Molecular and immunological characterisation of *Acanthocheilonema viteae* chitinase

Dissertation
zur Erlangerung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin
von
Babila Julius Tachu (M.Sc. Biochemistry)
geboren am 14.02.74 in Limbe (Victoria), Kamerun

Präsident in Vertretung der Humboldt-Universität zu Berlin
Prof. Dr. Hans Jürgen Prömel
Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I
Prof. Thomas Buckhout, PhD

Gutachter/innen: 1. Prof. Dr. Richard Lucius
2. Prof. Dr. Wolfgang Höhne
3. Prof. Dr. Rainer Borriss

Tag der Einreichung: 29.11.2005

Tag der mündlichen Prüfung: 04.05.2006

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die vorliegende Dissertation selbständig und ohne unerlaubte Hilfe aufgefertigt zu haben.

Bei einer Abfassung der Dissertation wurden keine anderen als die im Text aufgeführten Hilfsmittel benutzt.

Berlin den 23.11.2005

Babila Julius Tachu

Table of Contents

Dedication

This work is dedicated to:

Elizabeth Abimwoe and the one born after she transisted

Feh Susanna

Jeremiah Tchinda and Tanyi-Mbombo

Acknowledgements

I wish to extend my sincere gratitude to all those who contributed to the success of this work and my sojourn in the Chair of Molecular Parasitology, Humboldt-University of Berlin.

I wish to thank Professor Dr. Richard Lucius for his supervision, moral support and for offering me the possibility to do my thesis at the Chair of Molecular Parastology.

I wish to thank Prof. Dr. Vincent P.K. Titanji and the members of his research team at the Biotechnology Unit, University of Buea for their support and cooperation.

My particular thanks go to Prof. Dr. W. Höhne and Dr. Rudolf Volkmer-Engert of Charité Medical faculty, Humboldt-University of Berlin for designing and synthesizing the synthetic peptides, respectively.

My thanks also go to Dr. Jörg Hirzmann (University of Gießen) for providing the *A. viteae* genomic library used in this study.

My sincere gratitude goes to Dr. Thomas Pogonka for the close supervision and for regular and helpful discussions on the progress of this work.

I thank the staff and students at the Chair of Molecular Parasitology Berlin, for their support and discussions. Karin Biermann, Grit Meusel and Bettina Sonenburg are particularly thanked for their help in animal experiments. Many thanks go to Smitha Pillai and Michal Sereda for their moral support and collaboration. My thanks go to Chris, Jörg, Florian, Oliver and Katja for their interest in the work.

I also wish to thank my friends and family. Cécile Barra, Aysun Kara and Sabrina Werl are thanked for reviving my down-trodden spirit.

My special gratitude goes to the German Academic Exchange Services (DAAD) for financing my stay and study, as well as Dr. Bernd Kalinna and Dr. Susanna Hartmann for employing me within the framework of their projects.

Summary

Chitinases are enzymes that break down chitin, a homopolymer of N-acetylglucosamine (GlcNAc) to its monomers. They are important molecules in the life cycle of parasitic nematodes representing important drug and vaccine targets. In this thesis, three genomic chitinase sequences (I, II and III) were obtained by characterising nine clones from a genomic library of *Acanthocheilonema viteae*. Southern blots independently confirmed the existence of three chitinase genes in *A. viteae*. The organisation of all three genomic sequences is similar, with the greatest difference occurring in exons encoding the serine-threonine rich domain of chitinases: this domain is about ten times larger in sequence III compared to I, but is absent in sequence II. Sequence I and III had features of regularly transcribed genes: start ATG, followed by an open reading frame, stop codon and polyadenylation signal. Sequence II lacked the first exon with start ATG. Screening of cDNA libraries from adult female *A. viteae* worms and L3 (third-stage larvae), respectively, as well as reverse transcriptase PCR (RT-PCR) showed transcripts in uterine microfilariae, L3 and L4 (fourth-stage larvae) for gene I only.

The N-terminal fragment of *A. viteae* chitinase was cloned into an expression vector and expressed in *Escherichia coli*. About 80% of the expressed chitinase were found in inclusion bodies and were purified under denaturing conditions. The soluble fraction was about 20% and could be purified under native conditions. Chitinase purified from inclusion bodies was 13-fold less active compared to soluble chitinase.

Synthetic peptides (P1 and P2) were designed from the active site of *A. viteae* chitinase, and used in parallel with chitinase from inclusion bodies in vaccination experiments using the *A. viteae* / *Meriones unguiculatus* filariasis model. Vaccination with recombinant protein led to a 29 % significant reduction in adult worm burden in a single experiment. Vaccination with P1 and P2 led to an overall non significant reduction in adult worm burden in two independent experiments. In the P1 group, there was a consistent reduction (39%, $p>0.05$ and 45%, $p<0.05$) in mf load that attained significance only in the second experiment. In the P2 group, there was no reduction in mf burden in the first experiment, but a significant reduction (75%, $p<0.05$) in the second.

These results suggest that filarial chitinases are potential targets for transmission blocking drugs and vaccines.

Zusammenfassung

Chitinasen sind Enzyme, die Chitin, ein Homopolymer des N-Acetylglucosamin (GlcNAc) in seine Monomere spalten. Diese sind wichtige Moleküle im Lebenszyklus der parasitischen Fadenwürmer, welche bedeutende Medikamenten- und Impfstoffziele darstellen. In der vorliegenden Arbeit wurden durch die Charakterisierung von 9 Klonen aus einer Genbank von *Acanthocheilonema viteae* drei genomische Chitinasesequenzen (I, II und III) gefunden. Diese Anzahl an Chitinasegenen wurde durch Southern-Blots bestätigt. Der Aufbau aller drei Sequenzen ist sehr ähnlich. Die größten Unterschiede sind in den Exons zu finden, welche die Serin-Threonin-reiche Domäne der Chitinasen codieren. Diese Domäne ist in Sequenz III ca. 10fach länger als in Sequenz I. In Sequenz II ist sie nicht vorhanden. Sequenz I und III hatten Eigenschaften eines regulär transkribierten Genes: eine Startcodon, gefolgt von einem offener Leserahmen, einem Stopcodon und einem Polyadenylierungssignal. Sequenz II fehlt das erste Exon mit dem Startcodon. Bei Durchmusterung einer cDNA-Bibliothek adulter *A. viteae* Würmer bzw. L3 Stadien, sowie durch Reverser Transkriptase-PCR (RT-PCR) wurden in den Mikrofilarien, L3 und L4 Transkripte für Gen I gefunden, jedoch nicht für Gen III.

Das N-terminale Fragment der *A. viteae*-Chitinase wurde in ein Expressionsplasmid ligiert und in *E. coli* exprimiert. Ungefähr 80 % der exprimierten Chitinase lagen in Einschluss-Körpern vor. Dieser Anteil konnte unter denaturierenden Bedingungen aufgereinigt werden. Der lösliche Anteil konnte unter nativen Bedigungen aufgereinigt werden. Die aus den Einschluss-Körpern aufgereinigte Chitinase zeigte im Vergleich zur löslichen Chitinase eine 13fach verminderte Aktivität.

Vom aktiven Zentrum der Chitinase wurden Peptide synthetisiert. Diese wurden parallel mit Chitinase aus den Einschluss-Körpern für Vakzinierungsexperimente im *A. viteae / Meriones unguiculatus*-Filarien-Modell genutzt. Die Reduzierung der Wurmlast um 29 % nach einer Immunisierung mit rekombinantem Protein zeigte eine tendenziell schützende Kapazität des verwendeten Proteins. In zwei unabhängigen Experimenten konnte nach Immunisierung mit zwei synthetischen Peptiden (P1 und P2) eine bedeutende Reduktion der Wurmlast beobachtet werden. In der P1-Gruppe gab es eine gleichmäßige Reduktion der Mf-Last, die nur im zweiten Experiment signifikant war (39 %, $p > 0,05$ und 45 %, $p < 0,05$). In der P2-Gruppe konnte in einem von zwei Experimenten eine signifikante Reduktion der Mf-Last erzielt werden (75 %, $p < 0,05$).

Diese Ergebnisse lassen vermuten, daß Filarien-Chitinasen potenzielle Ziele für Medikamente und Impfstoffe sind.

1 Introduction

Infection with filarial parasites affect about 200 million people worldwide (Saint André et al., 2002). *Onchocerca volvulus*, the cause of onchocerciasis, is a major health problem in 36 endemic countries of Africa, Latin America and the Arabian Peninsula; it is responsible for one million visually impaired subjects, 270,000 bilaterally blind cases and 6.5 million subjects suffering from severe itching or dermatitis. Globally, 18 million people are infected with the parasite, while some 120 million others are at risk of infection (http://www.who.int/int-fs/en/fact095.html). Lymphatic filariasis, caused by the parasites *Brugia malayi* and *Wuchereria bancrofti*, affects about 120 million people worldwide, of which 40 million are incapacitated and disfigured. It is a major health problem in India, Africa, South Asia and the Pacific (http://www.who.int/mediacentre/fact sheets/fs102/en/).

Chemotherapeutic treatment and vector control are two strategies used to fight filariasis, but none of these strategies can have a long-lasting impact in endemic areas (Hougard et al., 1997; Cook et al., 2001, Richards et al., 2001). Vaccination is a promising additional control measure, and various strategies were used to identify and validate preventive vaccine candidates for filariasis (Abraham et al., 2002).

The search for chemotherapeutically and immunologically relevant target molecules led among others to a group of enzymatically active molecules, the chitinases.

1.1 Structure and function of chitinases

Chitinases (classified as EC 3.2.1.14 by the International Union of Biochemistry and Molecular Biology) are enzymes that break down chitin, a homopolymer of N-acetylglucosamine (GlcNAc) to its monomers. They are ubiquitous in the plant and animal kingdom (Flach et al., 1992) and play important structural, physiological, metabolic and defensive roles (Cohen-Kupiec et al., 1998; Felse et al., 1999). Chitinases are glycosyl hydrolases, a family that is subdivided according to their hydrolysis mechanisms and amino acid sequence similarities of catalytic domains (Henrissat, 1999; Henrissat and Bairoch, 1993, 1996) (http://afmb.cnrs-mrs.fr/CAZY). Chitinases belong to class 18 and 19 of this grouping. Family 18 chitinases are endochitinases that cleave chitin by a retaining mechanism through which the beta-linked polymer is cleaved to release beta anomer products. Family 19 chitinases are mostly found in plants, some bacteria and nematodes and cleave chitin by an inversion or retention mechanism (Robertus and Monzingo, 1999; Fukamizo, 2000;

Honda et al., 2004). Family 19 class IA/I and IB/II enzymes differ in the presence (IA/I) or absence (IB/II) of an N-terminal chitin-binding domain. Chitinases also have a carbohydrate-binding module that promotes adsorption of the enzyme to insoluble chitin. The chitin binding domain of filarial family 18 chitinases (EC 3.2.1.14) are in the carbohydrate-binding module family 14, and contain six conserved cysteins that probably form three disulfide bridges (Henrissat and Bairoch, 1993, 1996) (http://afmb.cnrs-mrs.fr/CAZY/).

Although the overall sequence similarity (average pair wise identity) between family 18 chitinases is only 21% (http://www.sanger.ac.uk/Software/Pfam), their active site regions contain many residues that are fully or highly conserved, and their structure is a TIM barrel fold as typified by chitinase-1 of *Coccidiodes immitis* (Figure 1) (Hollis et al., 1998; 2000; Perakis et al., 1994; Terwischa van Scheltinga et al., 1994; van Aalten et al., 2001).

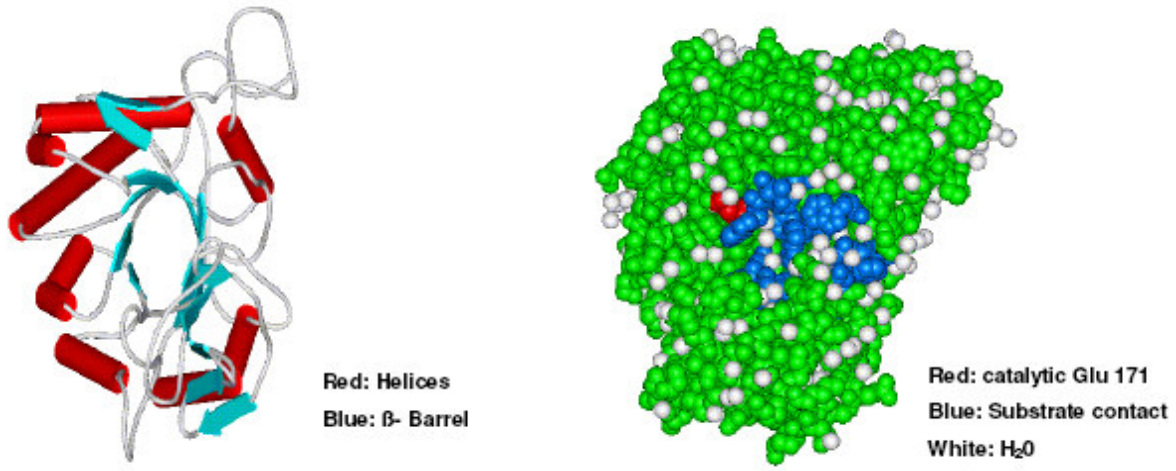


Fig. 1: Structure of chitinase-1 from *Coccidiodes immitis*.

A TIM-barrel (β /α)8) structure is shown on the left, and a model showing the catalytic Glu and positions of substrate contact is shown on the right. Diagrams were kindly provided by Prof. Dr. W. Höhne, Humboldt-University of Berlin)

In ChiB from *Serratia marcescens*, Tyr10, Ser93, Asp140, Asp142, Glu144, Tyr214, Asp215 and their corresponding subsets in other family 18 chitinases have been shown to be important for catalysis. The catalytically active residue is a glutamate in a stretch of highly conserved amino acids having the consensus DXDXE motif

(Watanabe et al., 1993; Bortone et al., 2002). The other residues important for catalysis have both mechanistic and chemical roles (Synstad et al., 2004). The carbohydrate binding clefts are present on the C-terminal side of the β strands in the (β /α)8 barrel (Fusseti et al., 2003).

Chitinases have been described for the infective larval (L3) stage of *A. viteae* (Adam et al., 1996; Wu et al., 1996) and *O. volvulus* (Wu et al., 1996), as well as in the microfilarial stage of *B. malayi*, *B. pahangi* (Fuhrman et al., 1992) and in infective larvae (L3) and microfilariae of *W. bancrofti* (Raghavan et al., 1994). L3 chitinases of *A. viteae* and *O. volvulus* are accumulated in the glandular oesophagus of the worms in their insect vectors, and released when the larvae are transmitted to the hosts and eventually diminish following moulting from L3 to L4 (Wu et al., 1996). Micorfilaria-specific chitinases (Fuhrman et al., 1992) could be recognised by a monoclonal antibody in extracts of microfilariae following several days of maturation in the vertebrate host as blood-borne microfilariae (Furhman et al., 1987). Interestingly, the appearance of these chitinases corresponds with the parasite's ability to infect the insect host (Fuhrman et al., 1992). To understand the possible role of these chitinases in the filariae, it is worth taking a brief look into the biology and biochemistry of the filariae relative to chitinase.

The filariae can be divided into two groups based on their mode of embryogenesis and eggshell remodelling, which arises from the two distinct patterns of intrauterine embryonic development (Rogers et al., 1976; Ellis et al., 1978). In filarial species like *D. immitis*, *O. volvulus* and *A. viteae*, the developing embryo hatches from its eggshell within the maternal uterus, and is subsequently shed as a microfilaria with no extracuticular structures or sheath. In *W. bancrofti*, *B. malayi*, and *B. pahangi*, the eggshell is remodelled as a sheath that eventually becomes part of the microfilaria. Another important difference is the route of migration of filarial microfilariae in their arthropod hosts. *B. malayi* and *B. pahangi* microfilariae penetrate the midgut of the mosquito vector and enter the haemocoel, while microfilariae of *D. immitis* enter the Malpighian tubules through the lumen at the posterior end of the gut.

The presence of chitin, the substrate of chitinases, has been biochemically documented in various nematode species and tissues. Chitin has been shown to be a component of the egg shells of many nematodes including the filariae (Brydon et

al., 1987). In addition, chitin has been demonstrated on the sheath of *B. malayi* microfilariae (Kaushal et al., 1984) and *W. bancrofti* microfilariae (Araujo et al., 1993).

In view of the biology and location of filarial chitinases, microfilaria chitinases of Brugian and Bancroftian filariae have been proposed to have a function in ex-sheathing the microfilaria when they are taken up in the arthropod vectors (Fuhrman et al., 1992). *A. viteae* and *O. volvulus* L3 chitinase are thought to have roles in ecdysis during post-infective development of the filariae, and in worm migration through the interstitial tissues of the host (Adam et al., 2001; Wu et al., 1996; 2001).

1.2 Filarial chitinases are potential targets for interventions

Filarial chitinases are potential targets for drug and immune interventions. Several observations implicate filarial chitinases to have roles in immune protection. In endemic areas, there are putatively immune (PIs) individuals (Elson et al., 1994; Turaga et al., 2000) with no current or past evidence of filarial infections. Interestingly, the sera of individuals putatively immune for *W. bancrofti* infection recognised a 43kD *B. malayi* L3 chitinase-like molecule, as opposed to the sera of microfilaremic individuals (Freedman et al., 1989). A monoclonal antibody that mediated microfilaria clearance upon transfer recognised two *B. malayi* microfilaria-specific chitinases (Canlas et al., 1984a; Canlas et al., 1984b; Southworth et al., 1996). In addition, Vaccination of jirds with irradiated *A. viteae* L3 led to 90% protection (Lucius et al., 1991) and the sera of vaccinated rodents dominantly recognised *A. viteae* L3 chitinase in blots, implicating L3 chitinase as an immunodominant antigen playing an important role in the immune protection of the animals (Adam et al., 1996). These observations suggest that filarial chitinases may play a role in protective immunity.

Further more, immunisation of mice with *O. volvulus* chitinase DNA led to a significant reduction of 53% of worm burden following challenge infection (Harrison et al., 1999), experimentally confirming the relevance of filarial chitinases as vaccine molecules.

An alternative approach to immunisation with recombinant proteins would be the use of synthetic peptides designed from the active sites of filarial chitinases. Immunisation with such peptides could lead to the production of neutralising antibodies that will inhibit enzyme activity in the parasites (da Costa et al., 1999; Hirota et al., 2001; Fujii et al., 2004). A major short coming is the absence of a crystal

structure for filarial chitinase. In the absence of a crystal structure, homology modelling could be used to design peptides. The highest homology between filarial chitinases and a chitinase with known structure is between *A. viteae* chitinase and chitinase-1 of *C. immitis* (Höhne, personal communication). Using homology modelling, a structure of the active site of filarial chitinases could be obtained with *C. immitis* chitinase as template (Figure 2) (Höhne, personal communication).

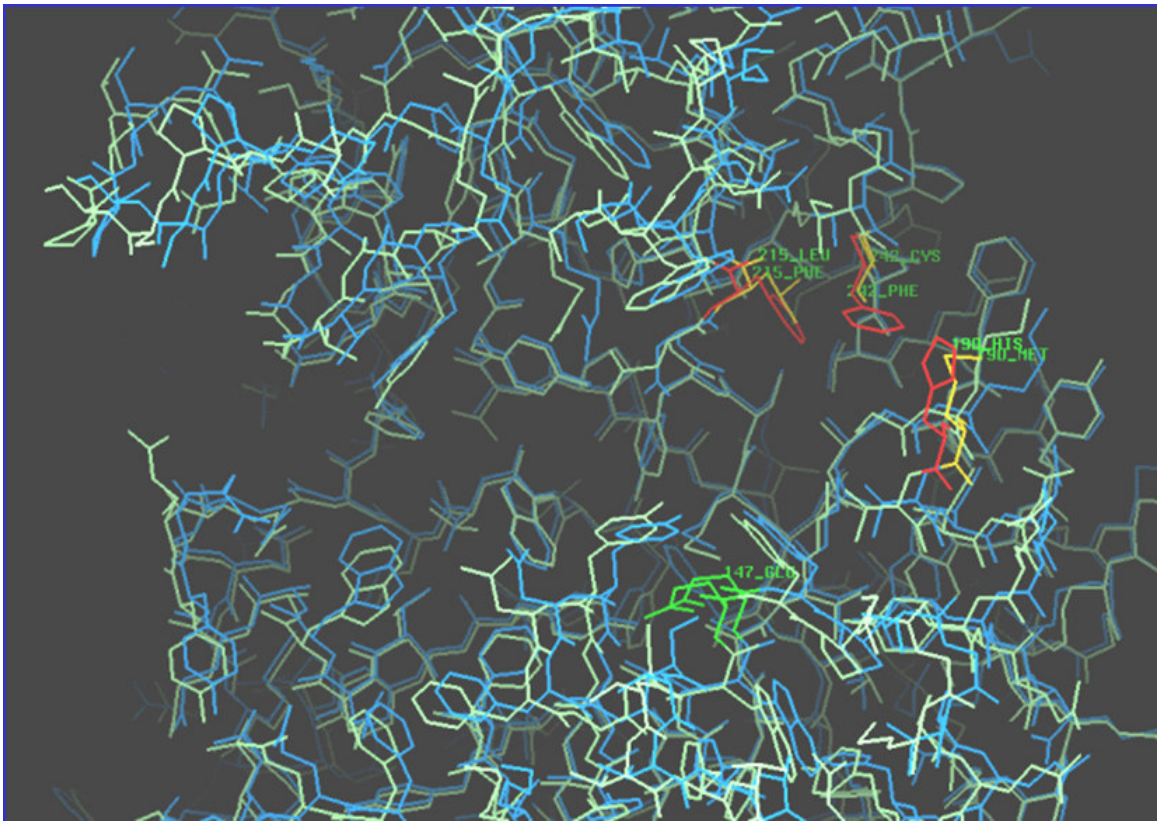


Fig. 2: Structural model of the active sites of *Acanthocheilonema viteae* (light green) and *Onchocerca volvulus* (blue) chitinases.

Non identical amino acids are shown in yellow for *A. viteae* and red for *O. volvulus* chitinase. The catalytic Glu is shown in green for both chitinases. Diagram was kindly provided by Prof. Dr. W. Höhne, Humboldt-University of Berlin.

No direct evidence exists to show that filarial chitinases are targets for drug intervention. However, chitinase molecules have been shown to be associated with parasite transmission and moulting (Wu et al., 1996; Fuhrman et al., 1987) and so will represent targets for transmission blocking drugs or vaccines. So far, the most effective inhibitor against family 18 chitinases is allosamidin (Sakuda et al., 1987). Due the cost of allosamidin, it is necessary to search for new and cheaper substitute

inhibitors. A prerequisite to the evaluation of potential inhibitors will be the production of soluble homogenous chitinase.

1.3 *A. viteae* as a model to investigate parasitological parameters of filarial diseases

Research in filariasis is dependent on the use of laboratory animal models owing to the inability to measure adult parasite population in humans. The filarial parasites are host-specific and so the first type of model system involves the use of parasites in surrogate models. The *Brugia spp.* / BALB/c mouse system has been used as a chemotherapeutic (Devaney et al., 1985) and immunological (Carlow and Philipp, 1987) model for the brugian filariasis. A similar approach with *Onchocerca spp.* involves the implantation of *Onchocerca spp.* in subcutaneous chambers in CBA/J or DBA/2J mice (Townson and Bianco, 1982; Abraham et al., 1992). These systems have a set back in that they rely on studies of a parasite in its non-natural host, a disadvantage that can be overcome by the use of full life-cycle models of the filaria *Brugia pahangi* in cats (Grenfell et al., 1991), bovine *Onchocerca spp*. or rodent filariae, like *Acanthocheilonema viteae* and *Litomosoides carinii,* in their natural hosts (reviewed in Abraham et al., 2002). Owing to the lack of adequate immunoreagents the underlying immunological mechanisms can however not be investigated in these model systems. The *B. pahangi* / cat model serves as a model for *Brugia* and *Wuchereria* in humans (Grenfell et al., 1991). This model system can be used to elucidate relationships between infection, immunity and disease states in lymphatic filariasis.

A. viteae in its natural host *Meriones unguiculatus* serves as a model for onchocerciasis, the disease caused by *O. volvulus*. This filarial model has some parallels to *O. volvulus* in that both reside in the subcutaneous tissue of their hosts and are therefore in the same immunological compartment. However, they do not form nodules and eye lesions like *O. volvulus*. Further more, both parasites share an array of antigenic similarities as demonstrated by anti-*O. volvulus* monoclonal antibodies (Nogami et al., 1986), and a high homology in corresponding molecules that have so far been characterised from both parasites. In addition, cross-protection between species has been shown in filariasis (Storey et al., 1982; Geiger et al., 1996) so that vaccine candidates established in one system could be tested in others. The *A. viteae* / *Meriones* system allows the study of resistance to challenge infection following immunisation (Abraham et al., 2002).

In the *A. viteae* / jird-model, it has been shown that immunisation with irradiated *A. viteae* L3 led to 90% protection against challenge infection, while immunisation with excretory-secretory products (ESP) led to 70% protection (Lucius et al., 1991). Parallel results were also obtained using irradiated L3s in other filariasis models (Lange et al., 1993; Taylor et al., 1994; Johnson et al., 1998; Trees et al., 2000) and it could be shown in this model that immunisation with irradiated L3 could also lead to resistance against homologous challenge infection (Abraham et al., 2002).

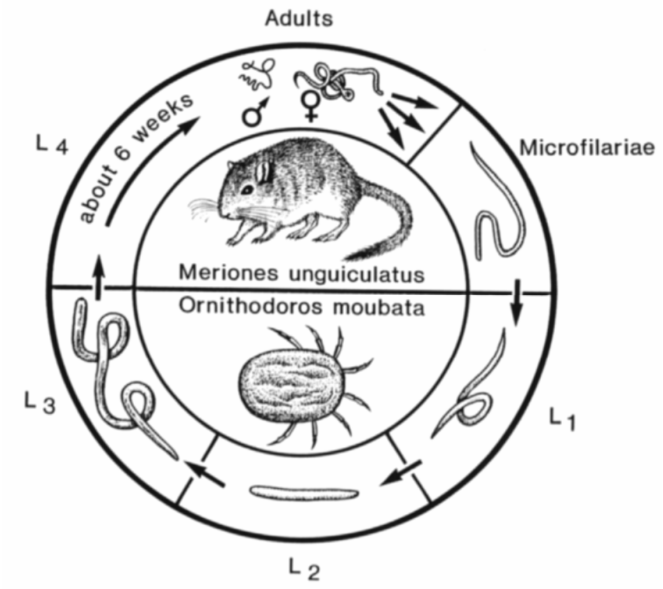


Fig. 3: Life cycle of Acanthocheilonema viteae

L1, L2, L3 and L4 are successive larval stages. Source: Archive of the Chair of Molecular Parasitology, Humboldt-University, Berlin.

1.4 Aims and objectives of the study

The aims of this study were three-fold. Firstly, *A. viteae* chitinase genes had to be isolated and characterised. To verify the hypothesis that there are several *A. viteae* chitinase genes, a cDNA library of female worms and L3, as well as cDNA from uterine microfilariae were screened for chitinase genes. Chitinase genes were investigated at the genomic level for copy number and gene organisation.

The second aim was the production of homogenous and enzymatically active *A. viteae* chitinase for use in inhibitor and structural studies.

The third aim was to evaluate the protective potentials of *A. viteae* chitinase and its synthetic peptides in immunisation studies. Since *A. viteae* chitinase activity has been demonstrated in two life stages of the filaria (microfilariae and infective larvae)

which could be targets for transmission blocking vaccines, another specific aim of this study was to design synthetic peptides from three-dimensional model of *A. viteae* chitinase for use in vaccination, in parallel with *A. viteae* N-terminal glycosyl hydrolase domain of chitinase.

2 Results

2.1 Characterisation of *A. viteae* chitinase genes

2.1.1 Determination of the gene structure of *A. viteae* chitinase genes

2.1.1.1 Isolation and characterisation of A. viteae *chitinase sequences from a genomic phage library*

A prerequisite to the determination of the complete genomic sequence of *A. viteae* chitinase was the isolation of clonal genomic recombinants containing the complete gene. To fulfil this requirement, an *A. viteae* genomic library in lambda dash II (Stratagene) was screened in plaque hybridisations using a DIG-labelled N-terminal cDNA sequence corresponding to nucleotides 52-1,117 of *A. viteae* L3 chitinase (Adam et al., 1996; Gene bank number U14638). The *A. viteae* genomic library (provided by Dr. Jörg Hirzmann, University of Gießen, Germany) was constructed in λ dash II using gDNA from a mixed population of adult worms. In the first screening at high stringency, $4x10^4$ plaque forming units (PFU) were investigated giving rise to 13 putatively positive recombinants. Following a second and tertiary screening round, 9 of the 13 plaques remained positive and were clonal.

2.1.1.2 Restriction analysis of genomic inserts

Inserts of recombinant phages were subcloned into the plasmid pBluescript for further analysis. The genomic inserts were digested out of the lambda arms, and subcloned directly into the Not I site of the plasmid, pBluescript. The advantage of such a strategy is the ease of preparing large amounts of genomic chitinase/ recombinant plasmid vectors for further analysis.

Restriction digestion of the pBluescript recombinants with Not I released genomic inserts ranging in size from 10-16 kb, as well as the 3 kb pBluescript plasmid (Figure 4). For restriction analysis, recombinant DNA was digested with Not I in combination with EcoRI (Figure 5). Three main restriction patterns were observed. The first pattern was found in clones 1, 2, 11 and 13, for which digestion of full-length versions (clones 1 or 13) gave rise to 4.5, 3.5, 2, 1.8, and 1 kb bands. The second pattern was found in clones 3, 4, 9, for which digestion of the full-length versions (clones 4 and 9) gave rise to fragments corresponding to 5, 4.5, 1.2, 0.9, 0.8, 0.6, and 0.4 kb. The third pattern was found in clones 5 and 12, for which digestion of the full-length version (clone 12) gave rise to 5.5, 2.5, 1.7, 1.2, 0.8, 0.6, and 0.4 kb bands. A pair of

full-length clones from each insert group (clones 1/13, 4/9, and 5/12) was used for further sequencing and analysis.

