., Peggie, M., Unterman, T.G., Williams, M.R., Cohen, P. (2002) Two novel phosphorylation sites on FKHR that are critical for its nuclear exclusion. EMBO J. 21, 2263-2271

Rhee, S.Y., Wood, V., Dolinski, K., Draghici, S. (2008) Use and misuse of the gene ontology annotations. Nat Rev Genet. 9, 509-515

Riddle, D.L., Albert, P.S. (1997) Genetic and environmental regulation of dauer larvae

development. In: *C. elegans* II, Cold Spring Harbor Laboratory Press, pp. 738–768.

Riddle, D.L., Swanson, M.M., Albert, P.S. (1981) Interacting genes in nematode dauer larva formation. Nature. 290, 668-671

Roberts, L.S., Jonovy, J., Jr. (2005) Nematodes: Strongylida, Bursate Rhabditians. In: Foundations of Parasitology (7th edition), McGraw-Hill, New York, pp. 417-418

Rochette-Egly, C. (2003) Nuclear receptors: integration of multiple signalling pathways through phosphorylation. Cell Signal. 15, 355-366

Rogers, W.P., Petronijevic, T. (1982) The infective stage and the development of nematodes. In: Symons, L.E.A., Donald, A.D.,Dineen, J.K. (Eds.), Biology and Control of Endoparasites, Academic Press, Australia, pp. 3-28

Rogers, W.P., Sommerville, R.I. (1963) The infective stage of nematode parasites and its significance in parasitism. Adv Parasit. 1, 109-177

Rokeach, L.A., Haselby, J.A. and Hoch, S.O. (1988) Molecular cloning of a cDNA encoding the human Sm-D autoantigen. Proc Natl Acad Sci U S A. 85, 4832–4836

Schad, G.A., Page, M.R. (1982) *Ancylostoma caninum*: adult worm removal, corticosteroid treatment, and resumed development of arrested larvae in dogs. Exp Parasitol. 54, 303-309

Schubert, L.A., Jeffery, E., Zhang, Y., Ramsdell, F., Ziegler, S.F. (2001) Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. J Biol Chem. 276, 37672– 37679

Sherman, B.T., Huang, D.W., Tan, Q., Guo, Y., Bour, S., Liu, D., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A. (2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics. 8, 426

Shostak, Y., Van Gilst, M.R., Antebi, A., Yamamoto, K.R. (2004) Identification of *C. elegans* DAF-12-binding sites, response elements, and target genes. Genes Dev. 18, 2529–2544

Shu, W., Yang, H., Zhang, L., Lu, M.M., Morrisey, E.E. (2001) Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem. 276, 27488– 27497

Stoltzfus, R.J., Dreyfuss, M.L., Chwaya, H.M., Albonico, M. (1997) Hookworm control as a strategy to prevent iron deficiency. Nutr Rev. 55, 223-232

Sulston, J., Hodgkin, J. (1988) Methods. In: Wood WB, editor. The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory; Cold Spring Harbor, NY. pp. 587–606

The C. elegans Sequencing Consortium. (1998) Genome sequence of the nematode *C. elegans*: A platform for investigating biology. Science. 282, 2012–2018

Thomas, H.S. (2009) The Cattle Health Handbook: Preventive Care, Disease Treatments & Emergency Procedures for Promoting the Well-Being of Your Beef Or Dairy Herd, Storey Books, pp133-134

Tissenbaum, H.A., Guarente, L. (2001) Increased dosage of a *sir-2* gene extends lifespan in *Caenorhabditis elegans*. Nature. 410, 227-230

Tissenbaum, H.A., Hawdon, J., Perregaux, M., Hotez, P., Guarente, L., Ruvkun, G. (2000) A common muscarinic pathway for diapause recovery in the distantly related nematode species *Caenorhabditis elegans* and *Ancylostoma caninum*. Proc Natl Acad Sci U S A. 97, 460-465.

Tissenbaum, H.A., Ruvkun, G. (1998) An insulin-like signaling pathway affects both longevity and reproduction in *Caenorhabditis elegans*. Genetics. 148, 703-717

Trivedi, S., Arasu, P. (2005) Evaluation of endogenous reference genes for real-time PCR quantification of gene expression in *Ancylostoma caninum*. Mol Biochem Parasitol. 143, 241-244

Tsai, W.C., Bhattacharyya, N., Han, L.Y., Hanover, J.A., Rechler, M.M. (2003) Insulin inhibition of transcription stimulated by the forkhead protein Foxo1 is not solely due to nuclear exclusion. Endocrinology. 144, 5615-5622

Tuerk, C., Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacterophage T4 DNA polymerase. Science. 249, 505-510

Tuteja, G., Kaestner, K.H. (2007a). Forkhead transcription factors II. Cell. 131, 192

Tuteja, G., Kaestner, K.H. (2007b). SnapShot: forkhead transcription factors I. Cell. 130, 1160

Vowels, J.J., Thomas, J.H. (1992) Genetic analysis of chemosensory control of dauer formation in *C. elegans*. Genetics. 130, 105–123