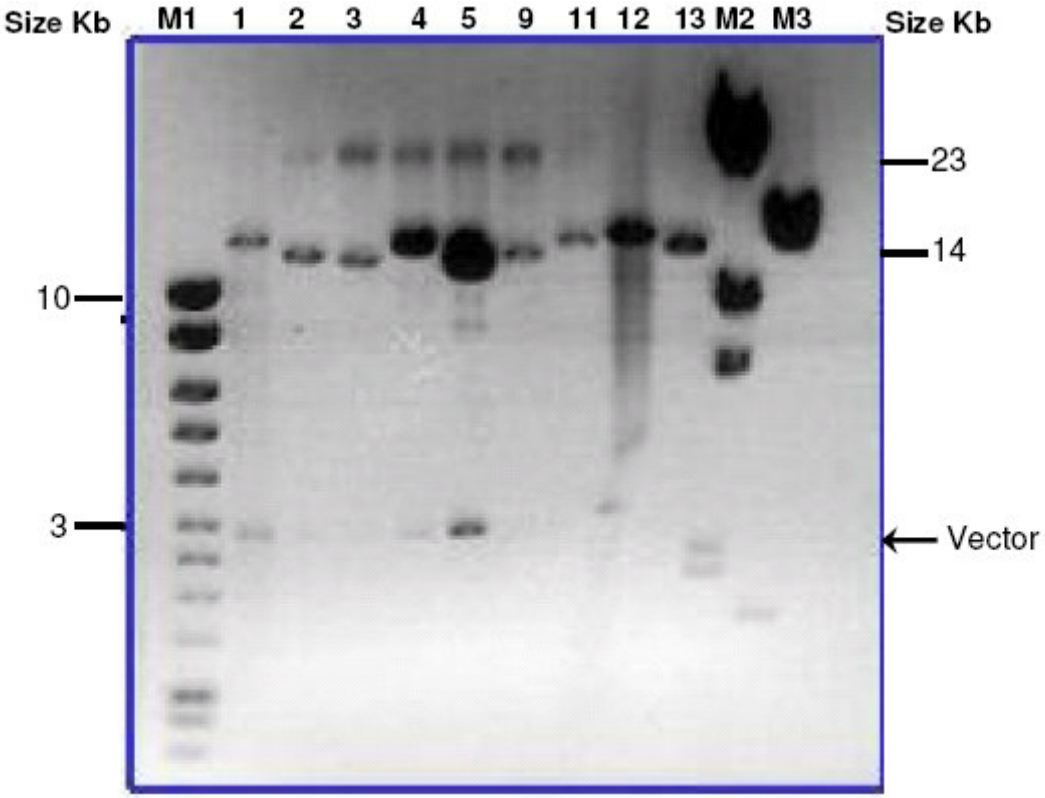


Fig. 4: Sizes of genomic inserts cloned into pBluescript

pBluescript recombinants were digested with Not I; Clone number 1-5, 9, 11-13: individual clones analysed, M1: low molecular size markers, M2: high molecular size markers, M3: 14 kb linearised plasmid.

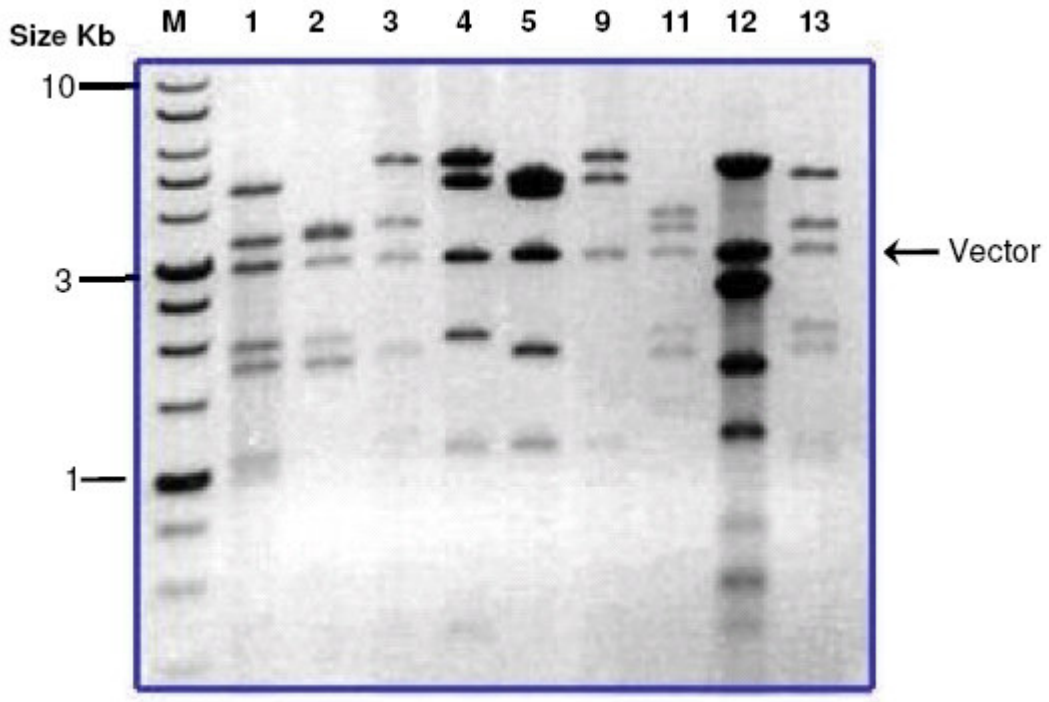


Fig. 5: Restriction analysis of *Acanthocheilonema viteae* genomic chitinase/ pBluescript recombinant plasmids.

pBluescript recombinants were digested with Not I and EcoR I; Clone number 1-5, 9, 11-13: individual clones analysed, M1: low molecular size markers.

2.1.1.3 Sequencing of A. viteae *genomic chitinase and determination of gene structure*

Since it was technically impractical to directly sequence inserts in lambda vectors, overlapping PCR fragments were directionally amplified from recombinant lambda DNA clones using gene-specific and vector-specific primers (Figure 6) and sequenced. Alternatively, constructs of genomic inserts in pBluescript were also sequenced directly by primer walking. Three representative genomic inserts from clones 1, 9 and 12, identified following restriction mapping were used for sequencing. Independent confirmatory sequencing runs were also done with clones 4, 5, and 13 to verify data obtained for clones 9, 12 and 1, respectively. Taking clone 12 as an example, gene-specific primers were used to sequence portions of the insert corresponding to the chitinase gene, using overlapping long PCR fragments (Figure 6) or plasmids with the insert. These sequencing strategies were also applied for genomic clones 1, 4, 5, 9 and 13.

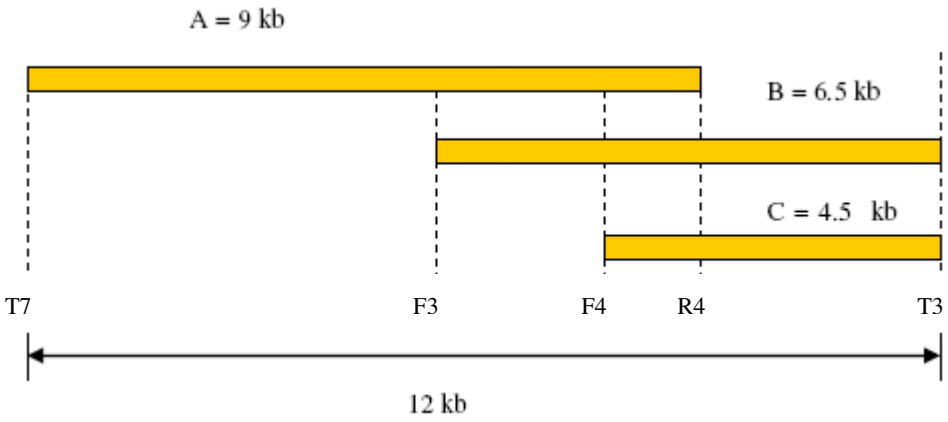


Fig. 6: Long-range PCR fragments used in sequencing of genomic insert 12.

Product A: amplified with primers T7-long range and R4 primer, product B: amplified with F3 and T3-long range primer, product C: amplified with F4 and T3-long primer.

Contiguous sequences were assembled from the nucleotide sequences obtained leading to the identification of three independent *A. viteae* chitinase genes (Figure 8). These sequences spanned about 14 kb for genomic clones 1/13, 14.5 kb for genomic clones 4/9, and 6 kb for genomic clones 5/12. The sequences were analysed and compared for the genomic organisation of the different chitinase genes using the published sequence for *A. viteae* L3 chitinase cDNA. A schematic diagram showing the extent to which the different insert groups were sequenced, and the genes predicted is shown in Figure 8. For further descriptive purposes, the chitinase genomic sequences have been named as gene sequence I, II, and III, respectively (Figure 8). Two sequences (I and III) have all features of a regularly transcribed gene: start methionine, followed by 12/13 exon sequences, and separated by 11/12 intron sequences respectively. Both sequences have a regular stop codon and a consensus polyadenylation signal. Generally, the splice donor and acceptor sites followed the common GT-AG rule (Senapathy et al., 1990), with the exception of introns e and n of gene II, and intron k of gene I that followed the GC-AG rule (Thanaraj and Clark, 2001) (Appendices 1,2 and 3) and offered for the possibility of alternative splicing.

To confirm the number of chitinase genes in the genome of *A. viteae*, Southern hybridisations were performed. A 700 bp dsDNA *A. viteae* chitinase probe was generated from *A. viteae* L3 chitinase cDNA, and hybridised under stringent

conditions to 2, 3 and 6 bands of *A. viteae* genomic DNA completely digested with Pvu II, Hind III and EcoR I, respectively (Figure 7).

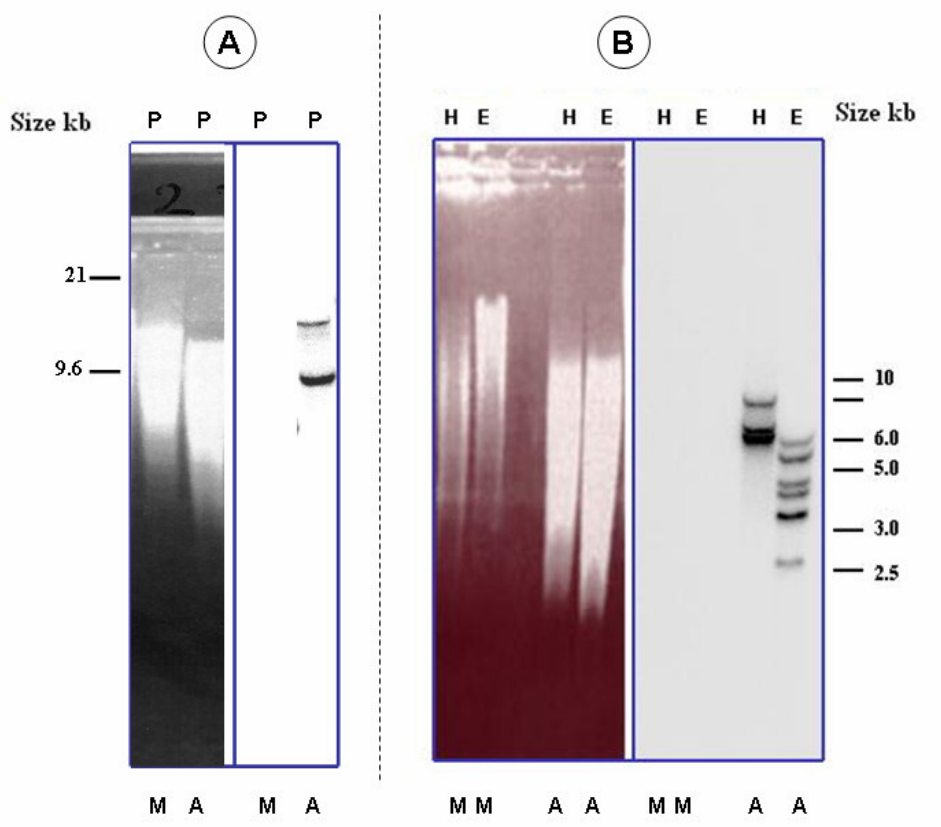


Fig. 7: Southern blot analysis of *Acanthocheilonema viteae* genomic DNA.

Meriones (M) and *A. viteae* (A) DNA (20 µg per lane) were respectively digested with Pvu II (P), EcoR I (E) and Hind III (H). Alternate results are shown for digestion of *Meriones* DNA as control, and *A. viteae* DNA on the left side of the picture, while hybridization results for both samples are on the right.

None of the restriction enzymes cut into the known cDNA sequence. In all three digested samples, the cDNA probe hybridised in different intensities to different bands. As completely digested DNA samples were used, the intensity of the bands seemed to be proportional to the number of probe molecules that bind to individual fragments. For the Pvu II digest that had two hybridisable bands, the band of approximately 15 Kb was recognised by less probe molecules than the fragment of 9.6 Kb, suggesting that there are more binding sites for the probe on the 9.6 Kb fragment. This would be consistent with a cluster of similar chitinase genes on the 9.6 restriction fragment. Further consideration of the Hind III hybridisation pattern suggests that there are at least three chitinase genes. Checking the corresponding

genomic sequences of chitinase gene I, II and III and adjacent regions up- and downstream, there are no Pvu II restriction sites, while there are two HindIII sites in each sequence that will each yield one hybridisable band within the size range observed. In addition, 6 hybridisable bands are expected for EcoRI, all falling within the range observed.

In summary, the Southern hybridisation results support the identification of three different chitinase genes in the *A. viteae* genome, for which two genes are clustered.

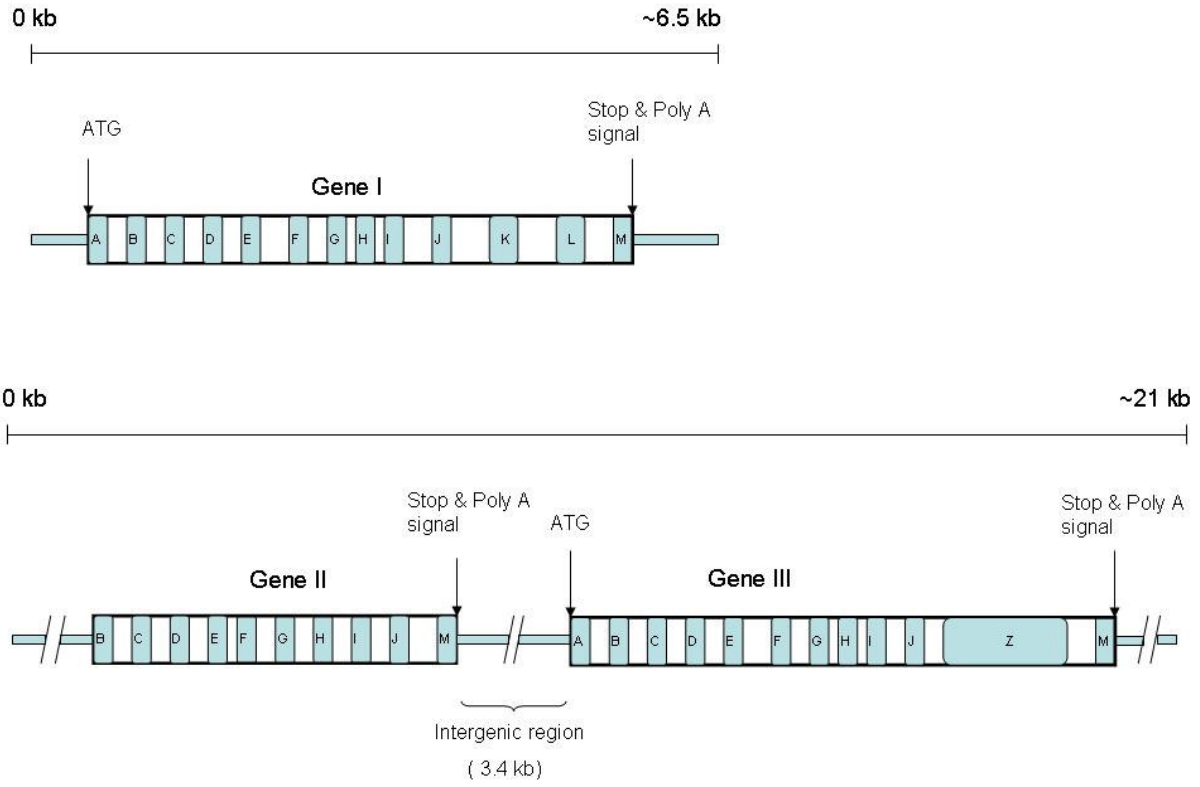


Fig. 8: Schematic diagram showing *Acanthocheilonema viteae* gene sequences.

Identified genes with lettered exons are shown in boxes, while intergenic regions are shown as grey lines. Gene sequences were obtained by sequencing different genomic clones as described in the text.

The 13 exons from gene sequence I have a 100% identity to the known cDNAs (Adam et al., 1996; Wu et al., 1996) as schematically represented in Figure 9. The signal sequence has two possible start methionines that have both been experimentally verified for gene I. The sequence can begin with the first start methionine (Wu et al., 1996) or with a second start methionine four amino acids

downstream (Adam et al., 1996). The Wu et al. sequence had additional 5' UTR of the cDNA which are found in the upstream region of the start ATG of gene I. This suggests that this sequence will be expressed in the worm. Individual exons build up different domains which represent different structural and functional parts of the molecule. Exon A contains the deduced N-terminal sequence necessary for the secretion of the protein. The sequence for a mature protein begins in exon B with the conserved YVRG sequence of filarial chitinases (Fuhrman et al., 1992). The active glutamate is found in exon E within the conserved amino acid sequence FDGFDLDW**E**YP (Synstad et al., 2004, Watanabe et al., 1993). The glycosyl hydrolase domain extends from exon B to exon J. This is followed by the serine-threonine region that has no direct relation to enzyme activity but has been demonstrated to facilitate enzyme secretion and to stabilize chitinase in the presence of proteolytic enzymes as shown in insect chitinases (Arakane et al., 2003). Finally exon M encodes a domain responsible for substrate binding (Fuhrman et al., 1992).

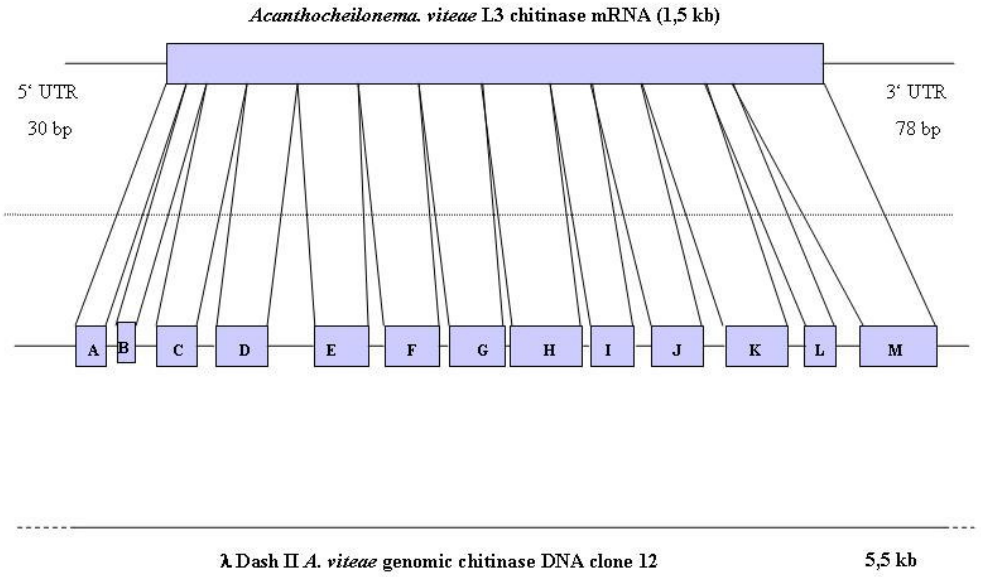


Fig. 9: Structure and organisation of *Acanthocheilonema viteae* L3 chitinase gene.

The upper panel shows the size of the cDNA, while the lower panel shows the positions and relative sizes of the 13 exons and 12 introns that make up the gene. Exons are shown as boxes, connected by introns shown as lines.

The exon / intron structure of gene sequence I and III are very similar. Therefore the exons and introns of gene I and gene III were compared as shown in Table 1. The amino acid sequences of the different exons show a high similarity of 77 to 97%.

Tab. 1: Comparison of exon/intron sizes and identities of *Acanthocheilonema viteae* genomic chitinase sequences I and III.

Exon	Size (aa)		Identities (%)		Intron	Size (bp)		Identities (%)
	Gene I	Gene III	NA	AM		Gene I	Gene III	
A	17	17	85	83	a	293	338	74
B	15	15	86	71	b	353	347	78
C	31	31	84	64	c	244	231	69
D	47	47	77	71	d	826	166	10
E	49	49	83	64	e	213	622	22
F	40	40	96	92	f	98	228	87
G	40	40	96	80	g	73	72	94
H	62	62	92	87	h	93	93	100
I	40	40	97	97	i	240	292	51
J	45	45	94	91	j	321	333	90
K	51	abs.	-	-	k	289	abs.	-
L	25	abs.	-	-	z	abs.	1276	-
Z	abs.	462	-	-	l	367	abs.	-
M	57	57	70	62				

Dashes represent identities not determined due to absence of comparable sequence in one of the genomic sequences. NA= nucleic acid; AM = amino acid; abs. = absent.

Overall, sequences I and III showed a relatively high homology. However, the greatest difference is in the region of exons K and L of gene sequence I. Exons K and L of gene I encode the serine-threonine rich region. This region is represented in gene III by one large exon named Z, which is also serine-threonine rich, but is about 6 times larger. The sequence of exon L can be aligned to the sequence of exon Z 19 times. Two different transcripts are therefore possible from both sequences: a 1.5 Kb transcript from gene I and a 2.7 Kb transcript for gene III giving a theoretical protein molecular weight of 55 kDa and 100 kDa, respectively.

Gene sequence II is apparently an incomplete gene without a regular start methionine, but having a stop codon and a consensus sequence for polyadenylation (Figure 8). An additional difference in gene II is the absence of the exons encoding for the serine-threonine rich region.

When gene II was compared with gene I and III, there was a higher amino acid identity between the exons of genes I and II (Table 2) relative to genes II and III

(Table 3). The identities of the intron sequences range from 10 to 100 % for comparable pairs.

Tab. 2: Comparison of exon/intron sizes and identities of *Acanthocheilonema viteae* genomic chitinase sequences I and II.

	Size (aa)		Identities (%)			Size (bp)		
Exon	Gene I	Gene II	NA	AM	Intron	Gene I	Gene II	Identities (%)
A	17	abs.	-	-	a	293	abs.	-
B	15	28	87	93	b	353	406	58
C	31	31	89	87	c	244	763	19
D	47	47	99	100	d	826	859	94
E	49	49	100	100	e	212	212	100
F	40	40	100	100	f	98	96	97
G	40	40	99	97.5	g	73	69	94
H	62	62	99	96	h	93	93	96
I	40	40	100	100	i	240	240	99
J	45	45	99	92	j	321	543	26
K	51	abs.	-	-	k	289	abs.	-
L	25	abs.	-	-	l	367	abs.	-
M	57	72	56	60				

Dashes represent identities not determined due to absence of comparable sequence in one of the genomic sequences. NA= nucleic acid; AM = amino acid; abs. = absent.

Tab. 3: Comparison of exon/intron sizes and identities of *Acanthocheilonema viteae* genomic chitinase sequences II and III.

Exon	Size (aa) Gene II	Size (aa) Gene III	Identities (%) NA	Identities (%) AM	Intron	Size (bp) Gene II	Size (bp) Gene III	Identities (%)
A	17	abs.	-	-	a	abs.	338	-
B	28	15	70	62	b	406	347	62
C	31	31	80	76	c	763	231	18
D	47	47	78	55	d	859	166	10
E	49	49	95	74	e	212	622	22
F	40	40	90	87	f	96	228	88
G	40	40	96	90	g	69	72	98
H	62	62	96	95	h	93	93	96
I	40	40	97	97	i	240	292	51
J	45	45	98	80	z	abs.	1276	-
Z	abs.	462	-	-	j	543	333	28
M	72	57	53	45				

Dashes represent identities not determined due to absence of comparable sequence in one of the genomic sequences. NA= nucleic acid; AM = amino acid; abs. = absent.

In conclusion, the characterisation of nine genomic clones from an *A. vieae* phage library and Southern blot experiments revealed the existence of three different chitinase genes for which at least two are closely associated and might form a functional transcript. The major differences between the three genes are the serine-threonine rich region (exons K and L versus Z) almost at the 3' end of the genes.

2.2 Comparative analysis of *A. viteae* genomic chitinase sequences with genomic data bases of nematodes

At the moment, information is available from two nematode genomic data bases: the whole genomic information from the finished *C. elegans* project (*C. elegans* sequencing consortium, 1998) and the on-going sequencing project from the *Brugia malayi* genome (available at http://www.tigr.org). A comparison with the genomic sequence of chitinase gene I led to the identification of Tigr assembly 14274 that had the genomic sequence of *B. malayi* chitinase (Fuhrman et al., 1992). This sequence, like *A. viteae* chitinase, is organised into 13 exons and 12 introns, with similar chitinase motifs and domains being represented on the same exons.

Tab. 4: Comparison of exons/ intron sizes and identities between *Acanthocheilonema viteae* L3 (Gene I) and *Brugia malayi MF 1* chitinases

Exon	Size (aa)		Identities (%)	
	A. viteae	*B. malayi*	NA	AM
A	20	21	68	52
B	15	15	85	93
C	31	31	74	80
D	47	47	90	77
E	49	49	94	74
F	40	40	81	82
G	40	40	92	75
H	62	62	90	79
I	40	40	85	82
J	45	45	92	77
K	51	25	32	27
L	25	15	33	24
M	57	57	70	68

Intron	Size (bp)		
	Av Chit	Bm Chit	Identities (%)
a	293	373	49
b	353	354	54
c	244	224	60
d	826	160	10
e	213	138	54
f	98	117	52
g	73	113	46
h	93	91	51
i	240	233	52
j	321	358	52
k	289	287	44
l	367	655	29

In both genomic sequences, exon A encodes for the secretory N-terminal signal sequence, exons B-J for the glycosyl hydrolase domain, exons K and L for the serine/threonine rich linker and exon M encodes the chitin binding domain. The size of all exons, except exons K and L are essentially similar (Table 4). The distance between the ATG start codon and the stop codon is 4,566 bases as compared to 4989 bases for *A. viteae* chitinase. The introns of *B. malayi* chitinase range in size from 91 bp (intron 8) to 655 bp (intron 12) as compared to 73 bp (intron 7) to 826 bp (intron 4) for *A. viteae* L3 chitinase (Table 4). The identities between introns of both genes range from 10% to 60%. The identities of nucleotides in corresponding exons are remarkably high, except for exons K and L that have a bare 30% identity. This kind of variation was also noticed for the analysed sequences of *A. viteae* chitinase genes. This high level of homology between the chitinase genes of these related filarial species is also shown by the conservation of most of the intron phases, with the exception of intron k (Jean et al., 2001; Guiliano et al., 2002).

The situation in *A. viteae* chitinase gene sequence II in which there is no start methionine is comparable to a *Brugia malayi* genomic chitinase sequence in BAC

clone 1133599. In both sequences, there is no regular first exon A, and exon B has no regular 5' exon/intron boundary. While *A. viteae* chitinase gene II has the last exon, a regular stop codon and polyadenylation signal, the *B. malayi* sequence goes only as far as exon J. The remaining part of the sequence has homologies to nematode retroviral elements like *Ascaris lumbricoides* proretrovirus-like element (Felder et al., 1994).

In addition to these sequences, there are other large genomic assemblies with anotated genomic chitinase sequences in different genomic neighbourhoods. Tigr assembly 14932 has two chitinase genes in tandem as found on *A. viteae* genomic clones 1 and 9 (Gene II and III). The first gene (512) on this cluster has 8 exons and 7 introns and should encode only the glycosyl hydrolase domain, while the second gene (514) has 10 exons and 9 introns and should encode for all domains as found in *A. viteae* chitinase gene I.

In conclusion, *B. malayi* has two annotated chitinase genes, which have similar organisation and domain architechture like *A. viteae* chitinase gene I. This however does not exclude the existence of other chitinase genes owing to the presence of additional 14 genomic assemblies with exons having homologies to filarial chitinases.

A search in the data bank of the free-living nematode *Caenorhabditis elegans*, WormBase (http://www.wormbase.org/), revealed an alignment with the highest probability for the *C. elegans* sequence C04F6.3. The gene, unlike all three described *A. viteae* chitinase, has only four exons spanning 2.22 kb of genomic DNA (Figure 10). The potential cDNA encodes a protein with 617 amino acids and a molecular weight of 66.9 kDa. Like filarial chitinases, this chitinase contains one glycoside hydrolase (family 18) domain, and a chitin-binding domain. Comparison of the gene structure revealed a comparable domain for exon A responsible for the secretion of the functional protein. The glycosyl hydrolase domain containing the catalytic site of chitinase, represented in *A viteae* by 9 different exons (B-J) is represented by exons 2 and 3 of the *C. elegans* chitinase gene. Thereby, exon 3 has an unusual size of 1313 nucleotides. Exon 3 encodes not only the glycosyl hydrolase domain, but contains also part of the serine-threonine linker region. Exon four of the *C. elegans* genomic chitinase sequence is clearly associated with the chitin binding domain as also represented by exon M of the *A. viteae* genomic sequences.

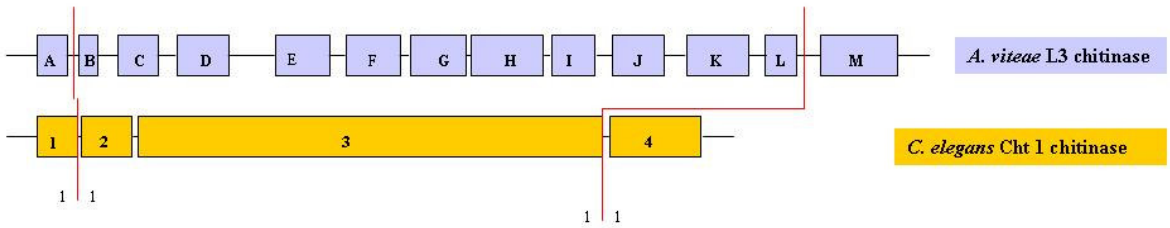


Fig. 10: Comparison of genomic organisation between *Acanthocheilonema viteae* and *Caenorhabditis elegans* chitinase genes.

Exons are shown as coloured boxes separated by lines representing introns. Red lines show two conserved intron phases with the phases indicated about the line. Intron phases are defined relative to the position of introns falling between codons (phase 0), within codon after first (phase 1) or second (phase 2) nucleotides, respectively (Jean et al., 2001a).

Despite the high degree of similarity of certain domains, there is a bigger difference in exon/ intron organisation and position of introns. But, in addition to the high homology with filarial chitinases from *B. malayi* and *A. viteae*, is also the fact that two of the three possible intron phases are conserved (Figure 10), revealing the high degree of relationship between the worm chitinase gene sequences.

2.3 *A. viteae* chitinase transcripts

The analysis of the genomic *A. viteae* chitinase genes showed the existence of at least two potential transcripts from sequences I and III. In order to determine *A. viteae* chitinase transcripts, different approaches were used like the screening of *A. viteae* cDNA libraries as well as the specific amplification of *A. viteae* chitinase nucleic acid sequences from reverse transcribed RNA material. In addition, western blot experiments were performed with material from L3 stages and compared to already existing data. Finally, using a bioinformatic approach, the resulting data were compared with information from various nematode data bases.

2.3.1 Screening of *A. viteae* L3 and adult female cDNA libraries for chitinase transcripts

In order to identify different *A. viteae* chitinase transcripts and to check the stage specificity, two cDNA libraries made from the L3s and from gravid female worms were separately screened in plaque hybridisations using an *A. viteae* cDNA specific probe. The *A. viteae* chitinase probe was a 700 nucleic acid sequence from *A. viteae* chitinase L3 cDNA, corresponding to the exons F to J of *A. viteae* chitinase gene sequence I. The cDNA probe had a 95% identity to the nucleic acids of the

corresponding exons in the *A. viteae* gene sequence III. For both cDNA libraries, 40,000 independent recombinants were screened. The screening procedure resulted in four positive plaques from the screening of the L3 cDNA library, but no positive plaques in the case of the female cDNA library. The resulting four clones of the L3 cDNA revealed a similar size of 1,500 bp for the individual inserts (data not shown). The insert of two of these four clones from the L3 library were *in vivo* excised and sequenced. Both sequences corresponded to the already published cDNA sequence for the L3 stage with minor differences. There are two published L3 sequences (Adam et al., 1996; Wu et al., 1996). These are similar along the length of the molecule with the exception of the extreme 5' end. The cDNA from Wu et al. has an additional alternative start methionine, followed by four more amino acids with an alternative start methionine. Both independently sequenced cDNA inserts from the cDNA library showed identical 5' ends as published by Wu et al. (Appendices 1, 2 and 3). In spite of the published nucleic acid sequences from Adam and Wu et al., there are some minor differences in the chitinase insert sequences identified from the L3 library (Table 5).

Tab. 5: Nucleotide differences between published *A. viteae* chitinase and the chitinase sequence used in this study

A. viteae chitinase		Changes		
Published	This work	Structural	Codon*	Functional
CAA	CAG	A to G transition (wobble)	29	none
GCA	ACA	G to A transition, A to T	406	none
ACC	CCC	A to C transversion, T to P	437	none
---	CAG	Insertion; Q	438/439	none
CAC	CGC	A to G transition, H to R	477	none
CAC	CAT	C to T transition (wobble)	493	none
CTT	TTT	C to T transition, L to F	501	none

The different nucleotides are indicated and the structural changes resulting in different amino acids are indicated in single letter code. Codon numbering is relative to *Acanthocheilonema viteae* cDNA chitinase (Adam et al., 1996)

Apart of the wobble in codon 29, all other differences are found in the variable serine-threonine rich region. The only difference that could have come from alternative splicing as deduced from genomic gene I was the insertion between codons 438/439. All the other differences were confirmed by sequencing independent clones and could represent allelic differences. When all the changes are considered, it may be surmised that the two sequences could code for chitinases with slightly different physical properties, but with similar enzymatic properties since the changes do not occur in sites responsible for enzyme activity.

2.3.2 Identification of *A. viteae* chitinase transcripts

A total of 21 putatively positive plaques were obtained by screening a female *A. viteae* cDNA library. However, all clones did not yield any amplicons in additional PCR screens using vector and chitinase primers, nor did they turn out positive in secondary screens.

Since the screening of a female cDNA library failed to give any chitinase transcripts, the next step to identify different chitinase transcripts form different parasite stages involved the use of RNA from uterine microfilaria and L3 stages for RT-PCR analysis. Accordingly, RT-PCR was done with total RNA of uterine *A. viteae* microfilariae and L3 using primer pairs designed to amplify segments of *A. viteae* chitinase.

RT-PCR was done with L3 and uterine microfilarial cDNA using specific primers for gene sequences I and III located on exons I and M, respectively. In the first PCR amplification experiment, two faint bands of 1.8 and 0.6 kb were amplified with primers specific for gene III, and one amplification product of 0.6 kb resulted from the PCR with the gene I specific primers (Figure 11). A nested PCR with gene III specific primers did not yield specific products, while primers specific for gene I gave rise to a specific PCR product. Identical results were obtained with RNA from both L3 and uterine microfilariae. The 0.6 kb fragment amplified with the specific primers for gene I from the microfilarial stage was subcloned and sequenced and shown to be 100 % identical to the corresponding sequence for *A. viteae* L3 chitinase (Adam et al., 1996; Wu et al., 1996).

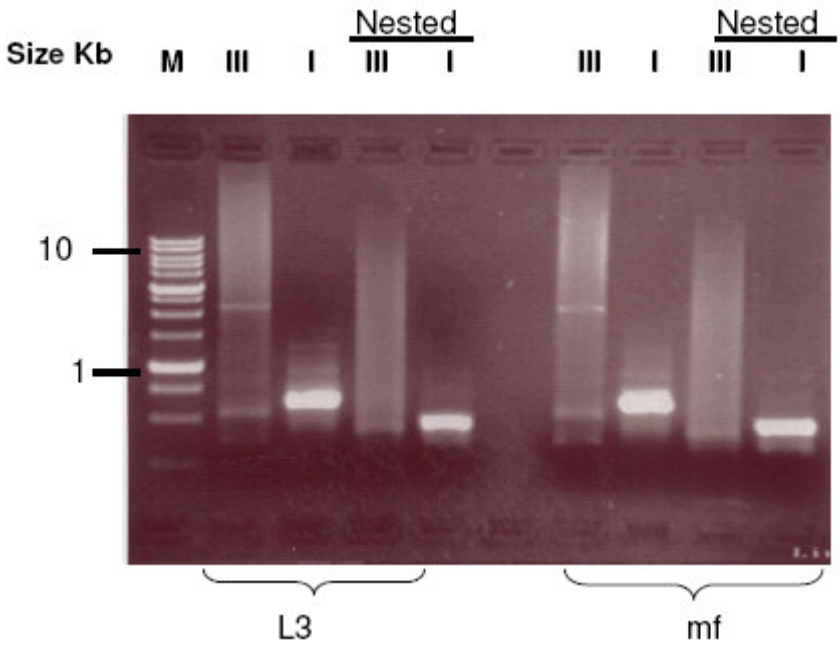


Fig. 11: Transcripts of *A. viteae* genomic sequences in L3 and microfilariae (mf)

RT-PCR results were obtained with primers specific to genomic sequences I and III. Corresponding sequence-specific primers (F8 and R3) were used in RT-PCRs, while a common forward primer (F5) was used with specific reverse primers (R3) in nested PCRs. M, Molecular size marker.

The fragment did not contain the terminal ends of the corresponding *A. viteae* L3 transcript. As the differences between sequences I and III lie particularly in the 3' end, a 3' RACE PCR was done with RNA from uterine microfilariae, L3 and L4 to determine possible variations in the 3' end. A major PCR product of 700 bp was obtained from all three stages as shown by the representative gel in Figure 12. The 700 bp product from mf cDNA was subcloned in pGEM-T vector, checked by sequencing and shown to be 100% identical to the already published 3' end of gene I. The products from L3 and L4 were confirmed by nested PCR to be transcripts of gene I.

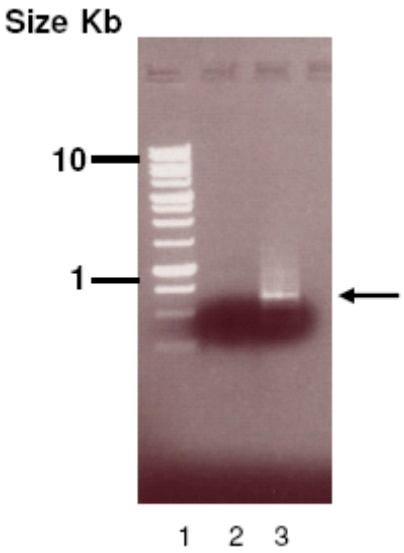


Fig. 12: Amplification of 3' end of cDNA chitinase from different stages using Rapid amplification of cDNA ends (RACE)

Lane 1, molecular size markers; lane 2, negative control; lane 3, amplified chitinase fragment (indicated by arrow).

In conclusion, a transcript for gene I could be amplified from uterine microfilariae, L3 and L4, while transcripts for gene III could not be found in these same stages. However, it can not be excluded that chitinase transcripts are found in other stages.

2.4 Identification of homologous chitinase sequences from nematode databases

A blastx analysis of public data bases with cDNA sequences corresponding to exons of genomic gene I revealed high homologies with orthologous transcripts of filarial (Table 6) and *C. elegans* nematodes (Table 8).

Tab. 6: Orthologous filarial transcripts to *A. viteae* gene sequence I

Sequence accession number	Source of chitinase sequence	Blastx e-value to gene I
L42010	*Acanthocheilonema viteae* (L3)	7.3e-267
AF250997	*Wuchereria bancrofti* (mf)	1.4e-197
M73689	*Brugia malayi* (mf)	5.0e-194
L42021	*Onchocerca volvulus* (L3)	2.1e-186
U59689	*Brugia pahangi* (mf)	4.1e-166
U59690	*Brugia pahangi* (mf)	8.4e-163
NM_076187 (C04F6.3)	*Caenorhabditis elegans* (embryos)	1.9e-93

With the exception of *Brugia pahangi*, all the other filariae have one described chitinase sequence. *B. pahangi* has two micorfilarial chitinase transcripts which differ in the serine threonine rich linker region (Figure 13). Comparison of the filarial amino acid sequences of *Wuchereria bancrofti, Brugia malayi, B. pahangi,* and *Onchocerca volvulus* to the *A. viteae* L3 chitinase sequence revealed an apparently high identity in the overall amino acid composition, with the exception of a region at the 3' end that corresponds to the serine-threonine rich liker region between the glycosyl hydrolase domain and substrate binding domain (Figure 13).

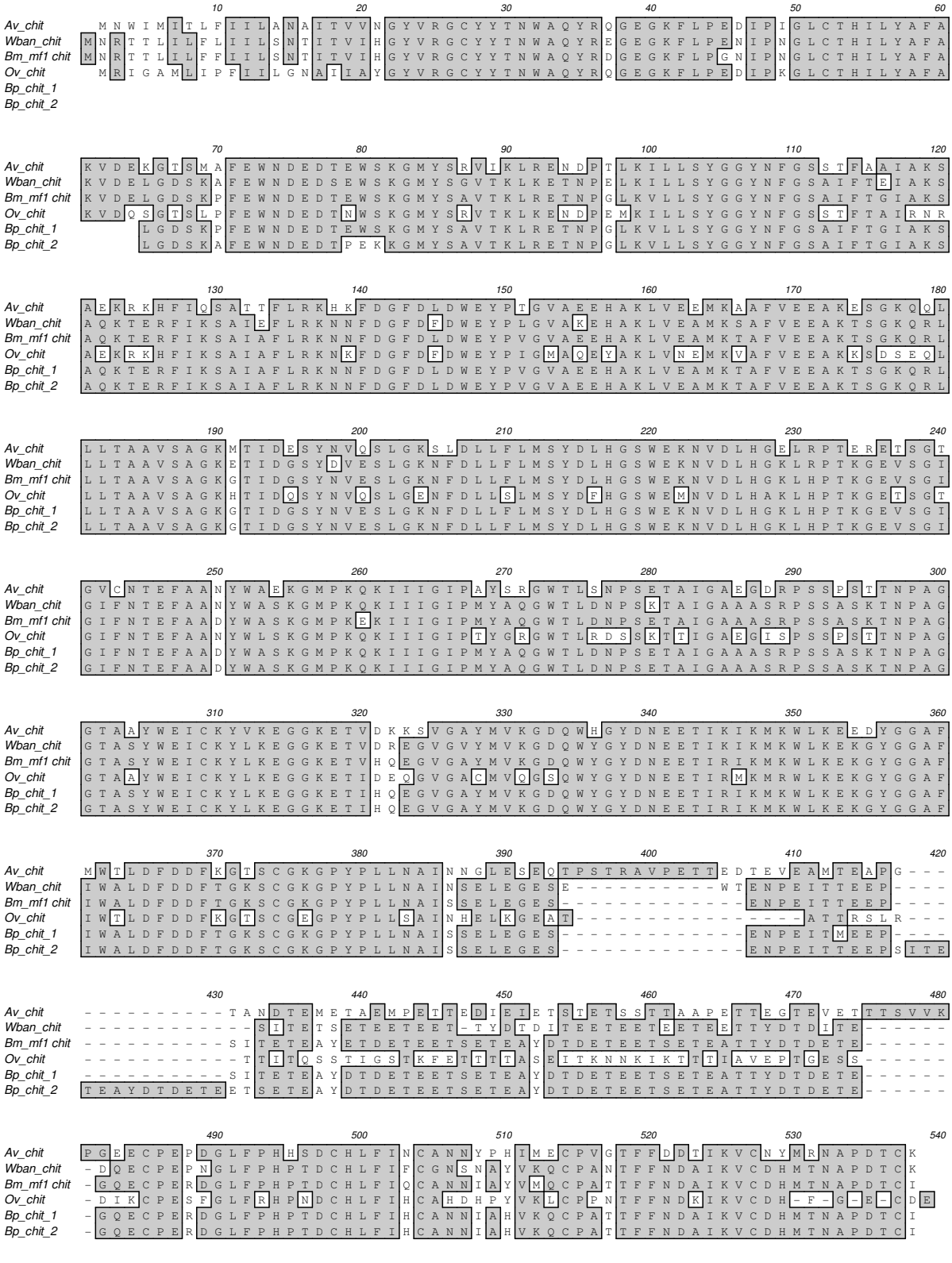


Fig. 13: Amino acid sequence comparison of different filarial chitinases.

Av_chit: *Acanthocheilonema viteae* chitinase (Wu et al., 1996); Wban_chit: *Wuchereria bancrofti* chitinase (Raghavan et al., 1994); Bm_mf1chit: *Brugia malayi* chitinase (Fuhrman et al., 1992); Ov_chit: *Onchocerca volvulus* chitinase (Wu et al., 1996); Bp_chit: *Brugia pahangi* chitinase (Arnold et al., 1996).

A detailed analysis of the EST databases of *B. malayi*, *W. bancrofti* and *O. volvulus* revealed that *B. malayi* databases contain the highest number of chitinase–like ESTs. A clustering of these ESTs resulted in three contigs. One of the three contigs in *B. malayi* is complete and corresponds to the already published sequences for *B. malayi* microfilaria chitinase (Fuhrman et al., 1992). The other contigs are not identical to the published chitinase sequence (Tab 7). In conclusion, *B. pahangi* has two published microfilarial chitinase mRNAs, *W. bancrofti* one microfilarial chitinase, *O. volvulus* and *A. viteae* each has one L3 chitinase mRNA. *B. malayi* has a total of three chitinase sequences with a possible difference in stage-specific expression.

Tab. 7: Brugia malayi ESTs

Cluster or Sequence	ESTs		Status of transcript
	Number	Stage	
BMC00298	AA022373 AA022418 AA246080 AA246111 AA598350 H48963 N44465	Microfilariae	*B. malayi* microfilaria chitinase (M 73689)
	AA257830	Female	
BMC12278	AA054927	Microfilariae	Not yet characterised: lysin (K) substitutes glutamate (E) at active site
BMC18355	CD570590	2-day irradiated L3	Not yet characterised. Identities to other chitinases: 25% to *A. viteae* chitinase; 32% to *B. malayi* microfilarial chitinase.

A blastx analysis of the *C. elegans* database led to 21 confirmed *C. elegans* sequences with some similarity to *A. viteae* chitinase (Table 8).

Tab. 8: *C. elegans* homologues of *A. viteae* chitinase gene sequence I

Wormpep number	Secretory signal sequence	Blastx e-value to sequence V
C04F6.3	yes	6.5e-102
F07G11.9	yes	1.9e-31
T13H5.3	yes	3.9e-25
Y22D7AL.14	no	2.2e-23
F15A4.8b	no	3.4e-20
F15A4.8a	no	1.5e-15
R09D1.10	no	1.5e-14
R09D1.5	no	2.9e-13
R09D1.8	no	3.2e-12
R09D1.3	no	1.1e-11
C45E5.3	no	2.7e-11
R09D1.11	no	8.4e-11
C25A8.4	yes	2.2e-10
R09D1.9	no	4.9e-10
C08H9.12	no	2.1e-09
C08H9.14	yes	3.1e-09
C08H9.13	no	4.3e-09
T19H5.1	no	9.2e-09
C08H9.11	no	1.5e-08
M176.8	no	1.7e-07
C08H9.7	no	3.9e-07

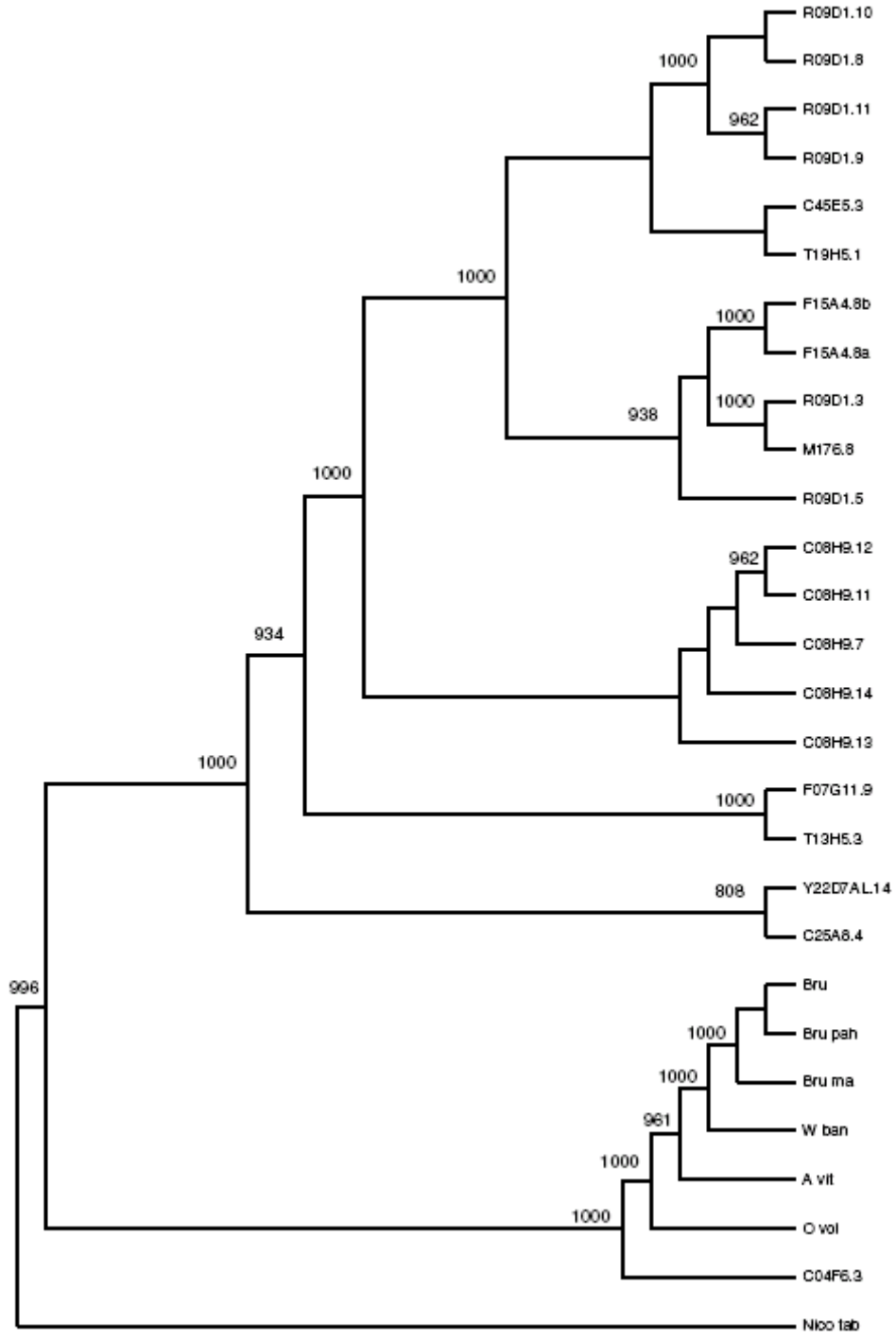


Fig. 14: Phylogenetic relationships of filarial chitinases with *C. elegans* chitinases

The bootstrapped tree was rooted using chitinase from *Nicotiana tabacum* (Nico tab) (Gene bank: X78325) as outgroup. Bootstrap values for 1000 replicates are shown where they are significant (>700). Chitinase sequences are from the following organisms: *Acanthocheilonema viteae* (A vit), *Brugia malayi* (Bru ma), *B. pahangi* (sequence 1= Bru; sequence 2 = Bru pah), *Onchocerca volvulus* (O vol) and *Wuchereria bancrofti* (W ban). All other sequences are from *Caenorhabditis elegans*.

C04F6.3 represents the *C. elegans* chitinase with highest homology and phylogenetic relationship to the filarial chitinases.

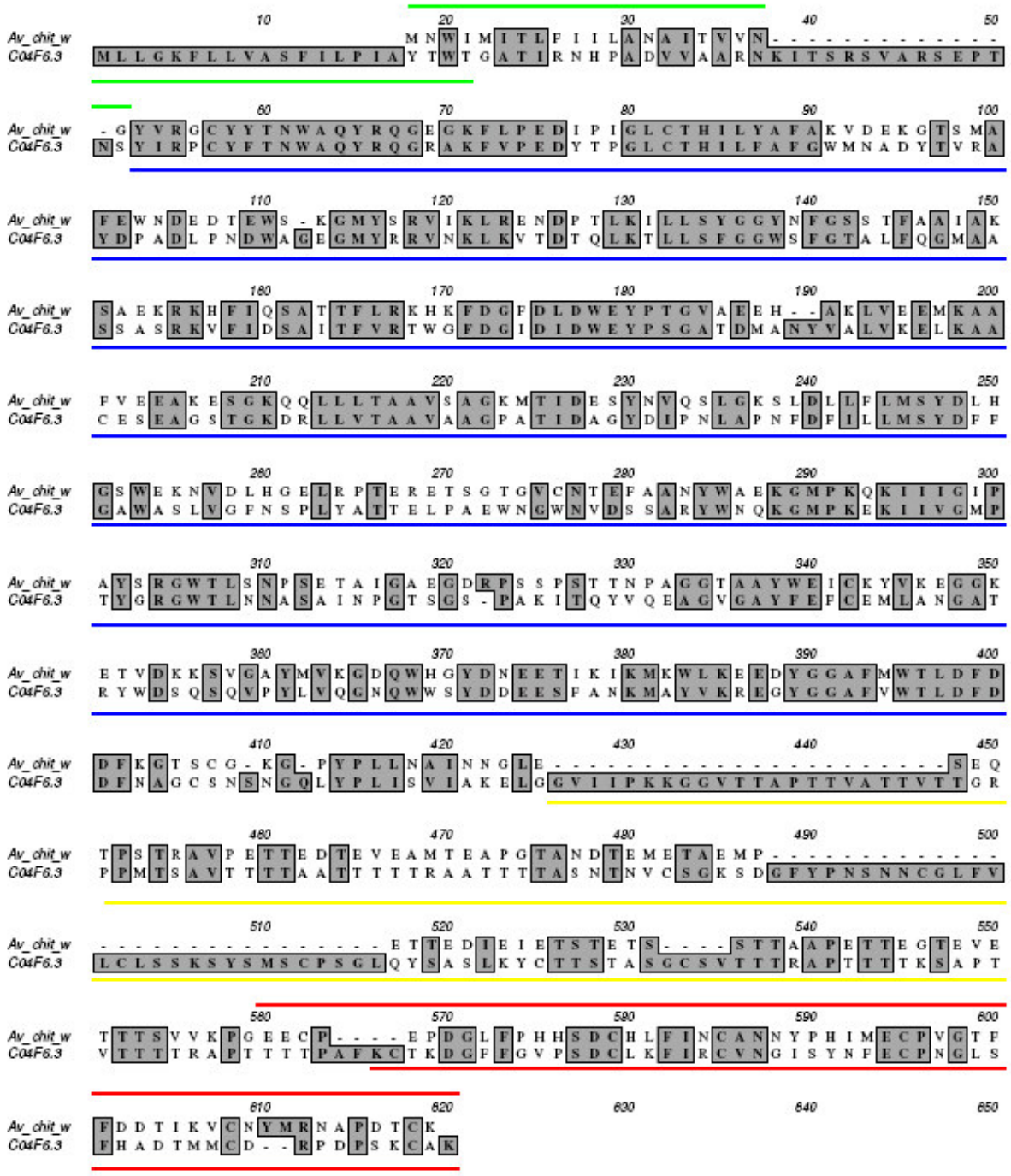


Fig. 15: Amino acid sequence comparison of *Acanthocheilonema viteae* chitinase with *Caenorhabditis elegans* sequence C04F6.3.

Regions corresponding to different domains are indicated with a line closely apposed to the sequence: Green, signal sequence; Blue, glycosyl hydrolase domain (the same for both sequences); Yellow, serine-threonine rich linker; Red, chitin binding domain.

Following phylogenetic analysis of confirmed *C. elegans* chitinase with other filarial chitinase sequences (Figure 14), *C. elegans* sequence C04F6.3 clustered with the filarial chitinases from *A. viteae*, *B. malayi*, *B. pahangi*, *O. volvulus* and *W. bancrofti*.

An alignment of C04F6.3 with *A. viteae* chitinase is shown in Figure 15. There is a great difference in similarity between the two derived protein sequences. Though the overall identity between both sequences is barely 34 %, a general analysis reveals the presence of the four important motifs described for family 18 chitinases: there is a secretory signal sequence, a glycosyl hydrolase domain, a serine-threonine rich linker region, and a substrate binding domain. The important consensus sequence of the active site is fully conserved.

In conclusion, *C. elegans* has several family 18 chitinase genes, one of which is closely related to the filarial chitinases.

2.5 Cloning, expression and purification of recombinant *A. viteae* chitinase

2.5.1 Cloning, periplasmatic/ cytoplasmatic expression and purification of N-terminal portion of *A. viteae* chitinase in the expression vector pET 22 b (+) in *E. coli*

The DNA sequence of the glycosyl hydrolase domain, corresponding to nucleotides 52-1,117 of *A. viteae* L3 chitinase was subcloned into the pET 22 b(+) vector, leading to the expression of a protein with an N-terminal vector-specific bacterial signal sequence and a C-terminal 6- His tag for purification. The bacterial signal sequence allows for potential periplasmatic localisation of the expressed protein and subsequent release of the bacterial signal sequence by bacterial signal peptidases. This construct was transformed into BL 21 (DE3) *E. coli* cells, and expressed as a protein of 43 kDa representing about 40 % (about 7 mg) of the total bacterial biomass, following induction with IPTG as shown in Figure 16. The protein could be exported into the periplasmatic space, released through osmotic shock and purified by affinity chromatography from the osmotic shock fluid (Figure 16). As shown on SDS-PAGE in Figure 16, there is a main band representing N-terminal chitinase without any signal sequence, while there is a minor band that runs slightly higher and should represent the soluble cytoplasmic fraction with the intact bacterial signal sequence. The total soluble protein purified represented about 15 % of the total protein expressed corresponding to 1.1 mg protein from 500 ml bacterial culture.

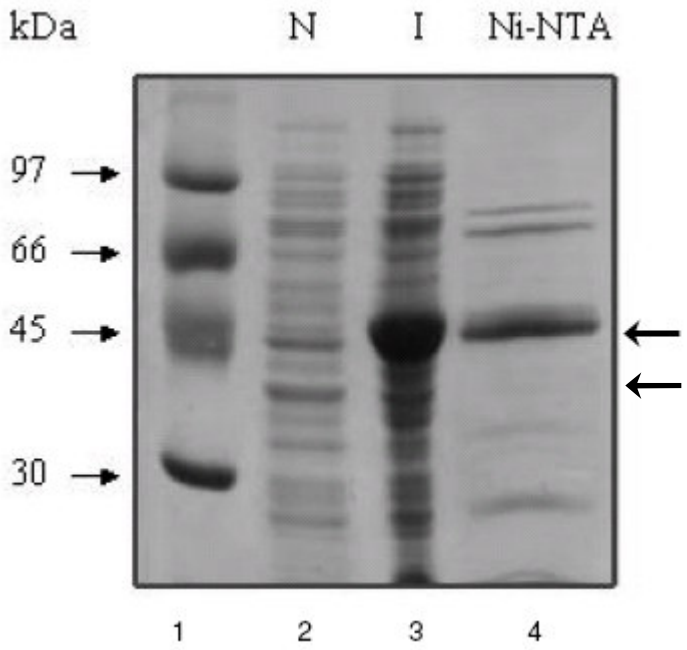


Fig. 16: Expression and purification of *A. viteae* chitinase from the periplasmatic space

Lane 1, low molecular weight marker; lane 2, lysate of non-induced bacteria (N); lane 3, lysate of induced bacteria (I); lane 4, eluted fraction from a Ni-NTA column of A. viteae chitinase (arrow)

The insoluble cytoplasmic fraction was solubilised in 8 M urea and also purified by affinity chromatography (Figure 17). The purified protein in urea was then dialysed successively against decreasing concentrations of urea in phosphate buffer. A greater part of the protein (4mg) precipitated during dialysis giving rise to just over 1mg soluble protein per 500 ml bacteria culture.

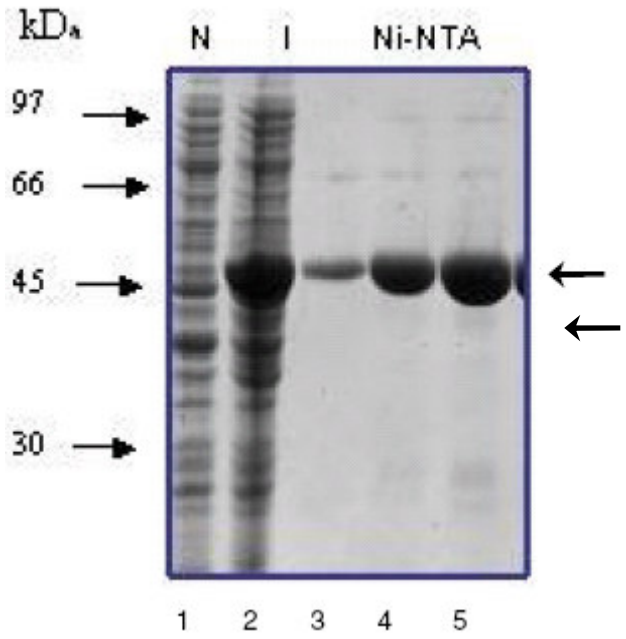


Fig. 17: Expression and purification of *A. viteae* chitinase under denaturing conditions

Lane 1, lysate of non-induced bacteria (N); lane 2, lysate of induced bacteria (I); lanes 3-5, fractions of A. viteae chitinase eluted from Ni-NTA column.