Van Der Heide, L.P., Hoekman, M.F., Smidt, M.P. (2004) The ins and outs of FoxO shuttling: mechanisms of FoxO translocation and transcriptional regulation. Biochem J. 380, 297–309

Vlahos, C., Matter, W., Hui, K., Brown, R. (1994) A specific inhibitor of phosphatidylinositol 3-kinase, 2-(4- morpholinyl)-8-phenyl-4H-1-benzopyran-4-one (LY294002). J Biol Chem. 269, 5241–5248

Walker, G.A., Thompson, F.J., Brawley, A., Scanlon, T., Devaney, E. (2003) Heat shock factor functions at the convergence of the stress response and developmental pathways in *Caenorhabditis elegans*. FASEB J. 17, 1960-1962

Walker, G.A., White, T.M., McColl, G., Jenkins, N.L., Babich, S., Candido, E.P., Johnson, T.E., Lithgow, G.J. (2001) Heat shock protein accumulation is up-regulated in a long-lived mutant of *Caenorhabditis elegans.* J Gerontol A Biol Sci Med Sci. 56, B281-287.

Walson, J.L., Stewart, B.T., Sangaré, L., Mbogo, L.W., Otieno, P.A., Piper, B.K., Richardson, B.A., John-Stewart. G. (2010) Prevalence and correlates of helminth co-infection in Kenyan HIV-1 infected adults. PLoS Negl Trop Dis. 4, e644

Wang, M.C., Bohmann, D., Jasper, H. (2005) JNK extends life span and limits growth by antagonizing cellular and organism-wide responses to insulin signaling. Cell. 121, 115-125

Wang, W., Shakes, D.C. (1994) Isolation and sequence analysis of a *Caenorhabditis elegans* cDNA which encodes a 14-3-3 homologue. Gene.147, 215-218

Wang, W., Shakes, D.C. (1996) Molecular evolution of the 14-3-3 protein family. J Mol Evol. 43, 384-398

Wang, W., Shakes, D.C. (1997) Expression patterns and transcript processing of *ftt-1* and *ftt-2*, two *C. elegans* 14-3-3 homologues. J Mol Biol. 268, 619-630

Wang, Y., Oh, S.W., Deplancke, B., Luo, J., Walhout, A.J., Tissenbaum, H.A. (2006) *C. elegans* 14-3-3 proteins regulate life span and interact with SIR-2.1 and DAF-16/FOXO. Mech Ageing Dev. 127, 741-747

Wang, Z., Abubucker, S., Martin, J., Wilson, R.K., Hawdon, J., Mitreva, M. (2010) Characterizing *Ancylostoma caninum* transcriptome and exploring nematode parasitic adaptation. BMC Genomics. 11, 307

Webster, J.M., Eriksson, K. B., McNamar, D. G. (2008) An Anecdotal History of Nematology, Pensoft Publishers, pp.92

Weigel, D., Jurgens, G., Kuttner, F., Seifert, E., Jackle, H. (1989) The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. Cell. 57, 645-658

WHO (2004). The global burden of disease. (Geneva, World Health Organization),

pp.60

WHO (2010). Working to overcome the global impact of neglected tropical diseases (Geneva, World Health Organization), pp.136

Woods, Y.L., Rena, G., Morrice, N., Barthel, A., Becker, W., Guo, S., Unterman, T.G., Cohen, P. (2001) The kinase DYRK1A phosphorylates the transcription factor FKHR at Ser329 *in vitro*, a novel in vivo phosphorylation site. Biochem J. 355, 597-607

Yaffe, M.B. (2002) How do 14-3-3 proteins work?-- Gatekeeper phosphorylation and the molecular anvil hypothesis. FEBS Lett. 513, 53-57

Yang, Y., Hou, H., Haller, E.M., Nicosia, S.V. and Bai, W. (2005) Suppression of FOXO1 activity by FHL2 through SIRT1-mediated deacetylation. EMBO J.24, 1021–1032

Yoshida, Y., Kondo, K., Kurimoto, H., Fukutome, S., Shirasaka, S. (1974) Comparative studies on *Ancylostoma braziliense* and *Ancylostoma ceylanicum*. III. Life history in the definitive host. J Parasitol. 60, 636-641

Yu, H., Larsen, P.L. (2001) DAF-16-dependent and independent expression targets of DAF-2 insulin receptor-like pathway in Caenorhabditis elegans include FKBPs. J Mol Biol. 314, 1017-1028

Yuan, J. S., Wang, D., Stewart, C. N. Jr. (2008) Statistical methods for efficiency adjusted real-time PCR quantification. Biotechnol J. 3, 112-123

Zhang, B., Schmoyer, D., Kirov, S., Snoddy, J. (2004) GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. BMC Bioinformatics. 5, 16

Zhang, X, Gan, L., Pan. H., Guo, S., He, X., Olson, S.T., Mesecar, A., Adam, S., Unterman, T.G. (2002) Phosphorylation of serine 256 suppresses transactivation by FKHR (FOXO1) by multiple mechanisms. Direct and indirect effects on nuclear/ cytoplasmic shuttling and DNA binding. J Biol Chem. 277, 45276-45284