2.5.2 Activity of N-terminal chitinases

The activity of the N-terminal fragment of chitinase, purified from the cytoplasmic as well as from the periplasmatic space, was assessed using the substrate 4-methylumbelliferyl ß-N',N'',N''' chitotrioside that releases the fluorescent product 4-methylumbelliferone upon cleavage by a glycosyl hydrolase. Recombinant N-terminal *A. viteae* chitinase, purified under denaturing conditions and dialysed against phosphate buffer, was shown to have chitinase activity that had about 50 times less specific activity (defined as relative fluorescent units per hour per ng protein) as compared to the positive control, which was Baculovirus expressed *O. volvulus* chitinase (New England Biolabs) (Figure 18, Table 8). The negative control protein, *A. viteae* cystatin (Av17) did not hydrolyse the substrate.

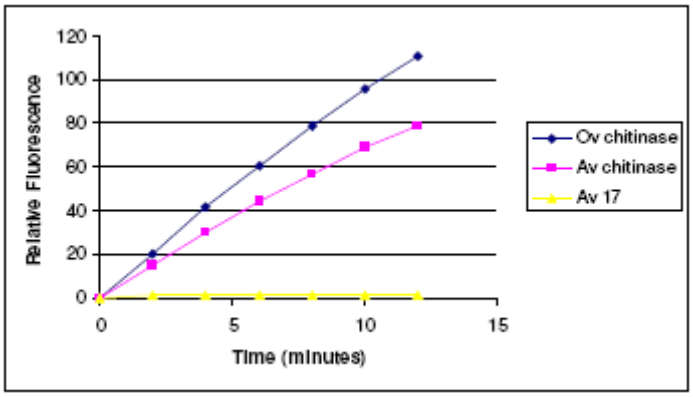


Fig. 18: Activity assay of purified cytoplasmic N-terminal chitinase

Activity assays were done with *A. viteae* chitinase (pink line), *O. volvulus* chitinase as positive control (blue), and Av 17 (filarial cystatin) as negative control. The amounts of protein assayed were 1.043 ng *O. volvulus* chitinase, 38.5 ng *A. viteae* chitinase and 38.5 ng Av 17.

Tab. 9: Specific activity of purified *A. viteae* and *O. volvulus* chitinases

Enzyme	Specific activity (Units*/hr/ng)
A. viteae N–terminal chitinase (Cytoplasma)	11.1
A. viteae N-terminal chitinase (Periplasma)	166.14
O. volvulus chitinase (NEB)	553.8

The specific activities of *A. viteae* and *O. volvulus* chitinases are shown. *The specific activity is reported as relative fluorescence units per hour per ng protein used.

The activity of *A. viteae* N-terminal chitinase purified from the periplasmatic space was originally determined using *Serratia marcescens* chitinase (Sigma) as standard. The relative fluorescence units were estimated to have about 30 % of the activity of baculoviral expressed *O. volvulus* chitinase (New England Biolabs) (data not shown).

2.6 Immunisation studies with *A. viteae* chitinase in *Meriones unguiculatus*

For the evaluation of protective potentials of different vaccines, *Meriones* were vaccinated and subsequently challenged with 70 L3s of *A. viteae*. Two preparations were used in this study, namely recombinantly expressed glycosyl hydrolase domain

(N-terminal fragment) of *A. viteae* chitinase, and synthetic peptides designed from the active site of *A. viteae* chitinase.

2.6.1 Immunisation studies with *A. viteae* N-terminal chitinase fragment

Recombinant N-terminal chitinase of *A. viteae* was used in vaccination experiments with STP and Alum as adjuvants. Immunisation experiments were done following the scheme on Figure 19. Ten-week-old *Meriones* were given a primary immunization subcutaneously at the flanks with 25 µg protein, followed by two boosters on the 2nd and 4th weeks before challenge. *Meriones* were bled for sera on weeks –6, 0, 4, 8 and 12, and the microfilaria load was checked on the 8th and 12 weeks following challenge. *Meriones* were sectioned for retrieval of adult worms on the 12th week after challenge (Figure 19).

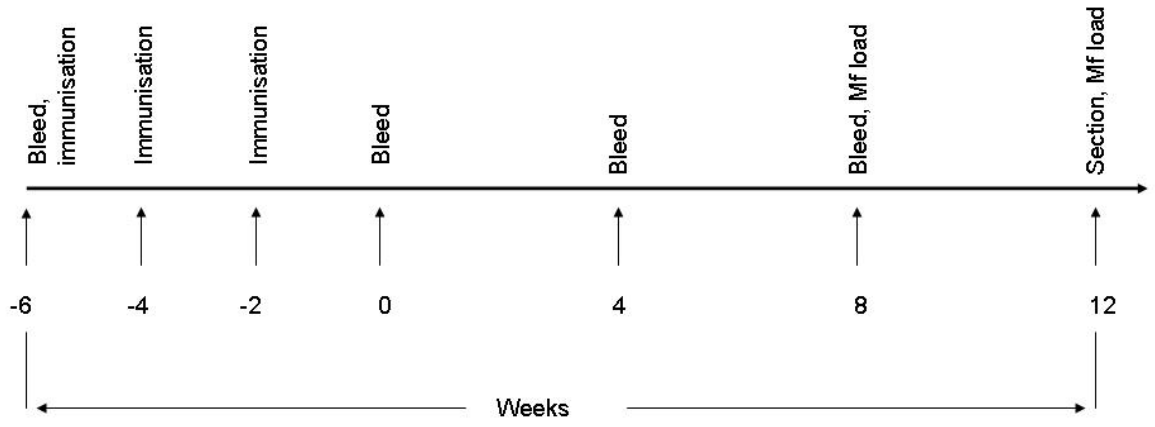


Fig. 19: Immunisation scheme of *Meriones unguiculatus* with experimental vaccines

Recombinant N-terminal *A. viteae* chitinase purified under denaturing conditions was used in the immunization of 10-week-old *Meriones* with STP and Alum as adjuvants. During the course of the experiment, one animal died from the chitinase/ Alum group, two from the chitinase/ STP group, three from the STP group and three from the PBS group (Table 10). Vaccination of *Meriones* with the glycosyl hydrolase domain of chitinase and STP as adjuvants led to a 29% significant overall reduction in worm burden ($p < 0.05$), while vaccination with Alum as adjuvant did not lead to any overall reduction in worm burden (Figure 20, Table 10).

Tab. 10: Summary of biometric data from immunization studies with N-terminal *Acanthocheilonema viteae* chitinase

Antigen/ control	Chitinase/ Alum	Chitinase/ STP	Alum	STP	PBS
Worms recovered	20.16±6.43	17.4±3.43	20.57±3.59	24.5±4.65	24.85±4.84
Female worms (Length)	13.83±4.4 (4.7±0.4)	11 ±2.5 (4.57±0.3)	13.57±3.4 (4.55±0.3)	14.25±2.7 (4.7±0.4)	13.42±3.4 (4.79±0.4)
Male worms (Length)	9.2±5.5 (2.08±0.1)	6.4±4.3 (2.35±0.1)	8.5±4.1 (2.2±0.1)	10.25±2.7 (2.19±0.1)	11.42±2.7 (2.2±0.2)
Mf density / ml blood	285±117.8	216.2±46.2	265±100.2	220±116	432.4±350
Surviving animals	6/7	5/7	7/7	4/7	7/10

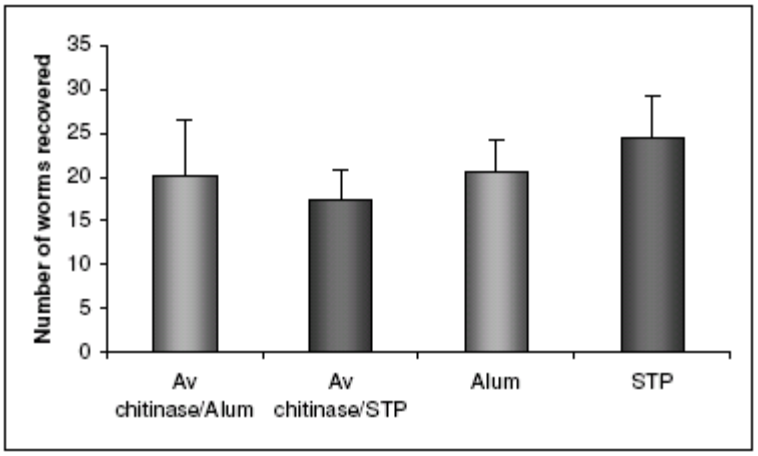


Fig. 20: Immunisation of *Meriones* unguiculatus with N-terminal *Acanthocheilonema viteae* chitinase

The number of worms recovered following immunization with N-terminal chitinase (*rA. viteae* chitinase) and STP and Alum as adjuvants are shown. Controls: Alum and STP.

When the microfilaria load was compared, there was no statistically significant difference in the microfilarial burden between experimental and control groups on the 10th week following challenge infection (Table 10). There was equally no statistically significant difference in the lengths of the worms obtained from experimental and control groups (Table 10).

2.6.2 Immunisation studies with synthetic peptides from the active site of *A. viteae* chitinase

To investigate whether antibodies towards the active site of *A. viteae* chitinase could inhibit chitinase and thus lead to protection against challenge infection, synthetic peptides were used in immunization studies. The synthetic peptides were designed from the active site of *A. viteae* chitinase. For the identification of relevant amino acids, the process of homology modeling was used. *Coccidiodes immitis* chitinase, a family 18 chitinase whose structure is known, was used as a template to predict the structure of *O. volvulus* chitinase. The corresponding amino acids were then obtained from *A. viteae* chitinase following alignment with *O. volvulus chitinase*. Homology modeling and synthesis of peptides were done respectively by Professor Höhne and Dr. Volkmer-Engert of the Charité Medical Faculty, Humboldt University of Berlin. Three candidate peptides were identified (see materials) and synthesized, and two turned out to be soluble. To investigate whether the two peptides could generate a specific immune response, they were coupled to a KLH carrier protein and 100 µg of coupled peptide/ KLH was used to immunize BALB/c mice with Alum as adjuvant. The sera of immunized mice could recognize full-length native *A. viteae* chitinase in western blots with *A. viteae* L3 total antigen (Figure 21).

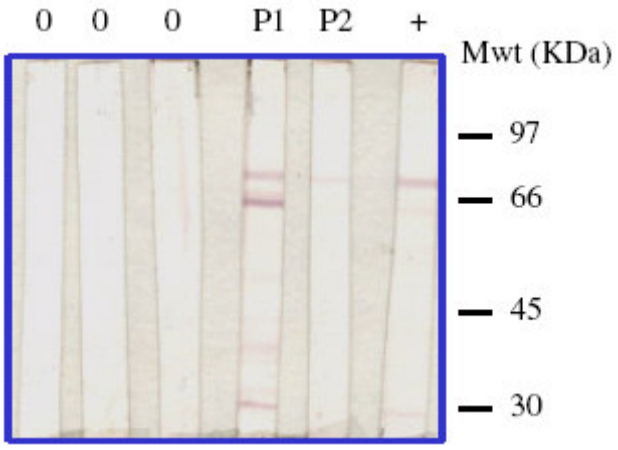


Fig. 21: Recognition of native *Acanthocheilonema viteae* L3 chitinase by sera of mice immunized with peptides.

Western blot reactions are shown for zero bleeds (0) and the reaction of antisera of peptides 1 and 2 to native chitinase. Positive control: Antisera to N-terminal chitinase protein.

The sera from animals immunized with P1 gave stronger reactions than sera from mice immunized with P2.The two synthetic peptides, P1 and P2, were then coupled with KLH and used for immunization. The coupled peptides (100 ug) were administered with alum as adjuvant at the flanks of the animals. Alum was used as adjuvant to shift the immune response towards the Th2 direction since the initial premise was to produce neutralizing antibodies to the active site of chitinase. Boosting, bleeding, challenge and section were done as in section 2.6.1.

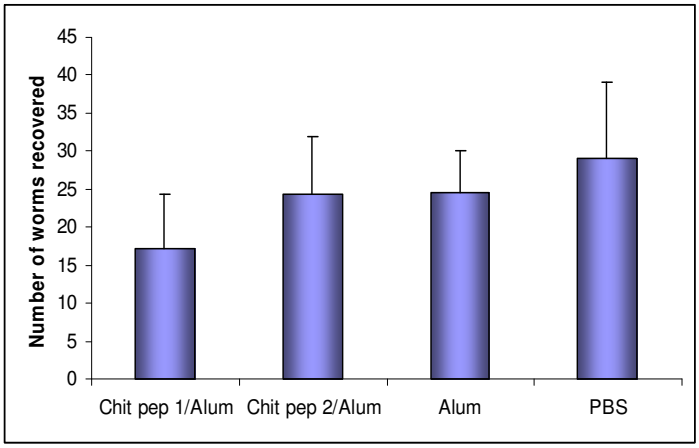
A summary of the biometric data obtained from both experiments is shown in Table 11.

Tab. 11: Summary of biometric data from immunization studies with synthetic peptides from *Acanthocheilonema viteae* chitinase

Experiment A				
Antigen/ control	Peptide 1/ Alum	Peptide 2/ Alum	Control alum	Control PBS
Worms recovered	17.2±7.1	24.3±7.6	24.5±5.6	29.5±10.1
Female worms (Length)	10.7±4.8 (4.99±0.15)	13.9±3.5 (5.16±0.46)	13.3±3.8 (5.08±0.21)	16.4±5.4 (5.25±0.65)
Male worms (Length)	6.5±2.2 (2.29±0.18)	10.4±4 (2.36±0.17)	11.2±2.8 (2.22±0.2)	13±5.2 (2.44±0.22)
Mf density / ml blood	274±146	574±215	439±54	456±152
Surviving animals	4/7	5/7	6/6	9/11

Experiment B			
Antigen/ control	Peptide 1/ Alum	Peptide 2/ Alum	Control PBS
Worms recovered	25±3.2	24.7±5.3	33.1±10.2
Female worms (Length)	15±4.2 (5.02±0.7)	11.8±4.6 (4.58±0.7)	19.1±4.8 (5.15±0.9)
Male worms (Length)	10.3±2.1 (2.32±0.3)	10±2.1 (2.31±0.3)	14.7±7.2 (2.38±0.4)
Mf density / ml blood	443.2±171.8	208.1±58.5	833±384.2
Surviving animals	6/6	6/6	9/11

A:



B:

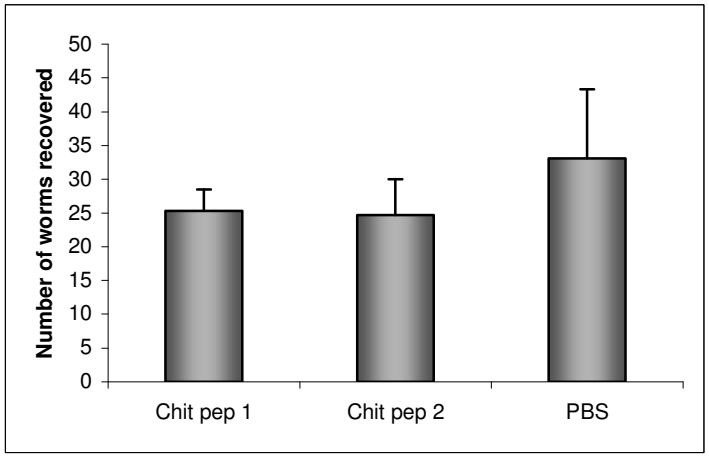


Fig. 22: Vaccination of *Meriones* with KLH-coupled chitinase peptides 1 and 2

The diagram shows the worm burdens following vaccination with 2 different chitinase peptides in two independent experiments, A and B.

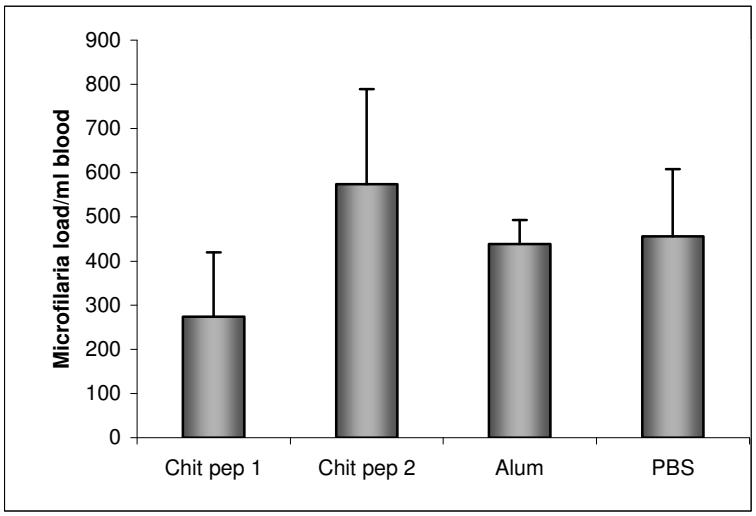
There was no significant difference in the lengths of the worms recovered from both experimental and control groups, indicating that immunization with the peptides may not have influenced the growth of the worms.

In two independent immunization experiments, data from animals vaccinated with peptide 1 suggested a trend towards decrease (33% and 23.4 %) in worm burden as compared to the control groups (Figure 22; Table 11). The first experiment was stopped at week 10 owing to death of animals in different groups (Table 11). Vaccination of *Meriones* with peptide 2 did not lead to an overall decrease in worm burdens in the first experiment, but showed a tendency towards decrease (25.5 %) in

worm burden in the second experiment (Figure 22). In the first experiment, the microfilaria load was significantly reduced ($p<0.05$) in the group of *Meriones* immunized with chitinase peptide 1 (Figure 23), while the chitinase peptide 2 group showed no tendency towards reduction.

In the second experiment, the microfilaria load was compared between experimental and control groups on the 12th week following challenge, and it could be shown that vaccination with chitinase peptide 1 and peptide 2 led to 45% and 75% significant reductions in microfilaria load ($p< 0.05$) , respectively (Figure 23).

A



B

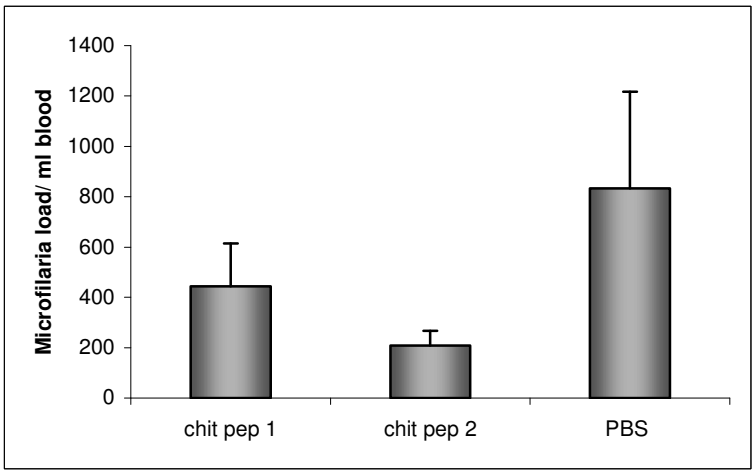


Fig. 23: Microfilaria burden from *Meriones* vaccinated with KLH-coupled chitinase peptides

The microfilaria burden following vaccination with chitinase peptides (chit pep) 1 and 2, as well as Alum and PBS controls are shown.

3 Discussion

3.1 Analysis of *A. viteae* chitinase genes and transcripts

One of the major goals of our study was the characterisation of *A. viteae* chitinase genes. This involves the description of the genomic structures as well as the characterisation of different transcripts of the individual stages of the life cycle of *A. viteae.*

According to our studies, the genome of *A. viteae* contains three different chitinase genes.

Southern blot analysis showed, particularly for the Pvu II digest, that there are at least two different chitinase genes. However, the different band intensities of especially the Pvu II digest could be explained by the fact that the probe bound at least twice more to close gene copies on a single DNA fragment. This indicates that there is a cluster with of a minimum of two chitinase genes. A similar suggestion was made by Arnold and colleagues (1996), who investigated gDNA from *B. malayi* for chitinase genes. The variation in the intensity of individual bands was also explained by the existence of a gene cluster. Further analysis of nine genomic sequences from a gene library of *A. viteae* worms confirmed the existence of three independent chitinase gene sequences. The sequences of two genomic clones (1 and 9) had a cluster of two genes as predicted from Southern hybridisation results, and an identical overlapping region with some minor differences. The corresponding genomic region of two inserts comprises a length of 6,845 bp, of which 93% was identical between two pairs of genes. The main differences arose from the beginning of insert 1, the end of insert 9, a repeating sequence of 25 nucleotides and a non-matching region in the intergenic region of both sequences. An additional argument for the identity of both sequences is the fact that the introns of both pairs of chitinase genes were 100% identical. This strongly indicates that the two pair of sequences are identical because intron sequences should show a higher mutation rate (Neafsey et al., 2005) since they are not under evolutionary pressure as exon sequences coding for particular amino acids responsible for the function of a protein. This hypothesis was confirmed by sequencing two independent clones (13 and 4) with similar restriction maps like clones 1 and 9. Analysis showed that the incomplete gene sequences at the beginning of clone 1 and at the end of clone 9 were truncated versions of gene sequences II and III, respectively. The discrepancies initially

observed could have been due to sequencing errors or to artefacts inherent in the construction of the genomic libraries.

In consequence, there were three independent genomic chitinase sequences that were very similar in the structure, and showed the highest degree of variation in exons for the serine threonine rich domain in the 3' part of the genes. A comparison with the structure of the *B. malayi* MF 1 chitinase gene showed that the whole structure of the genes is identical to the *A. viteae* genes, with the only difference being in the exons coding for the serine threonine rich domain. There are three other annotated chitinase genes in *B. malayi* having a broad range of organisations with some domains and exons fully, partially represented or absent (http://www.tigr.org). The absence of some exons in these genomic sequences could be seen as a footprint of unequal recombination occurring between exons of homologous and nonallelic chitinase genes (Maeda and Smithies, 1986). Interestingly, two of these chitinase genes are found in tandem on the same genomic assembly. Taken into account that a similar situation was demonstrated for *A. viteae*, it may be surmised that these genes exist in a cluster in these filarial parasites and are consequently a hot spot for recombination.

A. viteae chitinase gene sequence II, in contrast to sequences I and III, had no first exon with a start ATG and probably no functional transcript. A comparable sequence is found in the genome of *B. malayi*. Both sequences lack a first exon, and have an irregular second exon, a situation which may have arose by unequal crossing over resulting to the loss of the 5' exon with the start ATG (Maeda and Smithies, 1986). While *A. viteae* gene II has a regular stop codon and polyadenylation site, and may thus be considered as part of a true gene, the *B. malayi* sequence is definitely a pseudogene due to the presence of proretrovirus-like elements after the 10^{th} exon (Felder et al., 1994), and to an alteration of the reading frame.

The gene structure for the chitinase gene of the free living nematode *C. elegans* has a different distribution of exons coding for the enzymatic domain and for the serine-threonine rich linker structure. An unusually long exon codes for the glycosyl hydrolase domain responsible for the catalytic activity of the deduced protein, and for most of the serine threonine rich linker region. The exons coding for the secretory signal sequence and the chitin binding domain correspond to those in the filariae.

The *C. elegans* chitinase is an orthologue of the filarial chitinases, despite the lack of similar introns and conservation of all intron positions and phases. This is congruent with the clustering of all of these sequences in a phylogenetic tree, and to the fact that no other *C. elegans* chitinase has a better match to filarial chitinases than this *C. elegans* chitinase. A similar observation was made by comparing an 83 kb synthenic region between *B. malayi* and *C. elegans* and demonstrating that corresponding orthologous genes clearly had differences in gene structure (Guiliano et al., 2002). Apart of this chitinase, there are some other 30 expressed chitinase genes in *C. elegans,* whose implication and function have not been addressed (www.wormbase.org; Popovici et al., 1999).

It is not known why *C. elegans* has many more chitinase genes than the filarial nematodes. The *C. elegans* chitinase gene family is made up of many orthologues and homologues that may have arisen by duplication and diversion of a common ancestor gene (Popovici et al., 1999). Though the function of all the members of this gene family is not yet known, diversion may have led to new copies with different functions as well as tissue or substrate specificities. Some of the genes are very identical with up to 10 nonallelic copies clustered on one chromosome, while others have very low identities and are widely dispersed on different chromosomes. A general conclusion drawn from the *C. elegans* chitinases would be that there may be more members of a gene family than one could suspect on the basis of protein analysis and by the use of probes to fish out identical genes of the same family. This is supported by the fact that many of the chitinase genes share common motifs, but have very low sequence identities. In line with this hypothesis, there may be several other chitinase genes in the filarial nematodes which have not been characterised based on the low identities they share with known family members. Support for this hypothesis comes from the *B. malayi* irradiated L3 chitinase EST that has a bare 32 % identity to the published sequence for *B. malayi* chitinase. Another observation showing that such a phenomenon is present in the filarial genome is the description of two chitin synthases in *Brugia* with a bare 27% identity (Harris et al., 2000; Foster et al., 2005). The existence of orthologous genes with the same function in a parasite makes it difficult to effectively use these genes as vaccine or drug targets. Thus, the further characterisation of filarial chitinases as drug or vaccine targets will entail the identification and/or exclusion of further functional chitinase genes in the nematode of interest.

In keeping with the identification of three chitinase genes in *A. viteae* it was hypothesised that there could be several stage-specific *A. viteae* chitinase genes which may have different substrate specificities or functions. There are two published *A. viteae* chitinase cDNA sequences (Adam et al., and Wu et al., 1996) that differ only in the outmost 5' ends. The Wu et al., sequence begins with a start methionine and three additional amino acids which are absent in the Adam et al., sequence. Within the frame work of this study, data from cDNA and genomic sequences could confirm that the Wu et al., sequence is the variant that is expressed *in vivo*. The exons of *A. viteae* genomic gene I are 100% identical to the sequence of Wu and colleagues, particularly with regards to the 5' end. Transcripts could actually be found for this sequence in uterine microfilaria, blood microfilariae, L3 and L4, while no transcripts could be found for gene sequence III. It would however not be excluded that other *A. viteae* chitinase-like transcripts could be present in *A. viteae*, since western blot data show the existence of four molecules reacting with antibodies to *A. viteae* chitinase. It was shown that sera of *Meriones* protected by vaccination with irradiated L3 were mainly directed against L3 molecules of 205 kDa, 68 kDa and 17 KDa (Lucius et al., 1991). The 205 and 68 KDa molecules were characterised as chitinases (Adam et al., 1996). In addition, monoclonal antibodies produced from sera of animals immunised with irradiated attenuated L3 and Excretory secretory products (ESP) could recognise an array of molecules in different filarial stages in immunoblots (Adam et al., 1996). Amongst others, a 68 kDa chitinase could be shown for L3 and L4, while 220 kDa, 205 kDa and 140 kDa molecules were found in female worms, L3 and microfilariae, respectively. The authors proposed that the 205 kDa L3 molecule could be as a result of oligomers formed from the 68 kDa monomers, or post-synthetic modification of the monomer. The antigens from female worms were shown to be from uterine microfilariae (Adam et al., 1996). The implication of such a finding could be that there are possibly different stage-specific chitinases in microfilariae and L3 of *A. viteae*. While we have demonstrated that the same chitinase molecule (68 kDa) is expressed in mf, L3 and L4, it remains to be shown that there are other chitinase messages in different parasite stages.

Filarial chitinases had been exclusively attributed to the L3 and mf stages of *A. viteae* and *O. volvulus* (Wu et al., 1996; Adam et al., 1996) or to the microfilarial stages of *Brugia* (Fuhrman et al., 1992; Arnold et al., 1996). However, *Brugia* EST data point to

the existence of a novel chitinase transcript in *B. malayi* L3, a stage from which no chitinase gene had previously been characterised. In this study, we showed chitinase transcripts in uterine microfilaria and L3 and L4, but not in the cDNA library from adult female worms, even though the uterine microfilariae were derived from gravid female worms. A similar difficulty was found in the isolation of *B. malayi* chitinase from microfilaria (Southworth et al., 1996) and *O. volvulus* chitinase (Perler, personal communication) from a female cDNA library. In both cases, the reasons advanced were the low representation of chitinase transcripts in the total transcript population of the filarial stages looked into, or suboptimal isolation of targeted transcripts from these filarial stages (Southworth et al., 1996; Perler, personal communication).

A comparison of the deduced amino acid sequences from the three *A. viteae* chitinase gene sequences revealed a very high rate of identity except for the serine threonine rich linker region. The variation in the serine threonine rich linker region of *A. viteae* chitinase is also found when different filarial chitinase sequences are compared. It is not known whether the size and composition of the serine threonine rich linker are of particular relevance to the function of the protein. However, this region has several consensus sequences for glycosylation and when glycosylated will influence protein solubility, secretion and the resistance of the protein against proteases as has been demonstrated for insect chitinases (Arakane et al., 2003). Such properties could be of particular relevance to the host-specificity of parasites and their chitinase molecules.

The function of chitinases in filarial worms is not yet known. However, in protozoan parasites like *Plasmodium* and *Leishmania*, chitinase has been shown to be important for transmission by their role in lysis of the arthropod peritrophic matrix (Langer and Vinetz, 2001; Romalho-Ortigao and Traub-Cseko, 2003). While *Simulium* vectors of *O. volvulus* have been reported to have peritrophic membranes (Ramos et al., 1994; Demanou et al., 2003; Soumbey-Alley et al., 2004), such membranes have not been demonstrated in tick vectors of *A. viteae* (Sonenshine, 1991). It is therefore not known whether similar mechanisms operate in these vectors. However, L3 chitinases of *O. volvulus* and *A. viteae* could have a role in the

egress of these parasites from their arthropod vectors since their expression coincides with the transmission of the parasites (Wu et al., 1996).

If chitin is the unique substrate for filarial chitinases, it may be speculated that chitinases play a role in parasite transmission based on two observations. Firstly, *Brugia* microfilariae express chitinase; secondly, it has been shown that *Brugia* microfilariae not expressing chitinase could not penetrate the arthropod vector (Fuhrman et al., 1987). It is hypothesised that lectins or agglutinins in the arthropod midgut binds to the microfilariae and reduces their infectivity. Chitinase promotes infectivity by digesting its substrate and releasing N-acetylglucosamine (GlcNAc) that saturates the lectins and reduces their interaction with microfilariae. Such a hypothesis has been proven in other parasite systems by supplementing the infection blood meal with N-acetylglucosamine, the monomers of chitin, which in turn increases infectivity (Welburn et al., 1993). Such a hypothesis can be tested in the filariae by carrying out similar experiments.

Parallel to the role of insect chitinases in ecdysis (Kramer and Muthukrishnan, 1997), filarial chitinases are speculated to be involved in moulting and hatching, since filarial eggs and adult sheath have been shown to contain chitin (Harris et al., 2000; Brydon et al., 1987). Such a role in ecdysis was suspected for *A. viteae* chitinase by the expression of chitinase on the surface of microfilaria, just before they hatch from their modified eggshell, and the absence of chitinase in blood-borne microfilaria (Adam et al., 1996).

3.2 Expression, purification and activity of N-terminal chitinase

There are three possibilities to control filarial infections: vector control, chemotherapeutic control, and immune control through vaccines. Chitinase is an attractive target for vaccine and drug control, since it is an enzyme that is expressed in two key stages (microfilariae and L3s) in the life cycle of filariae. In addition, knock out experiments with the *C. elegans* chitinase orthologue show that this gene is important for the development of larvae that are the equivalent of microfilariae in the filarial nematodes (www.wormbase.org). Moreover, knowledge on the structure of filarial chitinases will help towards the development of selective inhibitors against these molecules. Chitinase inhibitors might at the same time also target other chitin-containing pathogens like fungi and arthropods. For these reasons, it was necessary to produce large amounts of soluble and active *A. viteae* chitinase.

A bacterial expression system was chosen for the expression of *A. viteae* chitinase. The chitinase gene was cloned into the pET 22 b (+) vector (Novagen) and could be expressed in a suitable *E. coli* host. Expression and purification of large amounts of soluble full-length molecule was not possible, and so the N-terminal portion corresponding to the glycosyl hydrolase domain of the molecule was expressed. Failure to obtain a full-length molecule could have been due to difficulties in disulfide bond in the substrate-binding domain of the molecule that contains 6 cysteins. Amongst other things, the expression of soluble active chitinases from filarial paraistes (Southworth et al., 1996; Drabner et al., 2001) and other parasites like *Leishmania* (Razek-Desoukey et al., 2001) has always been difficult due to the 6 cysteins in the chitin binding domain, which disturbs proper folding. To overcome these problems of expressing soluble proteins in heterologous genes in *E. coli*, different strategies are used like reduction in induction temperature (Schein and Noteborn, 1989; Razek-Desoukey, 2001), using different bacteria expression hosts (Vinetz et al., 1999), fusion with a solubility-enhancer tag (Davis et al., 1998), export of the protein into the periplasmatic space (Southworth et al., 1996), and expression of truncated forms or independent domains of the molecule (Southworth et al., 1996).

The glycosyl hydrolase domain of chitinases folds independently of the substrate-binding domain (Synstad et al., 2004; Fusetti et al., 2003), and both domains are independently stable and active (Henrissat and Bairoch, 1993, 1996; Arakane et al., 2003) (http://afmb.cnrs-mrsfr/CAZY)). In this perspective, studies on chitinase activity and inhibitor binding could be made with the N-terminal glycosyl hydrolase domain of the molecule. For this reason, the N-terminal glycosyl hydrolase domain of *A. viteae* chitinase was cloned into the vector pET 22 b (+) and expressed in BL 21 (DE3) *E. coli* cells. Most of the protein was insoluble and could be purified only under denaturing conditions. There was a marginal increase in protein solubility when expression was done at lower temperatures. Export into the periplasmatic space produced little amounts of soluble material, which was 16 times more active compared to the cytoplasmic protein. A comparable 12-fold increase in the activity of a *B. malayi* MBP fusion chitinase was obtained upon export of the protein into the periplasmatic space (Southworth et al., 1996; Venegas et al., 1996). The increased specific activity of chitinase exported into the periplasma could be attributed to the

better folding conditions in the periplasm as opposed to the cytoplasm (Rietsch et al., 1996; Raina and Missiakas, 1997; Sone et al., 1997).

Since most of the chitinase was expressed as insoluble material in the cytoplasm, bacteria hosts that enhance protein solubility in the cytoplasm were used for expression. Bacteria strains like Origami B (DE3) carry the *trxB* and *gor* mutations that enhance the formation of disulfide bonds in the cytoplasm (Derman et al., 1993; Aslund et al., 1999) and increase protein solubility. The use of this strain led to a 4-fold increase in chitinase solubility and a 13-fold increase in activity as compared to chitinase purified under denaturing conditions.

In order to obtain much more active *A. viteae* chitinase, current efforts are directed towards a eukaryotic expression system like insect cells. The reasons for this are two-fold. Firstly, glycosylation may lead to an increase in filarial chitinase activity (Fuhrman et al., 1992; 1995). Secondly, a commercially available eukaryotically expressed *O. volvulus* chitinase was shown to be 4-fold active compared to the *A. viteae* chitinase in this study. The eukaryotic expression system offers better folding conditions and the possibility of post-synthetic modifications, which are absent in the prokaryotic expression system (Vialard et al., 1995).

3.3 Immunological control

Several independent observations and experiments hint to the development of protective immunity against filariae. In natural endemic populations, there are putatively immune (PIs) individuals (Elson et al., 1994; Turaga et al., 2000) with no current or past evidence of filarial infections. In natural and surrogate model systems for filariasis, immunisation with irradiated infective stage larvae led to near-sterile immunity (Lucius et al., 1991; Prince et al., 1992; Lange et al., 1993; Taylor et al., 1994; Johnson et al., 1998; Trees et al., 2000). However, the expression of identified filarial vaccine candidates as *E. coli* expressed recombinant proteins or as DNA vaccines in various models led to varying rates of protection that did not match up to the almost complete protection rates achieved by the use of irradiated third-stage larvae (Lustigman et al., 2002; Cook et al., 2001; Abraham et al., 2002). This failure to mimic vaccination with attenuated parasites can be due amongst others to the absence of non-protein moieties, like in phosphorylcholine and carbohydrates in the recombinant vaccines (Nutman, 2002). In addition, native-protective antigens may have a different tertiary structure and thus a greater capacity to induce protective

immunity as opposed to recombinant proteins. This is consistent with observations in the cestodes for which the initial use of native vaccine antigens has led to the development of successful vaccines (Lightowlers et al., 2003). The paucity of starting material however limits such an approach in the filarial nematodes. Thus the use of heterologous nematode protein expression systems may lead to the faithful expression of native proteins.

An understanding of the protective mechanisms in PIs and the irradiated L3 vaccination models will help in targeted development of potential vaccines. The protective mechanisms in PIs is not clearly associated with any of the arms of immunity: antibody responses (Stewart et al., 1995; Boyer et al., 1991) as well as Th1 and Th2 responses (Turaga et al., 2000) have been described. On the other hand, protective immunity associated with vaccination using irradiated larvae was clearly dependent on Th2 responses and involved antibodies and eosinophils (Le Goff et al., 2000; Lange et al., 1994; Bleiss et al., 2002). A role for antibodies in natural systems is provided by the elevated levels of IgG1/ IgG3 in protected subjects (Stewart et al., 1995) and the demonstration that sera from both infected and protected subjects mediate adherence and killing of microfilariae (Brattig et al., 1991) and infective larvae (Johnson et al., 1994).

Filarial chitinases could be considered as targets for immune attack based on two observations: the sera of PIs differentially recognise filarial chitinase molecules as opposed to clinically infected individuals (Rhagavan et al., 1994); and the sera of rodents immunised with irradiated *A. viteae* L3s dominantly recognised L3 chitinase in blots (Adam et al., 1996). Moreover, a monoclonal antibody, that mediated microfilarial clearance upon transfer, recognised *B. malayi* microfilaria chitinases in western blot (Canlas et al., 1984a; Canlas et al., 1984b).

In addition to these observations, confirmatory vaccination studies were done with filarial chitinases from *O. volvulus* leading to a 53% significant reduction in worm burden (Harrison et al., 1999).

It may therefore be surmised that eliciting the right antibody response against chitinase would contribute to immune protection.

The catalytic and substrate binding domains of chitinase have been shown to have different immunological properties (Venegas et al., 1996; Wu et al., 2001). We hypothesized that an increased antibody response to the glycosyl hydrolase domain and/or the active site of *A. viteae* chitinase will lead to the selective inhibition of chitinase activity and function, and an inhibition in the development of stages associated with this enzyme. The use of antibodies to epitopes or peptides from active sites of enzymes has been shown to be a successful tool for vaccination in cases where the enzyme appears at critical points in the life cycle of a parasite. Passive transfer of monoclonal antibodies binding to the active site of glutathione S-transferase of the nematode *Schistosoma bovis* impaired egg development, and this property was in turn correlatated with the enzyme inhibition (Da Costa et al., 1999). Using a similar strategy, it has been shown that vaccination of rabbits with synthetic peptides from the active site of *Helicobacter pylori* urease produced neutralising antibodies, and was a useful vaccination tool (Hirota et al., 2001). Synthetic peptides can be obtained from active sites of enzymes by homology modelling using other well characterised enzymes (Fujii et al., 2004).

The N-terminal catalytic domain of chitinase, as well as synthetic peptides designed from the active site of *A. viteae* chitinase 3-D model, was used in immunisation studies. Immunisation with the whole catalytic domain with STP as adjuvant led to a significant reduction in worm burden in a unique experiment. When alum was used as adjuvant, there was no tendency towards reduction in worm burden.

Despite the fact that STP was shown to be the better adjuvant for immunisation, alum was used for immunisation studies involving synthetic peptides, since our premise was to induce neutralising antibodies to the active site of chitinase. Alum is an adjuvant that directs the immune resonse to the Th2 direction /humoral arm. Immunisation with the synthetic peptides showed a tendency towards an overall decrease in worm burden in two independent experiments. In these same experiments, the group vaccinated with peptide 1 had a consistent significant reduction in microfilaria load, while the group vaccinated with peptide 2 had a significant reduction in microfilaria load in just one experiment.

Immunisation experiments were done just once with recombinant protein, and would therefore have to be repeated to confirm the trend observed. On the other hand, two independent vaccination studies were done using the synthetic peptides. The trend towards reduction in worm / mf burden was inconsistent when both peptides were

used. It can therefore not be definitely concluded that vaccination with either synthetic peptides or recombinant protein conferred protection to challenge infection. Inconsistencies in the protective potential of highly immunogenic filarial proteins have been observed before (Peralta et al., 1999). In their study, Peralta and colleagues used differential screening with sera of a diverse group of filariasis patients to isolate highly immunogenic proteins from a *B. malayi* expression library. However, independent vaccination experiments with the same proteins produced inconsistent protection results (Peralta et al., 1999).

Vaccination with peptide 1 could consistently reduce the microfilaria burden. Vaccination with this peptide may interfere with hatching of eggs to microfilaria or the development and moulting of microfilaria and L3. Such a hypothesis will be consistent with the finding that chitinase is present on the cuticle of post-infective L3 of unsheathed filarial (Harrison et al., 1999) and on the surface of uterine microfilaria (Adam et al., 1996). If chitinase has a role in the the development of filarial worms, any process that interferes with enzyme activity could also distort worm development. Such a case could be expected by the production of antibodies to the active site of the enzyme by vaccination with synthetic peptides. Thus, these molecules may therefore have a role in the development of uterine microfilariae and post-infective L3 which is distorted by vaccination with chitinase. The possible mechanism for reduction in microfilaria burden is not clear, and should be understood by demonstrating the role of antibodies in the observed tendency towards reduction in worm burden.

3.4 Outlook

The prospects for the future are three-fold: determination of the biological function of *A. viteae* chitinase, and its possible influence on host physiology; development of inhibitors to worm chitinases, and use of different immunisation approaches.

The function of *A. viteae* chitinase is not well defined. Therefore, future studies would be aimed at determining the biological function and possible influence on the host physiology. Reverse genetic studies like RNAi would be used to clarify the function of this gene in *A. viteae*. Three chitinase genes were described in this study. Analysis demonstrated transcripts in microfilariae, L3 and L4 larvae of *A. viteae* for gene I only, but not for gene III. To verify whether functional transcripts could actually be

produced from gene III, an *in vitro* transcription / translation system would have to be used. In addition, other *A. viteae* stages will have to be screened for chitinase genes.

Soluble active *A. viteae* is required for the development of inhibitors to worm chitinases, and for crystallisation studies. The expression and purification of soluble active chitinase in prokaryotic systems was difficult since most of the protein was found in inclusion bodies. The expression of *A. viteae* chitinase in eukaryotic systems like yeast or insect cells would provide better alternatives. Moreover, the proteins will be post-synthetically modified in these systems. A comparison of the modification pattern with that of native *A. viteae* chitinase will illustrate the similarities in post-synthetic modification.

Vaccination with prokaryotically expressed *A. viteae* chitinase did not lead to significant reductions in worm burden. Since, native proteins have been shown to be better vaccines in cestodes (Lightowlers et al., 2003), *A. viteae* chitinase expressed in eukaryotic systems would be tested in immunisation experiments along side native and prokaryotically expressed *A. viteae* chitinase. This will give a clue as to whether the modified proteins could have better protective capacities as compared to non-modified proteins.

4 METHODS

4.1 Identification of recombinant cDNA and genomic clones from phage libraries by plaque hybridisation

4.1.1 Preparation of plating bacteria

Host bacteria for bacteriophages (XL 1 blue for lambda zap, and XL 1 blue MRA for lambda dash II) were grown overnight at 30°C, 150 rpm in 100 ml LB containing 0.2% maltose. Cells were pelleted at 5000 rpm for 10 minutes at 4°C and re-suspended in 20 ml 10 mM $MgSO_4$ to an OD_{600} of 2 (about $1.6x10^9$ cells per ml). The cell suspension was incubated at 37°C, 220 rpm for 1 hr, stored at 4°C and used for a maximum of three days.

4.1.2 Plating of lambda phage

An appropriate volume of phage suspension was mixed with 100 µl (for 88 mm Petri dishes) or 200 µl plating bacteria (for 120 mm petri dishes) and incubated with shaking (100 rpm) for 15 minutes at 37℃. The suspension was then mixed with 4 or 8 ml of molten Top agar (50℃) and carefully spread on agar plates. The plates were incubated for 30 minutes at room temperature to allow the top agar to harden, and were then incubated at 37℃ overnight. A uniform bacteria lawn could be observed interspersed with lambda plaques. The plates could be stored at 4℃ for several days.

4.1.3 Titration of phage libraries

To determine the concentration of a phage suspension in Plaque Forming Units (pfu) per ml, dilutions of a phage stock were made from 1:10 to 1:10, 000 in SM buffer, and 1 µl of each dilution was plated as described in section 4.1.2. Plates having between 50 to 350 plaque forming units were used to calculate the phage concentration in pfu.

4.1.4 Amplification and cryopreservation of phage stocks

Confluent plates of 3 x 10^4 pfu/plate were produced and incubated for a maximum of 8 hours at 37℃. SM buffer was then added to the plates and the plates were incubated with constant shaking for 2 hours at 4℃. The lysate was then aspirated and centrifuged for 5 minutes at 10, 000 rpm to pellet residual agarose particles and the titre was determined as in section 4.1.3. DMSO was then added to 1 ml aliquots of bacteriophage stock at a concentration of 7% v/v. The tube was mixed gently, plunged into liquid nitrogen and stored at -80℃ until used.

4.2 Screening of libraries

4.2.1 Screening of libraries with a DIG-labelled *A. viteae* chitinase probe

4.2.1.1 Screening of libraries by plaque hybridisation

Up to 2 x 10^4 independent plaques were plated on a 120 mm petri dish and cultured for 8 to 12 hours. The plates were incubated at 4℃for 30 minutes to allow the top agar to set. Nylon membranes were rinsed in water and then in 5x SSC and allowed to dry at room temperature. The membranes were carefully placed on the surface of the infected plate so that the entire membrane was in contact with the plate. After 30 minutes of incubation at 4℃, the membranes were marked with a hot needle for orientation. Plates were again incubated at 4℃ for 30 minutes, after which the

membranes were removed. Replica filters were also used to lift plaques from infected plates.

DNA on the filters was denatured by incubation with 0.5 M NaOH/ 1.5 M NaCl for 15 minutes at room temperature, and subsequently neutralised twice by incubation with 0.5 M Tris, 1.5 M NaCl pH 8.0, after which the filters were allowed to dry on a Whatmann filter paper. DNA on the filters was immobilised by covalently linking the nucleic acids to the membrane by UV cross-linking. Alternatively, the membranes were baked at 80°C for 2 hours. An N-terminal cDNA sequence corresponding to nucleotides 52-1,117 of *A. viteae* L3 chitinase (Adam et al., 1996; Genbank number U14638) was labelled with digoxigenin (DIG) using the PCR labelling and detection kit from Roche. Hybridisation and detection was done as described by the manufacturer (Boehringer, Mannheim). Hybridisation was done with 15 ng/ml of probe in 10 ml of hybridisation buffer at a temperature of 50°C overnight. Stringent washes were done as follows: membranes were washed for 2 x 15 minutes in ample 2x SSC, 0.1% SDS at RT, and then in 0.5x SSC, 0.1% SDS (pre-warmed to wash temperature) for 2x 30 minutes at 65°C under constant agitation. Chemiluminescent detection was done using CSPD (Boehringer Mannheim).

Agarose plugs containing putatively positive plaques were incubated in SM buffer overnight at 4°C to let the phage wander out. The phage suspensions were diluted 1:10 with H_2O and boiled for 5 mins, and 1 µl was used for PCR as in section 4.3.12.1. Primers were vector-specific (T3 or T7) and *A. viteae* chitinase-specific.

To obtain clonal recombinant phages, phage suspensions identified as positive from the primary and PCR screening rounds were subjected to three further rounds of plaque hybridisation. In order to increase the chances of picking single/ independent plaques, 100-fold less independent plaques were used per new screening round as the proportion of positive plaques increased.

4.2.1.2 In vivo *excision*

The cDNA libraries used in this study were constructed in the vector Lambda UniZap XR (Stratagene). This vector is designed to allow *in vivo* excision and recircularization to form a pBluescript phagemid containing the cloned insert. An ExAssist helper phage contains a mutation that prevents replication of the phage genome in the non suppressing *E. coli* host (SOLR) and thus allows only the excised phagemid to replicate (Short et al., 1988). *In vivo* excision was done according to the

manufacturer's instructions. Briefly, 5-ml cultures of XL 1 Blue MRF´ and SOLR cells were grown at 30°C overnight in LB broth. The cells were pelleted and resuspended in 2.5 ml 10 mM $MgSO_4$ to an OD_{600} of 1.0 ($8x10^8$ cells/ml). An excision mix was made of 200 µl of XL 1 bLue MRF´ cells, 250 µl phage stock ($1x10^5$ phage particles) and 1 µl ExAssist helper phage (titer: 1 $x10^7$ pfu/ml). The mixture was incubated for 3 hours at 37°C with shaking. During this period the excision reaction takes place leading to the formation of pBluescript phagemid packaged as filamentous particles. The tube was heated at 70°C for 20 minutes to lyse the lambda phage particles and cells. Following centrifugation (1000xg for 15 minutes), the supernatant containing the excised phage particles was then stored. To plate the cells, 100 ul of phage supernatant was added to 200 µl SOLR cells (OD_{600}=1) and incubated at 37°C for 15 minutes. The cells were then plated on LB-ampicillin plates and incubated overnight at 37°C. Bacteria colonies were isolated containing the pBluescript double-stranded phagemid with the cloned cDNA insert.

4.3 Purification and manipulation of DNA

4.3.1 Isolation of lambda DNA and sub-cloning of genomic inserts

DNA was typically purified from clonal plaques and used for restriction digestion analysis, PCR, Southern and subcloning. The genomic libraries used in this study were constructed in lambda dash II (Stratagene). This vector, unlike lambda UniZap, can not be *in vivo* excised. To circumvent this short-coming, the inserts were released by digestion and cloned into plasmids.

4.3.1.1 Preparation of crude phage lysates

Confluent plates of 3 x 10^4 pfu/plate were produced and incubated for a maximum of 8 hours at 37°C. SM buffer (10 ml) was then added to the plates and the plates were incubated with constant shaking for 2 hours at 4°C. The lysate was aspirated and centrifuged for 5 minutes at 10, 000 rpm to pellet agarose particles.

4.3.1.2 Isolation of phage DNA

DEAE cellulose columns were prepared and used for isolation of DNA. To prepare these columns, hydrochloric acid (0.05N) was slowly added to DE52 powder (Whartmann DE52 number 4057-050) with continuous stirring until a pH < 4.5 was attained. Concentrated NaOH was then added until the pH was 6.8. The resin was let to settle, the supernatant was aspirated and the ion exchanger was then washed 4 to

5 times with several volumes of LB-Medium. The slurry was adjusted to 60% resin, NaN_3 was added to a final concentration of 0.02% and the slurry was stored at 4°C for up to three months. Columns were prepared by pouring 8 ml DEAE slurry in a 10-ml disposable syringe resulting in a bed height of 6 cm. After the DEAE-cellulose had settled, the column was washed with 5 ml LB and subsequently equilibrated in LB.

The supernatant from two plates was loaded onto a 6-8 ml DEAE column. During this chromatographic step, bacterial DNA and RNA are separated through ion-exchange chromatography; the ion-exchange resin preferentially binds contaminants in the lysate such as *E. coli* DNA, RNA and proteins. The eluate containing the phage particles were then collected and pooled. The column was washed again with 2 ml LB- Medium and the eluate added to the first. The NaCl concentration of the eluate was adjusted to 70 mM and phages were precipitated with 2 volumes of ethanol (20 minutes at –20°C, and then centrifuged for 15 minutes at 12,000 rpm). The pellet was then washed with 2 volumes of 70% ethanol and subsequently dissolved in 2 ml TE with 0.2% SDS to release the phage DNA. The mixture was extracted twice with 2 volumes of Phenol (TE saturated). DNA was precipitated from the aqueous phase with 2 volumes of ethanol without any salt. The phage DNA pellet was then washed twice with 70% ethanol and dissolved in 200 µl TlowE. The yield was about 75 µg per 10 ml phage lysate.

4.3.1.3 Release and sub-cloning of genomic inserts into pBluescript II SK

Purified recombinant lambda DNA and pBluescript II SK (-) were separately digested with Not I and dephosphorylated as described in sections 4.3.8 and 4.3.9. Genomic inserts ranging in size between 10-14 kb were then ligated into pBluescript vector (3kb) using a 1:1 molar ratio of insert and vector (see section 4.3.10.2).

4.3.2 Small scale Plasmid isolation from *E. coli*

Plasmid DNA was isolated as described by Del Sal et al., 1989, with some slight modifications. This technique relies on the property of the detergent cetyl-trimethyl ammonium bromide (CTAB) of preferentially precipitating nucleic acids at low NaCl concentrations (< 0.6 M), while leaving proteins and polysaccharides in solution.

Essentially, a single bacterial colony harbouring a plasmid of interest was used to inoculate 6 ml of LB medium (with an appropriate antibiotic for selection) and grown over night at 37°C. Bacterial cells were recovered by centrifugation at 4,000 rpm for

10 minutes. The pellet was re-suspended in 200µl of STET buffer, 4µl of a 50-mg/ml lysozyme solution was added and the samples were incubated at room temperature for 5 minutes. Samples were boiled for 45 seconds and then centrifuged at 13,000 rpm for 10 minutes in an Eppendorf centrifuge. Pellets were recovered, 5 µl of RNase A (10 mg/ml) were added to the supernatants and the samples were incubated for 10 minutes at 68°C. Nucleic acids were precipitated by the addition of 10 µl of CTAB (5% w/v), followed by centrifugation at 14,000 rpm for 5 minutes. Pellets were re-suspended in 300 µl of 1.2 M NaCl and the plasmids were re-precipitated by the addition of 750 µl ethanol and incubation at - 20°C for 1 hour. Samples were then centrifuged at 14,000 rpm for 15 minutes and the pellets were washed twice in 70% ethanol and allowed to air dry. Plasmids were re-suspended in 40 µl H2O. The yield was between 20 to 40 µg of plasmid DNA.

4.3.3 Electrophoresis and detection of DNA on agarose gels

Agarose gel electrophoresis was used for the routine analysis of DNA. Agarose was cast in 1x TAE buffer containing 0.5 µg/ml EtBr. The DNA samples (0.1 to 5 µg) were dissolved in DNA loading buffer and separated in agarose gels at 10 volts per cm. The concentration of the agarose gel was relative to the size of DNA fragments to be separated, and was typically between 0.7 and 1.2%.

4.3.4 Isolation of DNA from agarose gels

Following electrophoresis, a DNA fragment was excised from agarose gels using a clean scalpel, and the DNA was purified using the NucleoSpin® Extraction Kit (Clontech) according to the manufacturer's instruction.

4.3.5 Isolation and concentration of DNA from aqueous solutions

4.3.5.1 Extraction with Nucleospin kit

The NucleoSpin DNA Extraction Kit (Clontech) was used according to the manufacturer’s instruction for the isolation of DNA from PCR reactions, gel fragments and other aqueous DNA solutions.

4.3.5.2 Phenol chloroform extraction

To free DNA from protein contaminants, the volume was adjusted to at least 300 µl with TE buffer, and extracted with an equal volume of Tris-equilibrated phenol:chloroform:isoamyl alcohol (25:24:1) (Sambrook et al., 1989). The aqueous

phase was extracted twice with chloroform to remove traces of phenol, and the DNA was concentrated by ethanol precipitation.

4.3.5.3 Precipitation of DNA

Ice cold sodium acetate, pH 4.5 was added to a DNA sample to a final concentration of 0.3 M, followed by the addition of 2.5 volumes of ice cold absolute ethanol. The sample was incubated at -70°C for 15 minutes and centrifuged at 12, 000 rpm for 15 minutes at 4°C. The pellet was washed in 70% ethanol, air-dried, and re-suspended in an appropriate buffer.

4.3.6 Isolation of high molecular weight genomic DNA from *A. viteae*

About 200 mg to 1 g of adult *A. viteae* were snap-frozen in liquid nitrogen and ground to powder using a mortar and a pestle. Up to 100 mg of powdered worm material was suspended in 1.2 ml of digestion buffer containing Proteinase K (100 µg/ml), and incubated with shaking at 50°C for 12 to 18 hours until the sample became viscous with a visible sludge. The samples were extracted three times with an equal volume of phenol/chloroform/isoamyl alcohol. The viscous aqueous phase was transferred each time with a wide-bore pipette (0.3-cm-dimeter orifice) into in to a new tube. RNA was digested by the addition of 20 µg/ml RNase A and incubating for 1 hour at 37° C. Genomic DNA was then precipitated by adding ½ volume of 7.5 M ammonium acetate and 2 (original) volumes of 100% ice-cold ethanol. Stringy precipitates of DNA were carefully transferred in to new tubes using a blunt spatula, and washed twice with ample amounts of 70% ethanol (2 times the original volume). The DNA pellet was air dried and re-suspended in TE buffer until dissolved. The concentration of the gDNA was determined by photometry and the DNA was analysed on a gel.

4.3.7 Determination of DNA Concentration

The concentration of DNA in solution was determined using a spectrophotometer at OD_{260}, and subsequently confirmed by comparing the intensity and size of DNA from a sample to those of standard size markers with known concentrations.

1 unit of absorbance of dsDNA at OD_{260} = 50 µg/ml dsDNA

1 unit of absorbance of ssDNA at OD_{260} = 33 µg/ml ssDNA or RNA

The OD_{280} was also measured and the ratio between the two ODs indicated the purity of the DNA solution. For pure DNA,

$$1.8 \leq \text{absorbance at } OD_{260} / \text{absorbance at } OD_{280} \leq 2.0$$

A value less than 1.8 indicates contamination with proteins or with aromatic substances like phenol, while a value greater than 2.0 indicates possible contamination with RNA.

4.3.8 Restriction digestion of DNA

DNA was digested at the optimal temperature of restriction enzymes according to the pipetting scheme below.

Tab. 12: Pipetting scheme for restriction digestion of DNA

Reagent	Volume (µl)
DNA (0.1 – 5 µg)	5
10 x restriction enzyme buffer	2
10 x Bovine Serum Albumin (BSA)	2
Restriction enzyme (3-20 units)	1
HPLC water	10
Volume	20

Following digestion, the mixture was either heated at 70°C for 15 minutes to inactivate the enzyme, or extracted with phenol chloroform.

4.3.9 5' Dephosphorylation of digested DNA

The removal of 5’ phosphate terminals from digested dDNAs ends prevents vector self-ligation and reduces background in ligations, especially when only one restriction site is used. Shrimp alkaline phosphatase (SAP) (0.5 units) was added to the restriction digests and incubated for 15 minutes at 37°C. Enzyme inactivation was achieved by heating the mixture at 65°C for 15 minutes.

4.3.10 Ligation of DNA fragments

4.3.10.1 Ligation of PCR fragments into T-overhang vectors

PCR products were ligated into the pGEM-T vector (Promega, Madison, WI) according to the following protocol.

The reaction mixture was incubated overnight at 16°C. The ligation reaction was used to transform competent bacteria and single clones were picked for the isolation of recombinant plasmid DNA that was sequenced and its insert subcloned into an appropriate vector where necessary.

Tab. 13: Pipetting scheme for ligation of PCR fragments into T-overhang vectors

Reagent	**Volume (**µl)
Vector (pGEM-T) DNA (50ng)	1
2X ligation buffer (Promega)	5
Insert DNA	3 (in three molar excess of the vector)
T4 DNA ligase	0.5
HPLC water	0.5
Reaction volume	10

4.3.10.2 Ligation of DNA fragments with sticky ends

Insert DNA was digested out of a vector using appropriate restriction enzymes, while the target vector was also digested with the same enzymes. Both were then ligated following the scheme below.

Tab. 14: Pipetting scheme for ligation of DNA fragments with sticky ends

Reagent	**Volume (**µl)
Vector DNA (50ng)	1
10 x ligation buffer (New England Biolabs)	2
Insert DNA	3 (in three molar excess of the vector)
T4 DNA ligase	1
HPLC water	13
Reaction volume	20

4.3.11 mRNA isolation, reverse transcription and RT-PCR

Worms (1g) were homogenised with a plastic pistil in 1000 µl of RNA isolation reagent (Peqlab) according to the manufacturer's instruction, and total RNA was isolated by phenol chloroform extraction followed by ethanol washes. Total RNA was reverse transcribed by Moloney murine leukemia virus reverse transcriptase (Promega) in the presence of random hexamer primers.

4.3.12 Polymerase chain reaction (PCR)

Thermostable DNA polymerases were used for the amplification of DNA as described (Saiki et al., 1988, Bej et al., 1991). Two types of PCR amplifications were used in this study: standard PCR for the amplification of DNA fragments up to 3 kb and Mid-range PCR for the amplification of fragments up to 9 Kb in size.

4.3.12.1 Standard PCR

These were done in 50 µl reaction volumes as shown in Table 15. Optimal annealing temperatures for primer pairs relative to a target template were obtained using the Oligo 4.0 software. PCR was performed on an MJ Research PT-200 DNA machine according to the thermal profile on Table 16.

Tab. 15: Pipetting scheme for standard PCR

Reagent	Volume (µl)
Forward primer (10 pmol/ul)	1
Reverse primer (10 pmol/ul)	1
dNTP mix (40 mM)	1
DNA template	1 – 5 (between 100 to 500 ng)
10 X reaction buffer	5
HPLC water	ad 50
Taq polymerase (5 U/ul)	0.5

Tab. 16: Thermal profile for standard PCR

Phase	**Temperature (°C)**	**Duration**	**Cycles**
Denaturation	94	2 min	1
Annealing	53-60	40 s	
Elongation	72	1 min per 1 kb	
Denaturation	94	20 s	30
Annealing	53-60	40 s	
Elongation	72	1 min per 1 kb	
Denaturation	94	20 s	1
Annealing	53-60	40 s	
Elongation	72	15 min	
	4		

4.3.12.2 Mid-range PCR

This PCR technique was used for the amplification of genomic DNA fragments between 3 and 9 kb (Barnes et al., 1994; Cheng et al., 1994). These were done using the Peq Lab mid-range PCR system as follows:

Tab. 17: Pipetting scheme for Mid range PCR

Reagent	Volume (µl)
Forward primer (10 pmol/ul)	2
Reverse primer (10 pmol/ul)	2
dNTP mix (40 mM)	1.75
DNA template	1 – 5 (between 100 to 500 ng)
10 X reaction buffer	5
HPLC water	ad 50
Mid-range polymerase (5 U/ul)	0.5

PCR was performed on an MJ Research PT-200 DNA machine as follows:

Tab. 18: Thermal profile for mid range PCR

Phase	Temperature (°C)	Duration	Cycles
Denaturation	94	2 min	1
Annealing	53-60	40 s	
Elongation	68	1 min per 1 kb	
Denaturation	94	20 s	30
Annealing	53-60	40 s	
Elongation	68	1.5 min per 1 kb	
Denaturation	94	20 s	1
Annealing	53-60	40 s	
Elongation	68	15 min	
	4		

4.4 Southern blotting

4.4.1 Digestion, electrophoresis, and blotting of genomic DNA

Genomic DNA (20 µg) were digested with different restriction enzymes in 100 µl reaction volumes as described in section 4.2.7, and electrophoresed in 0.7% agarose gels (section 4.2.2). The DNA was capillary-blotted onto a nylon membrane at room temperature in 20x SSC overnight (Sambrook et al., 1989). The membrane was then baked at 80°C for 2 hours to fix the DNA onto the membrane.

4.4.2 Radioactive labeling of *A. viteae* chitinase probe

A. viteae chitinase PCR fragments (25 ng) were labelled with 50 µCi α-^{32} P dCTP using the RadPrime DNA labelling system (Life Technologies). The pipetting scheme was as follows:

Tab. 19: Pipetting scheme for radioactive labeling of probes

Item	Volume (µl)
A. viteae chitinase PCR fragment (25ng)	5
dATP (500 µM)	1
dGTP (500µM)	1
dTTP (500µM)	1
Random primers solution	20
α-^{32} P dCTP (3000 Ci/mmol, 10mCi/ml)	5
Distilled water	16
Klenow fragment	1

The mixture was incubated at 37°C for 10 minutes and the reaction stopped. The radioactive probes were purified using Sephadex^{TM} MicroSpin G50 columns (Amersham Biosciences Europe GmbH). The labelling efficiency was about 30%.

4.4.3 Hybridisation and detection

Membranes were washed in 6x SSC at room temperature and prehybridised at 65°C in 10 ml hybridisation buffer (6x SSC, 5x Denhardt's reagent, 0.5% SDS, 100µg/ml salmon-sperm DNA) for 1 hour. The probe was denatured by boiling for 2 min in a water bath and added to the hybridisation buffer. Membranes were hybridised overnight at 65°C in a roller bottle. Stringent washes were done in 2x SSC, 0.5% SDS for 5 minutes at room temperature, 2x SSC, 0.1% SDS for 15 minutes at room temperature, and 0.1% SSC, 0.1% SDS for 2 hours at 65°C. The membranes were rinsed in 0.1x SSC, covered in Saran Wrap, and exposed to a phosphorimager plate (Fuji Film) for 3 hours, and the bands were scanned using a phosphoimager.

4.5 Microbiological methods

4.5.1 Preparation of competent *E. coli*

Competent *E. coli* cells were prepared essentially as described by Inoue et al. (1990). *E. coli* were streaked on LB agar plates with antibiotics and cultured overnight at 37°C. Twelve large colonies were picked with a sterile toothpick and used to inoculate 125 ml SOB in a 1-liter Erlenmeyer flask. Flasks were incubated at 18°C with vigorous shaking (220 rpm) and the bacteria grown to an OD of 0.6 (mid-log phase). Bacteria cultures were poured into Falcon tubes and incubated on ice for 10 minutes and then centrifuged at 2500x g (3000 rpm) in an Eppendorf 5403 bench top centrifuge. The pellet was resuspended in 40 ml ice-cold TB buffer, incubated on ice for 10 minutes and centrifuged as above. The pellet was then carefully resuspended in 10 ml of TB, and DMSO added drop-wise to a final concentration of 7%. The bacterial suspension was incubated for ten minutes on ice, after which 1 ml aliquots were snap-frozen in liquid nitrogen and stored at – 80°C for up to one month.

4.5.2 Transformation of competent *E. coli*

Eppendorf tubes (1.5 ml) were pre-chilled on ice and 10 ng of purified plasmid DNA in a total volume of 20 µl were pippeted into the tubes. Competent cells were thawed, and 200 µl were dispensed into the tubes on ice. The tubes were flicked gently to mix and incubated on ice for 30 minutes. The cells were then heat-shocked by heating for exactly 45 seconds in a 42°C water bath, and then incubated on ice for 2 minutes. Room temperature SOC medium (800 µl) was added to each tube on ice. The tubes were then incubated at 37°C while shaking vigorously (220 rpm) for one hour. Antibiotic selective agar plates were plated with 50 to 150 µl of the transformation mixture and incubated overnight at 37°C. Positive controls were competent cells transformed with 10 ng pBluescript SK (-) plasmid, while negative controls were competent cells transformed without DNA.

4.5.3 Screening of bacterial colonies for plasmids/ recombinant plasmids

Two methods were routinely employed to identify bacteria colonies that contain the recombinant plasmids or plasmids of interest.

4.5.3.1 Blue-white screening

Most bacteria strains (e.g. JM109,DH5 alpha and XL 1 Blue), used as hosts for initial cloning of a target gene, code for an inactive carboxy-terminal portion of ß-

galactosidase due to a mutation in the lac Z gene (lacZdeltaM15). On the other hand, many vectors contain the regulatory sequences and information that could be induced (e.g. by IPTG) to code for the missing amino terminal (alpha region) of the ß-galactosidase gene. Both host and plasmid encoded fragments are inactive. However, if a bacterial host is transformed with a plasmid expressing the missing amino terminal of the enzyme, both fragments associate to form an active enzyme in a process called alpha complementation. Such bacteria can be recognised because they metabolise the chromogenic substrate 5-bromo-4-chloro-3-indolyl-ß-D-galactoside (X-gal) in the presence of isopropyl-ß-D galactopyranoside (IPTG) to form blue colonies. Insertion of a target gene in the alpha peptide region coding for the ß-galactosidase enzyme results to insertional inactivation of the alpha peptide. Bacteria carrying plasmids with the target gene will form white colonies since they lack active ß-galactosidase. Thus, transformed bacteria were plated on LB-agar plates supplemented with 0.5 mM IPTG, 80 µg/ml X-gal and antibiotics.

4.5.3.2 Restriction analysis of isolated plasmids

In order to verify constructs of plasmids/ inserts resulting from cloning, plasmid DNA was isolated and analysed. Independently transformed bacterial colonies were picked and grown in 5 ml LB/ antibiotic overnight. Plasmid DNA was isolated (section 4.3.2) and digested (section 4.3.8) at restriction sites used for cloning of the insert DNA. The digested DNA was then analysed by agarose gel electrophoresis (Section 4.3.3).

4.5.4 Bacteria cultures and long-term storage of bacterial stocks

Bacteria host strains used in this study included XL 1 Blue, BL 21 (DE3), and Origami B (DE3). Bacteria were streaked on LB agar plates and grown overnight at 37°C. For long-term storage of bacteria, a single colony was grown to an OD_{600} of 0.6 and 1 part was mixed with 9 parts of Hogness freezing medium (36 mM K2HPO4, 13 mM KH2PO4, 20 mM sodium citrate, 10 mM MgSO4 and 40 % glycerol). Tubes with bacteria stocks were snap-frozen in liquid nitrogen and stored at -80°C. Bacteria recovery involved scraping the surface of a bacteria stock with a sterile loop and streaking an LB plate with the appropriate antibiotics.

4.5.5 Expression and purification of recombinant proteins from *E. coli*

The pET expression system was used for prokaryotic expression of recombinant N-terminal *A. viteae* chitinase. In this system, target genes are cloned under the control

of the T7 promoter which is not recognised by *E. coli* RNA polymerase. Some *E. coli* host strains (like BL21 (DE3) and Origami B (DE3)) have a chromosomal copy of the T7 RNA polymerase gene under control of *lacUV5*. Transfer of recombinant pET plasmids into such hosts results in an IPTG-inducible gene expression system. The cDNA encoding the N-terminal domain of chitinase was cloned into the pET 22 b (+) vector and transformed into Origami B (DE3 cells). The pET 22 b (+) vector has a *pelB* signal sequence that leads to the localisation of an expressed protein into the periplasmatic space of the *E. coli* host.

A single bacterial colony was picked and used to inoculate LB (100 µg/ml Ampicillin, 12.5 µg/ml Tetracyclin, 15 µg/ml Kanamycin and 1% glucose) and the culture was grown overnight at 37°C, 150 rpm. The overnight culture was diluted 1:50 in fresh LB medium and further grown at 37°C till an OD_{600} of 0.4. The culture was shifted to room temperature and further grown to an OD_{600} of 0.6. Protein expression was induced with 1mM IPTG for 4 hours. Bacteria were pelleted by centrifugation at 6000 rpm for 15 minutes at 4°C. The pellet was then used for the extraction and purification of periplasmatic proteins and proteins in inclusion bodies.

For extraction of periplasmatic fluid, bacteria pellet from a 500 ml culture was resuspended in 30 ml 30mM Tris-HCl pH 8, 20% sucrose. EDTA was added to a final concentration of 1 mM and the suspension stirred for 10 minutes at room temperature. Cells were collected by centrifugation at 10,000xg at 4°C for 10 minutes. The pellet was then re-suspended in 30 ml of ice-cold 5 mM $MgSO_4$ and stirred slowly on ice for 10 minutes to release the periplasmatic proteins. The suspension was centrifuged at 10,000xg to pellet the cells and the supernatant was used for purification of periplasmatically expressed *A. viteae* chitinase. The insoluble fraction in inclusion bodies was purified as described below.

The pellet was resuspended in lysis buffer A, 100 ug/ml lysozyme and incubated for 30 minutes at 4°C. The suspension was then sonicated for 1.5 min, 30 cycles. After centrifugation (13,000 rpm, 20 min, 4°C), the supernatant was transferred into a falcon tube, incubated with 1 ml Nickel chelate matrix (Qiagen, Hilden) and incubated for 1hr at room temperature with constant gentle shaking. The matrix with bound protein was then loaded onto a column and washes were done with 20 ml of buffers B and C, respectively. Elution was done in 20 ml of buffer E, and collected in 1ml fractions. The fractions were analysed by SDS-PAGE for protein and the positive

fractions were pooled. This pool was then dialysed in phosphate buffers with reducing amounts of urea and the protein concentration was determined. Aliquots of the protein were stored at -20°C until used.

4.6 Protein analytical methods

4.6.1 Determination of protein concentration

Protein concentration was determined by the BCA test using the BCA kit (Pierce, Rockford, USA). The basis of this reaction is the biuret reaction: reduction of Cu^{2+} to Cu^{1+} by a peptide bond under alkaline conditions. Chelation of two molecules of bicinchoninic acid (BCA) by the cuprous ion (Cu^{1+}) produces a water soluble complex, whose solution has a deep purple colour and an absorbance at 562 nm. A linear standard curve was made with 0.05 to 2mg BSA.

4.6.2 Sodium dodecly sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Analytical polyacrylamide gel electrophoresis in sodium dodecyl sulfate was performed as described by Laemmli (1970). Samples were heated for 5 minutes at 100°C in sample buffer containing 125 mM Tris-HCL, pH 8.3, 1% SDS, 2.5% ß-mercaptoethanol and 10% glycerol. Samples were applied at volumes of 10-30 µl to a 12.5% slab gel cast in a mini-gel apparatus (Hoefer). The electrophoresis was performed at 20mA with electrode buffer (25mM Tris-HCL, pH 8.3, 192mM Glycine and 0.1% SDS). The run was stopped when the bromophenol blue dye front reached the end of the gel. Protein bands were then revealed by staining with Coomassie Brilliant Blue R250, and the gels dried for a permanent record.

4.6.3 Chitinase activity assays

Chitinase assays were performed using 4-methylumbelliferyl ß-N, N', N''-triacetylchitotrioside $(GlcNA)_3UMB$. $(GlcNA)_3UMB$ was dissolved in reaction buffer (20mM $NaPO_4$, pH 6.0, 200mM NaCl, 1mM EDTA) and 95µl of 20 µM substrate solution were mixed with 5µl of enzyme and incubated for 2 to 10 minutes at 25°C. The reactions were stopped by adding 1.9 ml of 0.5 M glycine, pH 10.5. Release of free methylumbelliferone was measured by a fluorescence spectrophotometer at 360 nm excitation and 450 nM emission wavelengths. The fluorometer was scaled at the beginning of the experiments such that 20 nM substrate had a fluorescence reading of 10 units.

4.7 Immunochemical and immunological methods

4.7.1 Western blot

4.7.1.1 Transfer of proteins onto nitrocellulose membranes

Proteins separated by SDS-PAGE were immobilised onto nitrocellulose membranes by electrophoretic transfers (Towbin et al., 1979). A transfer cassette was assembled in transfer buffer using a sponge followed progressively by a 3 MM- Whatmann paper, nitrocellulose membrane (pore size 0.2 uM), SDS-PAGE gel, another Whatmann paper and finally a second sponge. The cassette was introduced into a transfer tank between electrodes so that the gel with the protein was towards the negatively charged cathode and the nitrocellulose membrane towards the positively charged anode. A constant current of 70 mA was then applied across the cassette at room temperature overnight. To determine the transfer efficiency, the membrane was reversibly stained in Ponceau S dye (0.1% Ponceau S, 7% trichloroacetic acid in distilled water). After complete washing of the Ponceau stain, unbound sites of the membrane were blocked by incubation for 45 minutes in 5% skimmed milk powder in PBS.

4.7.1.2 Immunodetection of immobilised proteins

Nitrocellulose membranes with immobilised proteins were incubated with a primary protein specific antibody for one hour at room temperature, followed by 3 x 5 minutes washes in wash buffer (PBS, 0.02% Tween 20). The membrane was then incubated in a secondary conjugate antibody for one hour followed by washes as above. Detection was done by using a substrate for the alkaline phosphatase enzyme conjugate, 5-Brom-4-chlor-3-indolylphosphate (BCIP) and tetrazolium chloride (NBT) in alkaline phosphatase buffer. All antibodies were diluted in blocking solution (5% skimmed milk powder in PBS).

4.7.2 Coupling of peptides to KLH

Synthetic peptides (20 amino acids long) were synthesised with two extra N-terminal cysteins and coupled to KLH (Keyhole limpet haemocyanin) to make them immunogenic. Coupling was done using the Imject maleimide activated mcKLH kit (Pierce USA) and the coupled proteins were used at a concentration of 100µg in vaccinations.

4.7.3 Immunisation studies

The protective potential of *A. viteae* N-terminal chitinase was evaluated in the *Meriones/A. viteae* natural host-parasite system. Eight to ten-week-old *Meriones* were anaesthesized with ketamin:xylazin:normal saline (1:1:8) and immunised with 25 µg recombinant *A. viteae* chitinase or 100 µg KLH-coupled synthetic *A. viteae* chitinase peptide in adjuvant. The animals were immunised three times in a two-weekly interval, after which they were challenged with 70 freshly isolated L3s by injecting in the neck region. Microfilaria load was verified at weeks 4, 8 and 12 post infection (pi), and the animals were dissected at week 12 pi for isolation of adult worms and determination of protective potential.

Two adjuvants were used for immunisation, depending on the immune reaction desired: Alum was used for humoral-mediated immune reactions, while STP was used for cell-mediated immune responses.

4.7.3.1 Bleeding of animals for production of sera

Mice and jirds were bled from the retro-orbital sinus using a microhaematocrit capillary, and the blood was stored at 4°C overnight. Blood samples were then centrifuged (10,000 rpm, 20 minutes, 25°C) to separate the sera from blot clots. Sera were stored in 100 ul aliquots at -20°C.

4.8 Parasitological methods

4.8.1 Maintenance of the life cycle of *A. viteae*

The life cycle of *A. viteae* was maintained in *Ornithodoros moubata* essentially as described by Lucius and Textor, 1995.

4.8.2 Quantification of microfilarial load in blood of jirds

Infected *Meriones* were anaesthesized with ketamin:xylazin:normal saline (1:1:8) and bled from the retro-orbital sinus using a heparinised glass capillary tube. 20 µl blood were mixed with 100µl Teepol (10% in H_2O), and the microfilaria load was counted using a Fuchs-Rosenthal counting chamber.

4.8.3 Isolation of filariae

4.8.3.1 Isolation of adult A. viteae

Meriones were anaesthetized and fully bled from the retro-orbital sinus. Following dissection of the jirds, adult *A. viteae* were isolated from the subcutaneous and

intramuscular tissues, the iunguinal and subscapular regions and some times in the thoracic chamber. Animal carcasses were then incubated in normal saline (0.9% NaCl) overnight to allow the rest of the worms to wander out.

4.8.3.2 Isolation of L3 from the vector Ornithodoros moubata

L3s were isolated using a Baermann apparatus comprising of a funnel with a tap. The inner part of the funnel was covered with bandage and filled with Ringer's solution. Ticks were cut medially and briefly rinsed in a Petri dish with Ringer's solution to remove rests of blood meal and loose tissue. The ticks were incubated in warm Ringer's solution in the funnels for 1 hr for the L3s to migrate into the solution. The tap was then opened to collect the L3s in 50 ml falcon tubes. The larvae were amply washed in Ringer's solution and rinsed in RPMI if they were to be used for culture.

4.8.3.3 Isolation of uterine microfilariae from gravid female worms

Adult female *A. viteae* worms were isolated on day 90 post-infection (average length, 6 cm; embryogenesis starts on day 28 post-infection) and cut into 3mm segments in Ringer's solution. Developmental stages were separated from the worm segments by filtration. The stages obtained included: oocytes attached to the rachis of the ovary, fertilized eggs, multicellular stages, invaginated stages, pretzel and ring stages, and elongated microfilariae.

4.9 Computer analysis and statistical methods

4.9.1 Analysis of DNA sequences

Genomic DNA sequences obtained from primer walking were assembled into contigs by using the program MacVector 7.2 (Accelrys, USA). Assembled sequences were fed into Artemis® (www.sanger.ac.uk/Software/Artemis/) for visualisation of sequence features, its six-frame translations and exon-intron prediction. Exons thus predicted were verified by comparison to a cDNA sequence and by further analysis using Netgene (www.cbs.dtu.dk/services/NetGene2/). *Brugia malayi* genomic chitinase DNA sequences were retrieved from the server for the *B. malayi* genome project at http://www.tigr.org/tdb/e2k1/bma1/ while EST sequence analysis was done at the server of the Filaria genome project (http://nema.cap.ed.ac.uk/fgp.html).

4.9.2 Statistical analysis

The worms recovered from immunised and non-immunised *Meriones* in immunisation experiments were compared using the Mann-Whitney U-test at the 95% significance level ($p<0.05$).

5 Materials

5.1 Commercial Kits and Enzymes

Collagenase A	Biochrom, Berlin
RNase A	Boehringer Mannheim, Mannheim
T4 DNA ligase	New England Biolabs, USA
DNA restriction enzymes	New England Biolabs, USA
Mid-range Polymerase	PeqLab, Erlangen
Sawady Taq Polymerase	PeqLab, Erlangen
Lysosyme	Sigma, Deisenhofen
Qiagen protein purification kit	Qiagen, Hilden
Nucleospin DNA extraction kit	Clontech, Heidelberg
Nucleospin plasmid prep kit	Clontech, Heidelberg
BCA protein assay Kit	Pierce, USA
KLH protein conjugation kit	Pierce, USA
PGEM-T^{TM} Cloning kit	Promega, Madison, USA
Random Primers DNA labeling System	Life Technologies, Karlsruhe
Nick Translation System	Life Technologies, Karlsruhe

5.2 Laboratory Equipment and consumables

Chromatographic columns	Pharmacia, Freiburg
Agarose gel electrophoresis apparatus	AGS, Heidelberg
SDS-PAGE apparatus	Pharmacia, Biotech
Video documentation apparatus E.A.S.Y.RH	Herolab, Wiesloch
Peltier Thermal Cycler, PTC-200	MJ-Research, Massachusetts
Bandelin Sonoplus HD 200 sonicator	Berlin
Spectrophotometer	Eppendorf
DE52 powder (DE52 number 4057-050)	Whatman, USA
Chromatography paper (3MM)	Whatman, USA
Dialysis membrane	Serva, Heidelberg
Nitrocellulose membrane (0.2 µm)	Schleicher & Schuell
Cell culture plates (Flat & round bottom)	Costar, Bodenheim

HPLC water	Roth, Karlsruhe
CSPD	Boehringer; Mannheim

5.3 Synthetic oligonuceotides

Oligogonucleotides were delivered as lyophilised powders and dissolved to final concentrations of 100 pmol/µl in water as per manufacturer's instruction. This stock solution was then dissolved 1:10 in water to obtain working solutions with concentrations of 10 pmol/ µl. A total of 10 pmol primer was used in standard PCRs, while 20 pmol were used in long range PCR.

5.3.1 PCR Primers

5.3.1.1 Primers used in long range PCR

Name	**Sense**	**Clone**	**Sequence**
T7- long		all	GTA ATA CGA CTC ACT ATA GGG C
T3-long		all	AAT TAA CCC TCA CTA AAG GGA AC
F3	+	all	CTACGTTCGCGGATGTTAC
F4	+	all	TGGGCTTGAAAGTGAGGTAAG
R4	-	all	TGTTTGCTCACTTTCAAGCCC
R7	-	12	TCCCAACTGCCGTGTAAATCA
R8	-	12	CATTTCCGACAGTAATACGAT
R5	-	1	CCGGCAGAAATACAATGCTTG

5.3.1.2 Primers used for standard PCR

5.3.2 Sequencing primers

5.3.2.1 Universal sequencing primers

5.3.2.1.1 Primers on pGEM-T vector

Name	Sense	Sequence
SP6	-	ATTTAG GTG ACA CTA TAG
T7	+	TAA TAC GAC TCA CTA TAG GG

5.3.2.2 Primers on pBluescript SK (-) vector

Name	Sense	Position	Sequence
T3	-	772-791	ATTAACCCTCACTAAAGGGA
T7	+	626-645	TAATACGACTCACTATAGGG
M13 forward	+	603-620	GTA AAACGACGGCCAGT
M13 reverse	-	811-828	CAGGAAACAGCTATGAC

5.3.3 Primers used for sequencing *A. viteae* chitinase genomic clone 1

Name	Sense	Position	Sequence
F3	+	5679-5698	CTACGTTCGCGGATGTTAC
F3walk	+	6107-6127	ACACCTATGCACTCACATTC
F4	+	8698-8719	TGGGCTTGAAAGTGAGGTAAG
F31	+	8075-8094	TCGCCATCGACAACGAATC
F5	+	8653-8674	AAGTTGTGGCAAAGGTCCATA
F6	+	6521-6542	CATCAGTTTTTGCTGTTCGTT
F7	+	8852-8873	GGACCATACGGAAAAATGATT
F7walk	+	9208-9229	CAGATAATACTGAAATGCCAG
R4	-	8698-8709	TGTTTGCTCACTTTCAAGCCC
R5	-	6025-6046	CCGGCAGAAATACAATGCTTG
R5walk	-	5473-5493	CAATAGATGCTACTCGACAG
R6	-	4837-4858	GGAATTCAATGAGTTATGCTG
R6walk	-	4280-4301	GACGTATAACACAAAGTACG
R7	-	7772-7792	TCCCAACTGCCGTGTAAATCA
M13 forward	-	vector	GTA AAA CGA CGG CCA GT
R3n1	-	12355-12376	TCAGGCGCTAAGAAAAGAGAG
R3n1walk	-	11550-11573	TAAGGCTCATTTAACTTCGATAT
R3n2	-	10722-10743	TGGACCATTTCGGGTTTTCTG
Rur6	-	3706-3727	ATTCTTTGCACCGCNTGTTTC
Rur6walk	-	2948-2968	TAGGTAAAGCGAAAGAGCTG
Rur6.1	-	2217-2239	GTGAAAATGCGGAAACGGAATA
Rur6.1walk	-	1392-1412	ACTTCCTGAGGCTCTCCCTG

5.3.4 Primers used for sequencing *A. viteae* chitinase genomic clone 9

Name	**Sense**	**Position**	**Sequence**
F4	+	6645-6666	TGGGCTTGAAAGTGAGGTAAG
F>e10_9	+	8187-8208	GGAGAGCCTCAGGAAGTAGTG
K9_4pups	-	9414-9435	ACTAGGAAATTGGGAAAATACAAG
Rb4_4pupstr	+	10265-10286	TGCACTGTCCCATTTATTGTC
K9_3pup.	-	10516-10537	CTTTGTACCGCTTGTTTCTTGTT
M13 forward	+	vector	GTA AAACGACGGCCA GT
M13 reverse	-	vector	CAGGAA ACAGCT ATG AC
K9-FV	+	780-801	GGAAATGCAAATACAAACGGC
K9-RV	-	12925-12946	AGGATGTGAGTGCATAGGTG
K9-V1	-	11359-11380	GTACTTTATCAAACGTACCGT
K9-V2	-	9973-9993	GAAATAAAATCAGGAAGTTGC
K9-V3	-	8632-8650	TCAATTCATTTGACTGCCAC
K9-V4	+	2228-2249	CGAATATTCGTTTATCCATTG
K9-F1	+	3773-3795	GTAGGAAATGTCACATAACATC
K9-F2	+	4408-4427	CACTGATCGATGTAATATGC
K9-F3	+	5227-5240	CTGTGACGATCACTACCAGC
K9-F4	+	6117-6138	GTGAATAGCAACTTCTTTAGC

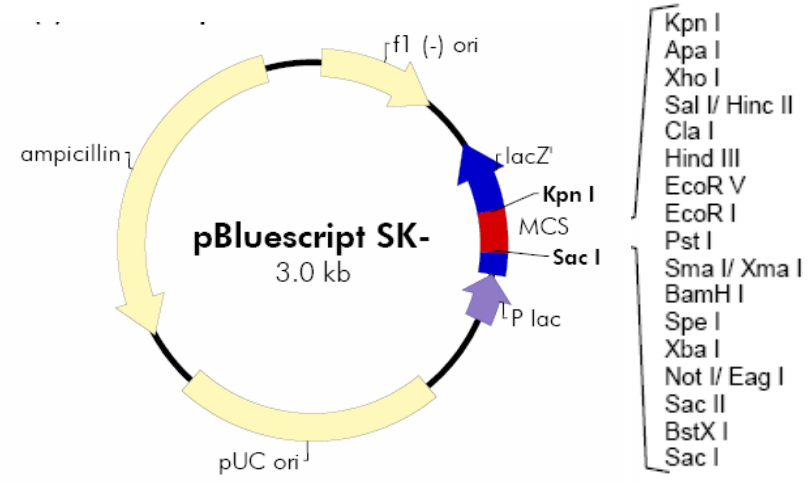
5.3.5 Primers used for sequencing *A. viteae* chitinase genomic clone 12

Name	Sense	Position	Sequence
F3	+	1374-1372	CTACGTTCGCGGATGTTAC
F3.walk12	+	2108-2129	GACGAAGATACCGAATGGTC
F4	+	4587-4607	TGGGCTTGAAAGTGAGGTAAG
F5	+	4542-4562	AAGTTGTGGCAAAGGTCCATA
F5.1	+	5288-5309	TACGTGCATCAAACACAGTC
F8	+	4130-4151	TGGCAAGGAAACGGTGGATAA
R4	-	4588-4608	TGTTTGCTCACTTTCAAGCCC
R7	-	3662-3683	TCCCAACTGCCGTGTAAATCA
R8	-	1709-1729	CATTTCCGACAGTAATACGAT
R13	-	543-564	AACCACGCTGAAGCCAAAATA
R13.1	-	950-971	GTAAGACTTGTACACTTCTG
F13	+	5836-5857	GTTTATTTCCGCATCACAGTG

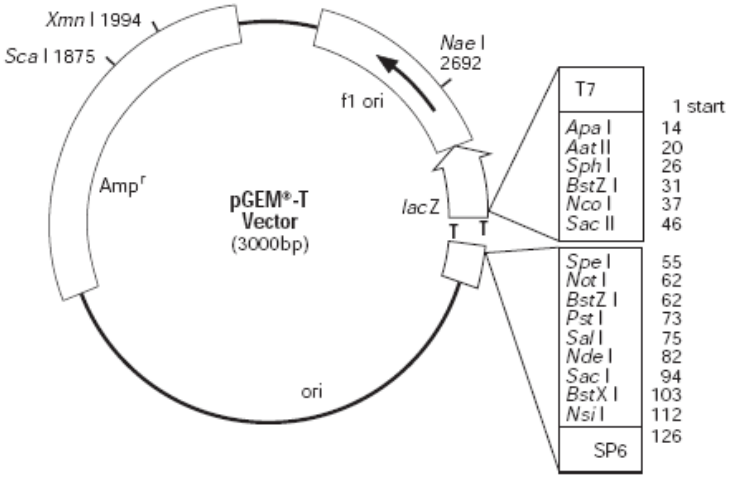
5.4 Plasmids

5.4.1 Cloning plasmids

5.4.1.1 Cloning plasmid for genomic inserts: pBluescript (Stratagen)

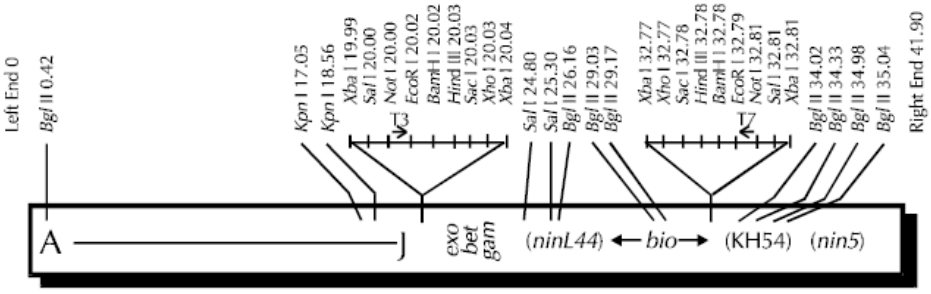


5.4.1.2 TA- overhang vector for cloning PCR products

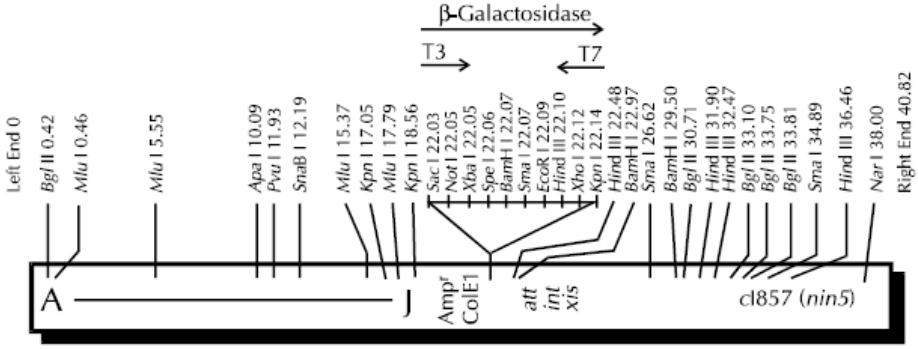


Source: Promega

5.4.2 Plasmids harbouring genomic and cDNA inserts: λ Dash II and λ Zap, respectively

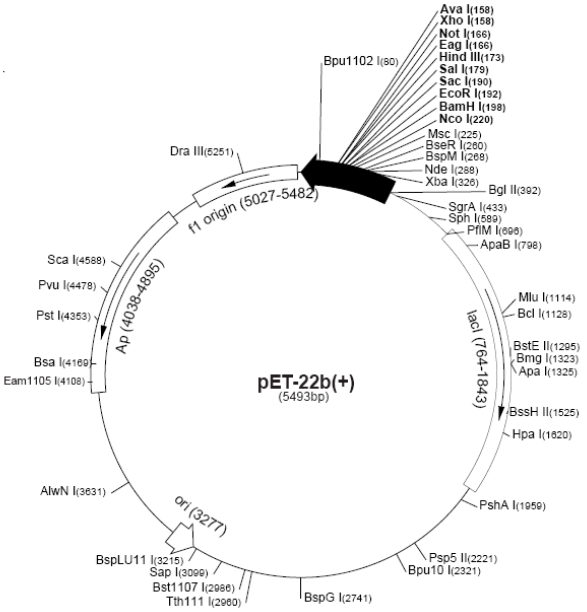


Lambda dash II replacement vector map (source: Stratagene)



Lambda Uni-Zap XR insertion vector

Expression vector:Vector maps of pET 22 b(+) expression vectors (Novagen)



5.5 Buffers and solutions

5.5.1 Molecular biology

5.5.1.1 Buffers for separation, purification and storage of DNA

TAE
40 mM Tris/Acetate pH 8.0
1 mM EDTA

TE Buffer
10 mM Tris/HCl, pH 7.5
1mM EDTA

TlowE Buffer
10 mM Tris/HCl, pH 7.5
0.1 mM EDTA

Phenol chloroform solution
50 % (v/v) Phenol, pH 7.5
48% (v/v) Chloroform
2% (v/v) isoamyl alcohol

10x Sample buffer
250 mg bromophenol blue
33 ml 150 mM Tris pH 7.6
60 ml glycerol
7 ml H2O

Chloroform Isoamyl alcohol
96% (v/v) Chloroform
4% (v/v) Isoamylalcohol

Ethidium Bromide stock solution
10 mg/ml in H_2O

5.5.1.2 Buffers for purification of plasmids: CTAB miniprep

RNase A 10 mg/ml
Lysozyme stock 50 mg/ml

STET
50 mM Tris-HCl, pH 8
8% sucrose
0.1% triton X-100
50 mM EDTA

CTAB solution
5 % CTAB in H_2O

5.5.1.3 Buffers for purification of lambda DNA

SM buffer	LB medium
5.8 g NaCl	see 5.6
2 g $MgSO_4.7H_2O$	
50 ml 1 M Tris.Cl, pH 7.5	
5 ml 2% (v/v) gelatine solution	
H_2O to 1 l, autoclave	

5.5.1.4 Buffers and reagents used for screening of libraries

Hybridisation	
Blocking solution	10 % (w/v) in MS 1 buffer
MS 1 buffer	0.1 M Maleic acid 0.15 M NaCl adjust pH 7.5 @ 20 °C; autoclave
20 x SSC	175,2 g NaCl 88.4 g Na-citrate adjust pH to 7 ; autoclave
Denaturing solution	0.5 M NaOH, 1.5 M NaCl
Neutralisation solution	0.5 M Tris/HCl, pH 7.5, 1.5 M NaCl
Prehybridisation buffer	50 ml 20 x SSC, 100 ml water, 20 ml blocking solution, 2.1 ml 10 % sarcosyl, 0.4 ml 10 % SDS add dH2O to 200 ml
Wash buffer I	2 x SSC, 0.5% SDS
Wash buffer II	2 x SSC, 0.1% SDS
Wash buffer III	0.1 x SSC, 0.1% SDS

Antibody reaction and development	
Wash buffer:	MS 1 + 0.3 % (v/V) Tween 20
MS 2 buffer:	block solution diluted 1: 10 in MS 1
MS 3 buffer:	0.1 M Tris, 0.1 M NaCl, 0.05 M MgCl, pH 9.5; autoclave
Antibody solution:	dilute 1: 7,500 with MS 2 buffer (from Boehringer)
CSPD solution:	dilute 1:200 in MS 3 buffer

5.5.1.5 Buffers and reagents used for Southern blotting

20 X SSC	see above
Prehybridisation buffer	6 x SSC
	5 x Denhardt's reagent
	0.5% SDS
	100 µg/ml Salmon sperm DNA
Wash buffers	See wash buffers I, II, and III above

5.6 Media and buffers for *E. coli*

LB Medium	Top Agar
10 g Bacto-Tryptone	0.75% agarose in LB
5 g Bacto-Yeast	
10 g NaCl	10x Hogness freezing medium: for dauer cultures
Add H_2O to 1l, pH 7.0	36 mM K2HPO4
	13 mM KH2PO4
LB Media with antibiotics	20 mM Sodium citrate
LB supplemented with:	10 mM MgSO4
100µg/ml Ampicillin and/or	40 % Glycerin
30 µg/ml Kanamycin and/or	
12.5 µg/ml Tetracyclin	
Agar plates	SOB Medium
LB medium supplemented with:	20 g/l Bacto Trypton
15 g/l agar and	0.5 g/l NaCl
Antibiotics (see above)	5g/l Bacto Yeast Extract
Dispensed into Petridishes	pH 7.0 with NaOH
	2.5 mM KCl
	10 mM $MgCl_2$ (added before use)

SOC Medium

25 mM Glucose in ready-to-use SOB medium

TB buffer (transformation buffer)

10 mM PIPES

15 mM $CaCl_2$

250 mM KCl

adjust pH to 6.7 with KOH

add 55 mM $MgCl_2$

sterile filter

5.7 Protein and immuno- chemistry

5.7.1 SDS-PAGE

Separating gel buffer

1.5 M Tris/HCl, pH 8.8

Stacking gel buffer

1 M Tris/HCl pH 6.8

10x SDS-PAGE running buffer

10 g SDS

30 g Tris

144 g glycin

In 800 ml H_2O

Add H_2O to 1l

2x SDS PAGE sample buffer

10 ml 1.5 M Tris/HCl (pH 6.8)

6 ml 20% SDS

30 ml glycerol

1.8 mg bromophenol blue

Add 168µl 3M DTT stock to 1 ml before use

Coomassie staining solution

0.2% Brilliant blue G-250

40% Methanol

1% acetic acid

Destaining solution

10% acetic acid

30% Ethanol

Preparation of gels

	Separating gel			Stacking gel
	14%	**12.5%**	**10%**	**4%**
Acrylamide stock (30%)	18.7 ml	16.7 ml	13.3 ml	2.7 ml
1 M Tris HCl, pH 6.8	---	---	---	5 ml
1.5 M Tris HCl, pH 8.8	10 ml	10 ml	10 ml	---
10% SDS	400µl	400 µl	400 µl	200 µl
10% APS	200 µl	200 µl	200 µl	300 µl
H_2O	10.7 ml	12.7 ml	16.1 ml	12 ml
TEMED	20 µl	20 µl	20 µl	20 µl

5.7.2 Western blot

Transfer buffer (1l)

2.9 g glycin

5.8 g Tris

0.37 SDS

In 200 ml Methanol

Adjust volume to 1l

Blocking buffer

5% skimmed milk powder in PBS

Washing buffer

0.2% Trixton X 100 in PBS

Ponceau S staining solution

2% Ponceau S

30% Trichloroacetic acid

30% Sulphobenzoic acid

Alkaline phophatase buffer

100 mM Tris-HCl, pH 9.5

100 mM NaCl

5 mM $MgCl_2$

Substrate stock solutions

NBT solution: 75 mg/ml in 70% DMF

BCIP solution: 50 mg/ml in 100% DMF

Alkaline phosphatase substrate (working solution)

66 µl NBT solution

33 µl BCIP solution

In 10 ml of alkaline phosphatase buffer

5.7.3 Buffers for native purification of N-terminal chitinase with Ni^{2+}-NTA

Lysis buffer	Elution buffer
50 mM NaH_2PO_4	50 mM NaH_2PO_4
300 mM NaCl	300 mM NaCl
10 mM imidazol	250 mM imidazole
pH to 8.0 with NaOH	pH to 8.0

Wash buffer
50 mM NaH_2PO_4
300 mM NaCl
20 mM imidazol
pH to 8.0

5.7.4 Buffers for denaturing purification of N-terminal chitinase with Ni^{2+}-NTA

Buffer A (1 liter)	Buffer C (1 liter)
100 mM NaH_2PO_4	100 mM NaH_2PO_4
10 mM Tris.Cl	10 mM Tris.Cl
8 M urea	8 M urea
adjust pH to 8.0 using NaOH	adjust pH to 5.9 using HCl

Buffer B (1 liter)	Buffer D (1 liter)
100 mM NaH_2PO_4	100 mM NaH_2PO_4
10 mM Tris.Cl	10 mM Tris.Cl
8 M urea	8 M urea
adjust pH to 6.3 using HCl	adjust pH to 4.5 using HCl

5.7.5 Synthetic peptides

B1 = Peptide 0: CGGLLSYGGYNFGSSTFTAIA
C1 = Peptide 1: CGGGFDLDWEYPTGVAEEHAK
D1 = Peptide 2: CGGTAAVSAGKMTIDESYNVQ

5.7.6 Buffers and solutions for chitinase enzyme activity assay

10x Enzyme assay buffer	Stop buffer
20 mM NaPO4, pH 6.0	0.5 M Glycin buffer, pH 11.0
0.2 M NaCl	
1 mM EDTA	

Substrate

20 mM 4-methylumbelliferyl N´,N´´,N´´´ß Chitotrioside in 1x reaction buffer

5.7.7 Stock solutions

10 X PBS	80 g NaCl 2.0 g KCl 2.4 g KH_2PO4 14.4 g Na_2HPO4 in 1 l dH_2O, pH 7.4
1 M Tris	121.14 g Tris In 800 ml H_2O 42 ml conc. HCl pH to 8.0 ad H_2O to 1l

5.8 Antibiotic stock solutions

Ampicillin sodium salt	100 mg/ml in dH_2O
Kanamycin sulfate	10 mg/ml in dH_2O
Tetracycline	10 mg/ml in 100% ethanol

5.9 *E. coli* host strains and plasmids

XL-1 Blue	Stratagene, Heidelberg
XL-1 Blue MRA	Stratagene, Heidelberg
BL 21 (DE 3)	Novagen, Darmstadt
Origami B (DE 3)	Novagen, Darmstadt
pET 22 b(+)	Novagen, Darmstadt
pET 41 a (+)	Novagen, Darmstadt

5.10 Databanks, softwares and online services

Blast server for ESTs: http://www.ebi.ac.uk/blast2/parasites

Sequence analysis: http://www.ebi.ac.uk/services

Brugia malayi genome: http://www.tigr.org/tdb/e2k1/bma1/

Filaria genome Project: http://nema.cap.ed.ac.uk/fgp.html

5.11 Softwares

Genomic DNA analysis, Artemis: www.sanger.ac.uk/Software/Artemis/

Clustalw: http://www.ebi.ac.uk/clustalw/

Netgene: www.cbs.dtu.dk/services/NetGene2/

MacVector 7.2: Accelrys, USA

Oligo 4.0:http://www.oligo.net/

6 Abbreviations

APS	Ammonium peroxidisulfate
A	Adenosine
APOC	African Programme for Onchocerciasis Control
Amp	Ampicillin
AP	Alkaline phosphatase
bp	Base pairs
BCA	2,2´Bicinchoninic acid
BCIP	5-Bromo-4-chloro-3-indolyphosphate disodium salt
BSA	Bovine serum albumin (Fraction V)
cDNA	Complementary DNA
CTAB	Cetyltrimethylamoniumbromide
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediamino tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EtBr	Ethidium bromide
EtOH	Ethanol
HPLC	High performance liquid chromatography
IPTG	Isopropyl-thio-ß-D-galactopyranoside
IU	International unit
kb	Kilo base
L_1	First stage larvae
L_2	Second stage larvae
L_3	Third stage larvae
L_4	Fourth stage larvae
LB	Lauria Bertani
LM	Low melting
2-ME	ß-Mercaptoethanol
mAb	Monoclonal antibody
Mf	Microfilariae
mRNA	Messenger RNA
NBT	Nitroblue tetrazoliumchloride
NC	Nitrocellulose

OCP	Onchocerciasis Control Program
OD	Optical density
ON	Overnight
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethyleneglycol
PIPES	1,4-Piperazinediethane sulfonic acid
PMSF	Phenylmethylsuöphonylflouride
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
RT	Room temperature
SDS	Sodium dodecylsulfate
TAE	Tris-acetate-EDTA
TB	Transformation buffer
TBS	Tris-buffered saline
TE	Tris-EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
TRIS	2-Amino-2-hydroxymethyl-1,3-propandiol
U	Units
UV	Ultraviolet
X-gal	5-Bromo-4-chloro-3-indoxyl-ß-D-galactoside

7 References

Abraham D, Eberhard ML, Lange AM, Yutanawiboonchai W, Perler FB, Lok JB. (1992): Identification of surrogate rodent hosts for larval Onchocerca lienalis and induction of protective immunity in a model system. J Parasitol. *78(3)*:447-453.

Abraham D, Lucius R, Trees AJ (2002): Immunity to Onchocerca spp. in animal hosts.Trends Parasitol. *18(4):*164-171.

Adam R, Kaltmann B, Rudin W, Friedrich T, Marti T, Lucius R. (1996): Identification of chitinase as the immunodominant filarial antigen recognized by sera of vaccinated rodents. J Biol Chem.*19;271(3):*1441-1447.

Adam, R., Drabner, B., and Lucius, R. (2001):in "Parasitic nematodes: Molecular biology, biochemistry and immunology, Edited by Kennedy, M. W. and Harnett, W." CABI Publishing, Wallingford, UK.

Arakane Y, Zhu Q, Matsumiya M, Muthukrishnan S, Kramer KJ. (2003): Properties of catalytic, linker and chitin-binding domains of insect chitinase. Insect Biochem Mol Biol. *33(6):*631-648.

Araujo AC, Souto-Padron T, de Souza W. (1993): Cytochemical localization of carbohydrate residues in microfilariae of Wuchereria bancrofti and Brugia malayi. J Histochem Cytochem. *41(4):*571-578.

Arnold K, Venegas A, Houseweart C, Fuhrman JA. (1996): Discrete transcripts encode multiple chitinase isoforms in Brugian microfilariae. Mol Biochem Parasitol. *80(2):*149-158.

Aslund F, Zheng M, Beckwith J, Storz G. (1999): Regulation of the OxyR transcription factor by hydrogen peroxide and the cellular thiol-disulfide status.Proc Natl Acad Sci U S A. *96(11):*6161-6165.

Barnes WM. (1994): PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. Proc Natl Acad Sci U S A. *91(6):*2216-2220.

Bej AK, Mahbubani MH, Atlas RM. (1991): Amplification of nucleic acids by polymerase chain reaction (PCR) and other methods and their applications. Crit Rev Biochem Mol Biol. *26(3-4):*301-334.

Bleiss W, Oberlander U, Hartmann S, Adam R, Marko A, Schonemeyer A, Lucius R. (2002): Protective immunity induced by irradiated third-stage larvae of the filaria Acanthocheilonema viteae is directed against challenge third-stage larvae before molting. J Parasitol. *88(2):*264-270.

Bortone K, Monzingo AF, Ernst S, Robertus JD. (2002): The structure of an allosamidin complex with the Coccidioides immitis chitinase defines a role for a second acid residue in substrate-assisted mechanism.J Mol Biol. *320(2):*293-302.

Boyer AE, Tsang VC, Eberhard ML, Zea-Flores G, Hightower A, Pilcher JB, Zea-Flores R, Zhou W, Reimer CB. (1991): Guatemalan human onchocerciasis. II. Evidence for IgG3 involvement in acquired immunity to Onchocerca volvulus and identification of possible immune-associated antigens. J Immunol. *146(11):*4001-4010.

Brattig NW, Tischendorf FW, Strote G, Medina-de la Garza CE. (1991): Eosinophil-larval-interaction in onchocerciasis: heterogeneity of in vitro adherence of eosinophils to infective third and fourth stage larvae and microfilariae of Onchocerca volvulus. Parasite Immunol. *13(1):*13-22.

Brydon LJ, Gooday GW, Chappell LH, King TP.(1987): Chitin in egg shells of Onchocerca gibsoni and Onchocerca volvulus.Mol Biochem Parasitol. *25(3):*267-272.

C. elegans Sequencing Consortium. (1999):Genome sequence of the nematode C. elegans: a platform for investigating biology. Science. Dec 11;282(5396):2012-2018. Review. Erratum in: Science 1999 Jan 1;283(5398):35. Science 1999 Mar 26;283(5410):2103. Science *3;285(5433):*1493.

Canlas M, Wadee A, Lamontagne L, Piessens WF.(1984): A monoclonal antibody to surface antigens on microfilariae of Brugia malayi reduces microfilaremia in infected jirds. Am J Trop Med Hyg. *33(3):*420-424.

Canlas MM, Piessens WF. (1984): Stage-specific and common antigens of Brugia malayi identified with monoclonal antibodies. J Immunol. *132(6):*3138-3141.

Carlow CK, Philipp M. (1987): Protective immunity to Brugia malayi larvae in BALB/c mice: potential of this model for the identification of protective antigens.Am J Trop Med Hyg. *37(3):*597-604.

Cheng S, Fockler C, Barnes WM, Higuchi R. (1994): Effective amplification of long targets from cloned inserts and human genomic DNA. Proc Natl Acad Sci U S A. *91(12):*5695-5699.

Cohen-Kupiec R, Chet I.(1998): The molecular biology of chitin digestion.Curr Opin Biotechnol. *9(3):*270-277.

Cook JA, Steel C, Ottesen EA. (2001): Towards a vaccine for onchocerciasis.Trends Parasitol. *17(12):*555-558.

da Costa AV, Lafitte S, Fontaine J, Bossus M, Gras-Masse H, Capron A, Grzych JM. (1999): Definition and mapping of epitopes recognized by specific monoclonal antibodies to Schistosoma bovis 28 kDa glutathione S-transferase: relation with anti-egg viability immunity. Parasite Immunol. *21(1):*9-18.

Davis GD, Elisee C, Newham DM, Harrison RG. (1999): New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol Bioeng. *20;65(4):*382-388.

Del Sal G, Manfioletti G, Schneider C. (1989): The CTAB-DNA precipitation method: a common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. Biotechniques. *7(5):*514-520.

Demanou M, Enyong P, Pion SD, Basanez MG, Boussinesq M. (2003): Experimental studies on the transmission of Onchocerca volvulus by its vector in the Sanaga valley (Cameroon): Simulium squamosum B. Intake of microfilariae and their migration to the haemocoel of the vector. Ann Trop Med Parasitol. *97(4):*381-402.

Derman AI, Prinz WA, Belin D, Beckwith J. (1993): Mutations that allow disulfide bond formation in the cytoplasm of Escherichia coli. Science.*10;262(5140):*1744-1747.

Devaney E, Howells RE, Smith G. (1985): Brugia pahangi in the BALB/C mouse: a model for testing filaricidal compounds. J Helminthol. *59(2):*95-99.

Drabner B, Reineke U, Schneider-Mergener J, Humphreys RE, Hartmann S, Lucius R. (2002): Identification of T helper cell-recognized epitopes in the chitinase of the filarial nematode Onchocerca volvulus. Vaccine. *20(31-32):*3685-3694.

Ellis DS, Rogers R, Bianco AE, Denham DA. (1978): Intrauterine development of the microfilariae of Dipetalonema viteae. J Helminthol. *52(1):*7-10.

Elson LH, Guderian RH, Araujo E, Bradley JE, Days A, Nutman TB.(1994): Immunity to onchocerciasis: identification of a putatively immune population in a hyperendemic area of Ecuador. J Infect Dis. *169(3):*588-594.

Felder H, Herzceg A, de Chastonay Y, Aeby P, Tobler H, Muller F. (1994): Tas, a retrotransposon from the parasitic nematode Ascaris lumbricoides. Gene. *149(2):*219-225.

Felse PA, Panda T. (1999): Regulation and cloning of microbial chitinase genes. Appl Microbiol Biotechnol. *51(2):*141-151.

Flach J, Pilet PE, Jolles P.(1992): What's new in chitinase research? Experientia. *48(8):*701-716.

Foster JM, Zhang Y, Kumar S, Carlow CK. (2005): Parasitic nematodes have two distinct chitin synthases. Mol Biochem Parasitol. *142(1):*126-132.

Freedman DO, Nutman TB, Ottesen EA. (1989): Protective immunity in bancroftian filariasis. Selective recognition of a 43-kD larval stage antigen by infection-free individuals in an endemic area. J Clin Invest. *83(1):*14-22.

Fuhrman JA, Lane WS, Smith RF, Piessens WF, Perler FB.(1992): Transmission-blocking antibodies recognize microfilarial chitinase in brugian lymphatic filariasis. Proc Natl Acad Sci U S A. *89(5):*1548-1552.

Fuhrman JA, Lee J, Dalamagas D. (1995): Structure and function of a family of chitinase isozymes from Brugian microfilariae. Exp Parasitol. *80(4):*672-680.

Fuhrman JA, Urioste SS, Hamill B, Spielman A, Piessens WF.(1987): Functional and antigenic maturation of Brugia malayi microfilariae. Am J Trop Med Hyg. *36(1):*70-74.

Fujii R, Morihara F, Fukushima K, Oku T, Hifumi E, Uda T. (2004): Recombinant antigen from Helicobacter pylori urease as vaccine against H. pylori-associated disease. Biotechnol Bioeng. *86(7):*737-746.

Fukamizo T.(2000): Chitinolytic enzymes: catalysis, substrate binding, and their application. Curr Protein Pept Sci. *1(1):*105-124.

Fusetti F, Pijning T, Kalk KH, Bos E, Dijkstra BW. (2003): Crystal structure and carbohydrate-binding properties of the human cartilage glycoprotein-39. J Biol Chem. *278(39):*37753-37760.

Geiger SM, Hoffmann W, Rapp J, Schulz-Key H, Eisenbeiss WF.(1996): Filariidae: cross-protection in filarial infections. Exp Parasitol. *83(3):*352-356.

Grenfell BT, Michael E, Denham DA. (1991): A model for the dynamics of human lymphatic filariasis. Parasitol Today. *7(11):*318-323.

Guiliano DB, Hall N, Jones SJ, Clark LN, Corton CH, Barrell BG, Blaxter ML. (2002): Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes. Genome Biol. *3(10):*RESEARCH0057.

Harris MT, Lai K, Arnold K, Martinez HF, Specht CA, Fuhrman JA. (2000): Chitin synthase in the filarial parasite, Brugia malayi. Mol Biochem Parasitol. *111(2):*351-362.

Harrison RA, Wu Y, Egerton G, Bianco AE. (1999): DNA immunisation with Onchocerca volvulus chitinase induces partial protection against challenge infection with L3 larvae in mice. Vaccine. *18(7-8):*647-655.

Henrissat B, Bairoch A (1993): New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. *293 (Pt 3):*781-788.

Henrissat B, Bairoch A.(1996): Updating the sequence-based classification of glycosyl hydrolases. Biochem J. *316 (Pt 2):*695-696.

Henrissat B.(1999): Classification of chitinases modules.EXS. *87:*137-156.

Hirota K, Nagata K, Norose Y, Futagami S, Nakagawa Y, Senpuku H, Kobayashi M, Takahashi H. (2001): Identification of an antigenic epitope in Helicobacter pylori urease that induces neutralizing antibody production. Infect Immun. *69(11):*6597-6603.

Hollis T, Monzingo AF, Bortone K, Ernst S, Cox R, Robertus JD. (2000): The X-ray structure of a chitinase from the pathogenic fungus Coccidioides immitis. Protein Sci. *9(3):*544-551.

Hollis T, Monzingo AF, Bortone K, Schelp E, Cox R, Robertus JD. (1998): Crystallization and preliminary X-ray analysis of a chitinase from the fungal pathogen Coccidioides immitis. Acta Crystallogr D Biol Crystallogr. *54(Pt 6 Pt 2):*1412-1413.

Honda Y, Kitaoka M, Hayashi K.(2004): Kinetic evidence related to substrate-assisted catalysis of family 18 chitinases.FEBS Lett. *567(2-3):*307-310.

Hougard JM, Yameogo L, Seketeli A, Boatin B, Dadzie KY.(1997): Twenty-two years of blackfly control in the onchocerciasis control programme in West Africa. Parasitol Today. *13(11):*425-431.

Inoue, H., Nojima, H and Okayama, H. (1990): High efficiency transformation of Escherichia coli with plasmids. Gene *96:* 23-28.

Jean, L., Long, M., Young, J., Pery, P. and Tomley, F. (2001): Aspartyl proteinase genes from apicomplexan parasites: evidence for evolution of the gene structure. Trends Parasitol. *17:* 491–498.

Johnson EH, Irvine M, Kass PH, Browne J, Abdullai M, Prince AM, Lustigman S. (1994): Onchocerca volvulus: in vitro cytotoxic effects of human neutrophils and serum on third-stage larvae. Trop Med Parasitol. *45(4):*331-335.

Johnson EH, Schynder-Candrian S, Rajan TV, Nelson FK, Lustigman S, Abraham D. (1998): Immune responses to third stage larvae of Onchocerca volvulus in interferon-gamma and interleukin-4 knockout mice. Parasite Immunol. *20(7):*319-324.

Kaushal NA, Simpson AJ, Hussain R, Ottesen EA. (1984): Brugia malayi: stage-specific expression of carbohydrates containing N-acetyl-D-glucosamine on the sheathed surfaces of microfilariae. Exp Parasitol. *58(2):*182-187.

Kramer KJ, Muthukrishnan S. (1997): Insect chitinases: molecular biology and potential use as biopesticides. Insect Biochem Mol Biol. *27(11):*887-900.

Lange AM, Yutanawiboonchai W, Lok JB, Trpis M, Abraham D. (1993): Induction of protective immunity against larval Onchocerca volvulus in a mouse model. Am J Trop Med Hyg. *49(6):*783-788.

Lange AM, Yutanawiboonchai W, Scott P, Abraham D. (1994): IL-4- and IL-5-dependent protective immunity to Onchocerca volvulus infective larvae in BALB/cBYJ mice. J Immunol. *153(1):*205-211.

Langer RC, Vinetz JM. (2001): Plasmodium ookinete-secreted chitinase and parasite penetration of the mosquito peritrophic matrix. Trends Parasitol. *17(6):*269-272.

Le Goff L, Loke P, Ali HF, Taylor DW, Allen JE. (2000): Interleukin-5 is essential for vaccine-mediated immunity but not innate resistance to a filarial parasite. Infect Immun. *68(5):*2513-2517.

Lightowlers MW, Colebrook AL, Gauci CG, Gauci SM, Kyngdon CT, Monkhouse JL, Vallejo Rodriquez C, Read AJ, Rolfe RA, Sato C. (2003): Vaccination against cestode parasites: anti-helminth vaccines that work and why. Vet Parasitol. *115(2):*83-123.

Lucius R, Textor G, Kern A, Kirsten C. (1991): Acanthocheilonema viteae: vaccination of jirds with irradiation-attenuated stage-3 larvae and with exported larval antigens. Exp Parasitol. *73(2):*184-196.

Lustigman S, James ER, Tawe W, Abraham D. (2002): Towards a recombinant antigen vaccine against Onchocerca volvulus. Trends Parasitol. *18(3):*135-141.

Maeda N, Smithies O. (1986): The evolution of multigene families: human haptoglobin genes. Annu Rev Genet. *20:* 81-108.

Neafsey DE, Hartl DL, Berriman M. (2005): Evolution of noncoding and silent coding sites in the Plasmodium falciparum and Plasmodium reichenowi genomes. Mol Biol Evol. *22(7):*1621-1626.

Nogami S, Hayashi Y, Tanaka M, Korenaga M, Tada I, Tanaka H. (1986): Antigenic similarity of Onchocerca volvulus to other helminths examined by monoclonal antibodies against O. volvulus. Jpn J Exp Med. *56(4):*177-183.

Nutman TB. (2002): Future directions for vaccine-related onchocerciasis research. Trends Parasitol. *18(6):*237-239.

Peralta ME, Schmitz KA, Rajan TV. (1999): Failure of highly immunogenic filarial proteins to provide host-protective immunity. Exp Parasitol. *91(4):*334-340.

Perrakis A, Tews I, Dauter Z, Oppenheim AB, Chet I, Wilson KS, Vorgias CE. (1994): Crystal structure of a bacterial chitinase at 2.3 A resolution. Structure. *2(12):*1169-1180.

Popovici C, Roubin R, Coulier F, Pontarotti P, Birnbaum D. (1999): The family of Caenorhabditis elegans tyrosine kinase receptors: similarities and differences with mammalian receptors. Genome Res. *9(11):*1026-1039.

Prince AM, Brotman B, Johnson EH Jr, Smith A, Pascual D, Lustigman S. (1992): Onchocerca volvulus: immunization of chimpanzees with X-irradiated third-stage (L3) larvae. Exp Parasitol. *74(3):*239-250.

Raghavan N, Freedman DO, Fitzgerald PC, Unnasch TR, Ottesen EA, Nutman TB. (1994): Cloning and characterization of a potentially protective chitinase-like recombinant antigen from Wuchereria bancrofti. Infect Immun. *62(5):*1901-1908.

Raina S, Missiakas D. (1997): Making and breaking disulfide bonds. Annu Rev Microbiol. *51:*179-202.

Ramalho-Ortigao JM, Traub-Cseko YM. (2003): Molecular characterization of Llchit1, a midgut chitinase cDNA from the leishmaniasis vector Lutzomyia longipalpis. Insect Biochem Mol Biol. *33(3):*279-287.

Ramos A, Mahowald A, Jacobs-Lorena M. (1994): Peritrophic matrix of the black fly Simulium vittatum: formation, structure, and analysis of its protein components. J Exp Zool. *268(4):*269-281.

Richards FO Jr, Boatin B, Sauerbrey M, Seketeli A (2001): Control of onchocerciasis today: status and challenges. Trends Parasitol. *17(12):*558-563.

Razek-Desouky A, Specht CA, Soong L, Vinetz JM. (2001): Leishmania donovani: expression and characterization of Escherichia coli-expressed recombinant chitinase LdCHT1. Exp Parasitol. *99(4):*220-225.

Rietsch A, Belin D, Martin N, Beckwith J. (1996): An in vivo pathway for disulfide bond isomerization in Escherichia coli. Proc Natl Acad Sci U S A. *93(23):*13048-13053.

Robertus JD, Monzingo AF.(1999): The structure and action of chitinases. EXS. *87:*125-135.

Rogers R, Ellis DS, Denham DA. (1976): Studies with Brugia pahangi. 14. Intrauterine development of the microfilaria and a comparison with other filarial species. J Helminthol. *50(4):*251-257.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. (1988): Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. *239(4839):*487-491.

Saint Andre A, Blackwell NM, Hall LR, Hoerauf A, Brattig NW, Volkmann L, Taylor MJ, Ford L, Hise AG, Lass JH, Diaconu E, Pearlman E. (2002): The role of endosymbiotic Wolbachia bacteria in the pathogenesis of river blindness. Science. *295(5561):*1892-1895.

Sakuda S, Isogai A, Matsumoto S, Suzuki A (1987): Search for microbial insect growth regulators. II. Allosamidin, a novel insect chitinase inhibitor. J Antibiot (Tokyo). *40(3):*296-300

Sambrook, J.; Fritsch, E. F. und Maniatis, T. (1989): Molecular cloning. A laboratory manual, Cold Spring Harbor Laboratory Press, New York.

Schein, C.H and Noteborn, MHM (1989): Bio/Technology *7:* 1141-1148

Seferian PG, Martinez ML. (2000): Immune stimulating activity of two new chitosan containing adjuvant formulations. Vaccine. *19(6):*661-668.

Senapathy P, Shapiro MB, Harris NL. (1990): Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol. *183:*252-278.

Short JM, Fernandez JM, Sorge JA, Huse WD. (1988): Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. Nucleic Acids Res. *16(15):*7583-7600.

Sone M, Akiyama Y, Ito K. (1997): Differential in vivo roles played by DsbA and DsbC in the formation of protein disulfide bonds. J Biol Chem. *272(16):*10349-10352.

Sonenshine, D.E (1991): Biology of Ticks. Oxford University Press, New York.

Soumbey-Alley E, Basanez MG, Bissan Y, Boatin BA, Remme JH, Nagelkerke NJ, de Vlas SJ, Borsboom GJ, Habbema JD. (2004): Uptake of Onchocerca volvulus (Nematoda: Onchocercidae) by Simulium (Diptera: Simuliidae) is not strongly dependent on the density of skin microfilariae in the human host. J Med Entomol. *41(1):*83-94.

Southworth MW, Fuhrman JA, Robbins PW, Beauregard K, Perler FB. (1996): Gene cloning and production of active recombinant Brugia malayi microfilarial chitinase.Gene. *177(1-2):*55-58.

Stewart GR, Elson L, Araujo E, Guderian R, Nutman TB, Bradley JE. (1995): Isotype-specific characterization of antibody responses to Onchocerca volvulus in putatively immune individuals. Parasite Immunol. *17(7):*371-380.

Storey DM, Al-Mukhtar AS. (1982): The development of normal and gamma-irradiated Litomosoides carinii larvae in cotton rats. Tropenmed Parasitol. *33(4):*223-226.

Synstad B, Gaseidnes S, Van Aalten DM, Vriend G, Nielsen JE, Eijsink VG. (2004): Mutational and computational analysis of the role of conserved residues in the active site of a family 18 chitinase. Eur J Biochem. *271(2):*253-262.

Taylor MJ, van Es RP, Shay K, Folkard SG, Townson S, Bianco AE. (1994): Protective immunity against Onchocerca volvulus and O. lienalis infective larvae in mice. Trop Med Parasitol. *45(1):*17-23.

Terwisscha van Scheltinga AC, Kalk KH, Beintema JJ, Dijkstra BW. (1994): Crystal structures of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. Structure. *2(12):*1181-1189.

Towbin H, Staehelin T, Gordon J. (1979): Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A. *76(9):*4350-4354.

Townson S, Bianco AE. (1982): Experimental infection of mice with the microfilariae of Onchocerca lienalis. Parasitology. *85 (Pt 2):*283-293.

Trees AJ, Graham SP, Renz A, Bianco AE, Tanya V. (2000): Onchocerca ochengi infections in cattle as a model for human onchocerciasis: recent developments. Parasitology. *120 Suppl:*S133-142.

Turaga PS, Tierney TJ, Bennett KE, McCarthy MC, Simonek SC, Enyong PA, Moukatte DW, Lustigman S. (2000): Immunity to onchocerciasis: cells from putatively immune individuals produce enhanced levels of interleukin-5, gamma interferon, and granulocyte-macrophage colony-stimulating factor in response to Onchocerca volvulus larval and male worm antigens. Infect Immun. *68(4):*1905-1911.

van Aalten DM, Komander D, Synstad B, Gaseidnes S, Peter MG, Eijsink VG. (2001): Structural insights into the catalytic mechanism of a family 18 exo-chitinase. Proc Natl Acad Sci U S A. *98(16):*8979-8984.

Venegas A, Goldstein JC, Beauregard K, Oles A, Abdulhayoglu N, Fuhrman JA. (1996): Expression of recombinant microfilarial chitinase and analysis of domain function. Mol Biochem Parasitol. *78(1-2):*149-159.

Vialard, J.E., Arif, B.M., and Richardson, C.D. (1995) in "Methods in Molecular Biology, Vol. 39: Baculovirus Expression Protocols," Humana Press Inc., Totowa, NJ.

Vinetz JM, Dave SK, Specht CA, Brameld KA, Xu B, Hayward R, Fidock DA. (1999): The chitinase PfCHT1 from the human malaria parasite Plasmodium falciparum lacks proenzyme and chitin-binding domains and displays unique substrate preferences. Proc Natl Acad Sci U S A. *96(24):*14061-14066.

Watanabe T, Kobori K, Miyashita K, Fujii T, Sakai H, Uchida M, Tanaka H. (1993): Identification of glutamic acid 204 and aspartic acid 200 in chitinase A1 of Bacillus circulans WL-12 as essential residues for chitinase activity. J Biol Chem. *268(25):*18567-18572.

Welburn SC, Arnold K, Maudlin I, Gooday GW. (1993): Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. Parasitology. *107 (Pt 2):*141-145.

Wu Y, Adam R, Williams SA, Bianco AE. (1996): Chitinase genes expressed by infective larvae of the filarial nematodes, Acanthocheilonema viteae and Onchocerca volvulus. Mol Biochem Parasitol. *75(2):*207-219.

Wu Y, Egerton G, Underwood AP, Sakuda S, Bianco AE. (2001): Expression and secretion of a larval-specific chitinase (family 18 glycosyl hydrolase) by the infective stages of the parasitic nematode, Onchocerca volvulus. J Biol Chem. *276(45):*42557-42564.

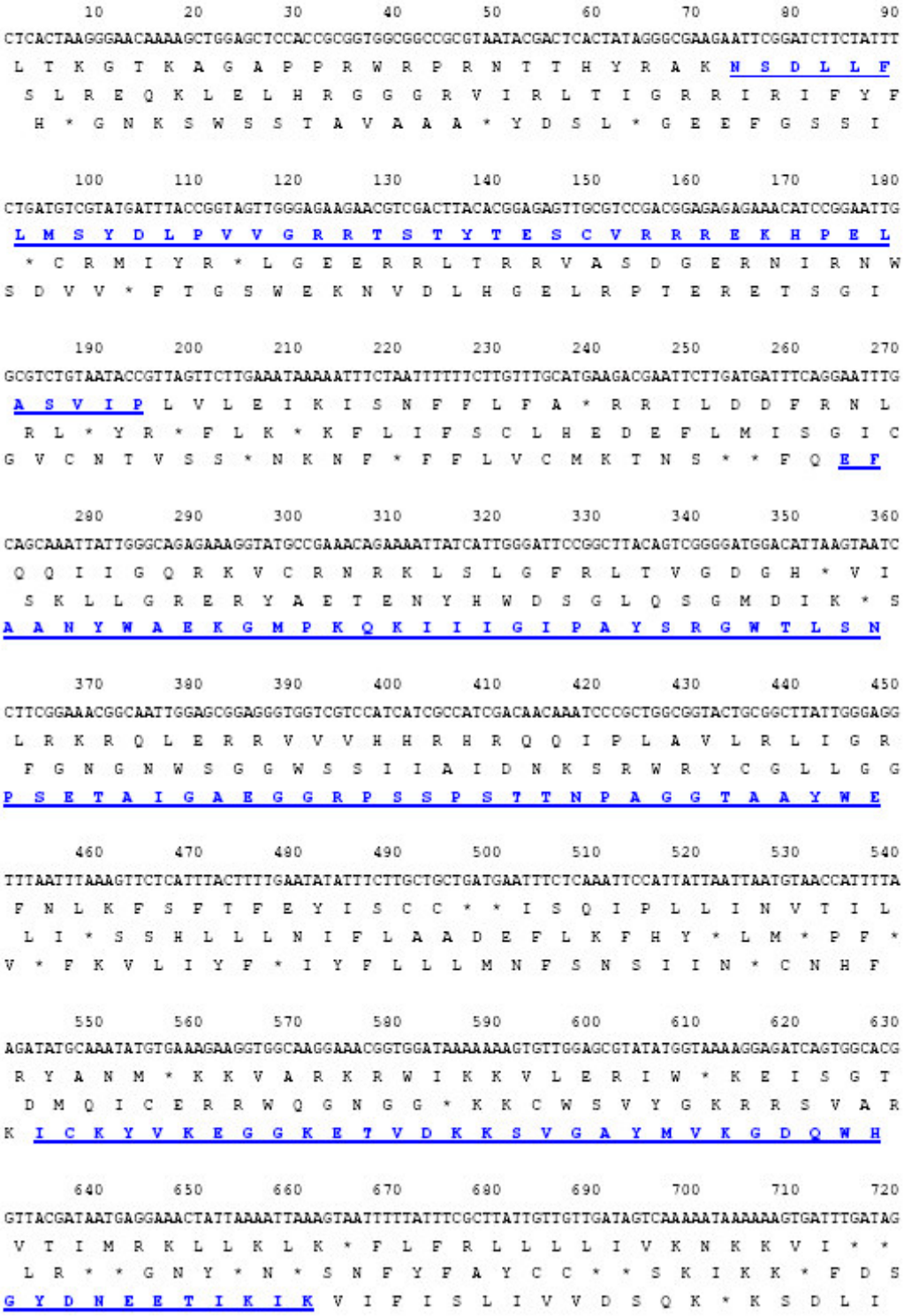
.

8 Appendix

Appendix 1: Insert of clone 1;Sequence Range: 1 to 13963

Gene II: 73...1797 (3' end)

Gene III: 5280...11878



```
      730       740       750       760       770       780       790       800       810
TCAAAAAATAAGAAAAGATAATAAGATAGTAGTCGAAAAATAAAAAATTAAAAAATAACGTTGCCAATATATCTTGCCTTAACTTTAATT
 S  K  N  K  K  R  *  *  D  S  S  R  K  I  K  N  *  K  I  T  L  P  I  Y  L  A  L  T  L  I
  Q  K  I  R  K  D  N  K  I  V  V  E  K  *  K  I  K  K  *  R  C  Q  Y  I  L  P  *  L  *  F
V  K  K  *  E  K  I  I  R  *  *  S  K  N  K  K  L  K  N  N  V  A  N  I  S  C  L  N  F  N

      820       830       840       850       860       870       880       890       900
TAAAGAGCACCATAAATTTTATAATGTTTAATAAGAAGAAATTAATTTCGGAGTTTACTCGTGTAATTGTTTTACTTTAGAATTAAAATT
 *  R  A  P  *  I  L  *  C  L  I  R  R  N  *  F  R  S  L  L  V  *  L  F  Y  F  R  I  K  I
  K  E  H  H  K  F  Y  N  V  *  *  E  E  I  N  F  G  V  Y  S  C  N  C  F  T  L  E  L  K  F
L  K  S  T  I  N  F  I  M  F  N  K  K  K  L  I  S  E  F  T  R  V  I  V  L  L  *  N  *  N

      910       920       930       940       950       960       970       980       990
TTCAGATGAAATGGCTGAAAGAGAAAGGTTATGGCGGTGCCTTCATATGGGCTCTTGATTTTGATGATTTTAAGGGTACAAGTTGTGGCA
 F  R  *  N  G  *  K  R  K  V  M  A  V  P  S  Y  G  L  L  I  L  M  I  L  R  V  Q  V  V  A
  S  D  E  M  A  E  R  E  R  L  W  R  C  L  H  M  G  S  *  F  *  *  F  *  G  Y  K  L  W  Q
F  Q  M  K  W  L  K  E  K  G  Y  G  G  A  F  I  W  A  L  D  F  D  D  F  K  G  T  S  C  G

     1000      1010      1020      1030      1040      1050      1060      1070      1080
AAGGTCCCTATCCACTGTTAAATGCTCTTAATAATGAGCTTCGAGACGGCTTCGTATTTTGGGTGAAGGATTTTACAACTTCGTAATGAT
 K  V  P  I  H  C  *  M  L  L  I  M  S  F  E  T  A  S  Y  F  G  *  R  I  L  Q  L  R  N  D
  R  S  L  S  T  V  K  C  S  *  *  *  A  S  R  R  L  R  I  L  G  E  G  F  Y  N  F  V  M  I
K  G  P  Y  P  L  L  N  A  L  N  N  E  L  R  D  G  F  V  F  W  V  K  D  F  T  T  S  *  *

     1090      1100      1110      1120      1130      1140      1150      1160      1170
TTGTTGCATATCATAAAAATGTTATAATCGGTCGGAACTTAAAGGATTGAAATTTAATCCGGAAATCCTAAAATTCATAATGTAATGCTG
 L  L  H  I  I  K  M  L  *  S  V  G  T  *  R  I  E  I  *  S  G  N  P  K  I  H  N  V  M  L
  C  C  I  S  *  K  C  Y  N  R  S  E  L  K  G  L  K  F  N  P  E  I  L  K  F  I  M  *  C  C
F  V  A  Y  H  K  N  V  I  I  G  R  N  L  K  D  *  N  L  I  R  K  S  *  N  S  *  C  N  A

     1180      1190      1200      1210      1220      1230      1240      1250      1260
TATTTTAAACAGTATTGCAAACGCAACTGGTGCATTTTTGGATATTCTTTGCTTTCTTACAAATGGTCCGCAAAAAATGAATCACGACAA
 Y  F  K  Q  Y  C  K  R  N  W  C  I  F  G  Y  S  L  L  S  Y  K  W  S  A  K  N  E  S  R  Q
  I  L  N  S  I  A  N  A  T  G  A  F  L  D  I  L  C  F  L  T  N  G  P  Q  K  M  N  H  D  K
V  F  *  T  V  L  Q  T  Q  L  V  H  F  W  I  F  F  A  F  L  Q  M  V  R  K  K  *  I  T  T

     1270      1280      1290      1300      1310      1320      1330      1340      1350
GCCGAAATGTCTATTCAATTAAATTTTTATATTAATTTCGTTCTATACAGATTTTAGATTTAAAAATTAATAAAGCAATCATTTTTAGTC
 A  E  M  S  I  Q  L  N  F  Y  I  N  F  V  L  Y  R  F  *  I  *  K  L  I  K  Q  S  F  L  V
  P  K  C  L  F  N  *  I  F  I  L  I  S  F  Y  T  D  F  R  F  K  N  *  *  S  N  H  F  *  S
S  R  N  V  Y  S  I  K  F  L  Y  *  F  R  S  I  Q  I  L  D  L  K  I  N  K  A  I  I  F  S

     1360      1370      1380      1390      1400      1410      1420      1430      1440
ATCAATGTTGGCAAAATCAAATAATGAAATATTGAGGAAAACAGGGAGAGCCTCAGGAAGTAGTGAAATTGAAACGATCAAAAGTGGTTT
 I  N  V  G  K  I  K  *  *  N  I  E  E  N  R  E  S  L  R  K  *  *  N  *  N  D  Q  K  W  F
  S  M  L  A  K  S  N  N  E  I  L  R  K  T  G  R  A  S  G  S  S  E  I  E  T  I  K  S  G  F
H  Q  C  W  Q  N  Q  I  M  K  Y  *  G  K  Q  G  E  P  Q  E  V  V  K  L  K  R  S  K  V  V

     1450      1460      1470      1480      1490      1500      1510      1520      1530
TGAAACAGTACGAAGAACTTCTGGTTTGTTATCCGCAAGTTAAAATGTCGTTCTTGATCTGTTCTTTTGCTTTAAACACAGCGCAACTTT
 *  N  S  T  K  N  F  W  F  V  I  R  K  L  K  C  R  S  *  S  V  L  L  L  *  T  Q  R  N  F
  E  T  V  R  R  T  S  G  L  L  S  A  S  *  N  V  V  L  D  L  F  F  C  F  K  H  S  A  T  F
L  K  Q  Y  E  E  L  L  V  C  Y  P  Q  V  K  M  S  F  L  I  C  S  F  A  L  N  T  A  Q  L

     1540      1550      1560      1570      1580      1590      1600      1610      1620
TCGAGTTTGCTCTTCATGCATTTTGGATAGAATGGAAATGATTGTGGATAGATGAAAAGAAATTTAGGAAAATCAATGTTCTGCTGTAAC
 S  S  L  L  F  M  H  F  G  *  N  G  N  D  C  G  *  M  K  R  N  L  G  K  S  M  F  C  C  N
  R  V  C  S  S  C  I  L  D  R  M  E  M  I  V  D  R  *  K  E  I  *  E  N  Q  C  S  A  V  T
F  E  F  A  L  H  A  F  W  I  E  W  K  *  L  W  I  D  E  K  K  F  R  K  I  N  V  L  L  *
```

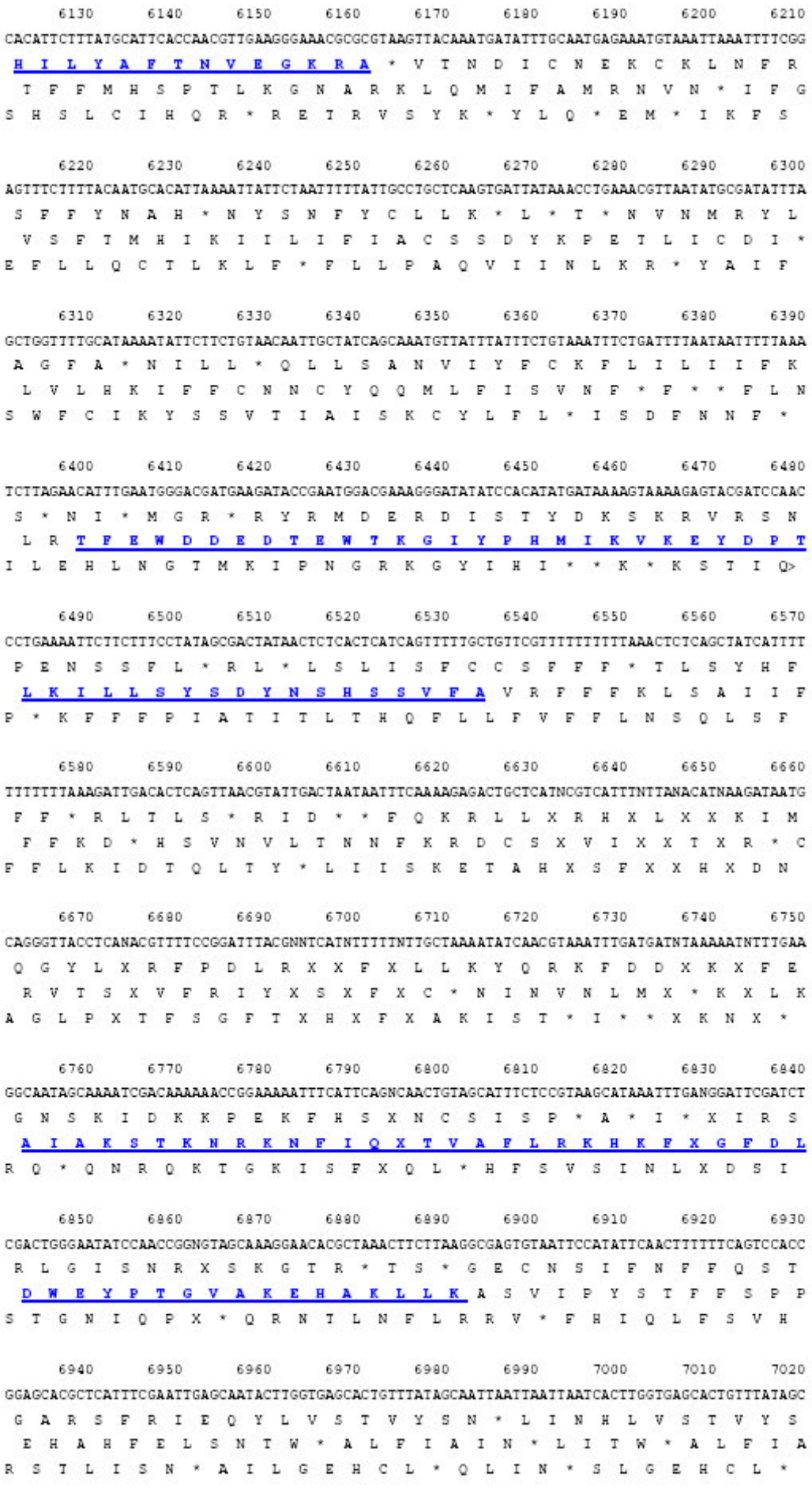
```
      1630       1640       1650       1660       1670       1680       1690       1700       1710
ATTTCAGATATAAAATGTTTGGAGTCATATGGTCTGTTCCCGCATCCTGACGATTGCCATTTATTTATTCATTGCGCAAATGACTATCCA
  I  S  D  I  K  C  L  E  S  Y  G  L  F  P  H  P  D  D  C  H  L  F  I  H  C  A  N  D  Y  P
   F  Q  I  *  N  V  W  S  H  M  V  C  S  R  I  L  T  I  A  I  Y  L  F  I  A  Q  M  T  I  H
H  F  R  Y  K  M  F  G  V  I  W  S  V  P  A  S  *  R  L  P  F  I  Y  S  L  R  K  *  L  S

      1720       1730       1740       1750       1760       1770       1780       1790       1800
TACGTGAAACAATGTCCGCTACATACCTTTTTTAATGATGATATCAAAATTTGCGATCACTTAGTAAATGCGCCGATCACATGCAAATGA
  Y  V  K  Q  C  P  L  H  T  F  F  N  D  D  I  K  I  C  D  H  L  V  N  A  P  I  T  C  K  *
   T  *  N  N  V  R  Y  I  P  F  L  M  M  I  S  K  F  A  I  T  *  *  M  R  R  S  H  A  N  D
I  R  E  T  M  S  A  T  Y  L  F  *  *  *  Y  Q  N  L  R  S  L  S  K  C  A  D  H  M  Q  M

      1810       1820       1830       1840       1850       1860       1870       1880       1890
TTTTCATATTACCAAGAATAAAAGTTTAATGTAAAAGTTGTGGCAGTCAAATGAATTGACCATACTCCTGTAATAATTCATATTATATAG
  F  S  Y  Y  Q  E  *  K  F  N  V  K  V  V  A  V  K  *  I  D  H  T  P  V  I  I  H  I  I  *
   F  H  I  T  K  N  K  S  L  M  *  K  L  W  Q  S  N  E  L  T  I  L  L  *  *  F  I  L  Y  R
I  F  I  L  P  R  I  K  V  *  C  K  S  C  G  S  Q  M  N  *  P  Y  S  C  N  N  S  Y  Y  I

      1900       1910       1920       1930       1940       1950       1960       1970       1980
GAGAATAAATATGACTATATGAGGACAAATGTCAAAAATGAGTCTAACGGAATTATTAAAAATGAAAAAATTTGTTATTATCTGTATTTT
  E  N  K  Y  D  Y  M  R  T  N  V  K  N  E  S  N  G  I  I  K  N  E  K  I  C  Y  Y  L  Y  F
   R  I  N  M  T  I  *  G  Q  M  S  K  M  S  L  T  E  L  L  K  M  K  K  F  V  I  I  C  I  L
G  E  *  I  *  L  Y  E  D  K  C  Q  K  *  V  *  R  N  Y  *  K  *  K  N  L  L  L  S  V  F

      1990       2000       2010       2020       2030       2040       2050       2060       2070
ATTTTGCGTAGCAAAAATCTTTCAACAGTTATCACCCATAACTTAGTAGAAAACATATGAATGATTCGGAAGATGAATAAATCACATGAA
  I  L  R  S  K  N  L  S  T  V  I  T  H  N  L  V  E  N  I  *  M  I  R  K  M  N  K  S  H  E
   F  C  V  A  K  I  F  Q  Q  L  S  P  I  T  *  *  K  T  Y  E  *  F  G  R  *  I  N  H  M  K
Y  F  A  *  Q  K  S  F  N  S  Y  H  P  *  L  S  R  K  H  M  N  D  S  E  D  E  *  I  T  *

      2080       2090       2100       2110       2120       2130       2140       2150       2160
AAGGATGAAATTCTGATAAAAAATAAGTACTCAATAAAAAAAGGATGAAATCATTATAAAAAAGGAAATTCCGATAAAAAATAAATACTCAA
  K  D  E  I  L  I  K  N  K  Y  S  I  K  K  D  E  I  I  I  K  R  K  F  R  *  K  I  N  T  Q
   R  M  K  F  *  *  K  I  S  T  Q  *  K  R  M  K  S  L  *  K  G  N  S  D  K  K  *  I  L  K
K  G  *  N  S  D  K  K  *  V  L  N  K  K  G  *  N  H  Y  K  K  E  I  P  I  K  N  K  Y  S

      2170       2180       2190       2200       2210       2220       2230       2240       2250
AGATAAAAATACTGTTCCATTGCATCAGGACTGGAGTATTTCCATCAAGAATCTTTTATTCCGTTTCCGCATTTTCACTGCTGCATAATT
  R  *  K  Y  C  S  I  A  S  G  L  E  Y  F  H  Q  E  S  F  I  P  F  P  H  F  H  C  C  I  I
   D  K  N  T  V  P  L  H  Q  D  W  S  I  S  I  K  N  L  L  F  R  F  R  I  F  T  A  A  *  F
K  I  K  I  L  F  H  C  I  R  T  G  V  F  P  S  R  I  F  Y  S  V  S  A  F  S  L  L  H  N

      2260       2270       2280       2290       2300       2310       2320       2330       2340
CCTTCCTTTTCTTTTGGCATTCCTTTTGTTTCCTTTACGACAATTCAGCTTAATTTTGATAAGAGAGTTTGTAGTTGGCTAATGTGATGA
  P  S  F  S  F  G  I  P  F  V  S  F  T  T  I  Q  L  N  F  D  K  R  V  C  S  W  L  M  *  *
   L  P  F  L  L  A  F  L  L  F  P  L  R  Q  F  S  L  I  L  I  R  E  F  V  V  G  *  C  D  D
S  F  L  F  F  W  H  S  F  C  F  L  Y  D  N  S  A  *  F  *  *  E  S  L  *  L  A  N  V  M

      2350       2360       2370       2380       2390       2400       2410       2420       2430
CACTCAATTCGAAAAAGGAAACAGTCAAAGGCGAATTTCAATTATTTTCAACAATTCTCATCATTGAATGATATGTGCGATAGAAGGAAT
  H  S  I  R  K  R  K  Q  S  K  A  N  F  N  Y  F  Q  Q  F  S  S  L  N  D  M  C  D  R  R  N
   T  Q  F  E  K  G  N  S  Q  R  R  I  S  I  I  F  N  N  S  H  H  *  M  I  C  A  I  E  G  I
T  L  N  S  K  K  E  T  V  K  G  E  F  Q  L  F  S  T  I  L  I  I  E  *  Y  V  R  *  K  E

      2440       2450       2460       2470       2480       2490       2500       2510       2520
TTTCATAATCTCAGATGAAGAAAAGTATTTTGCAGTTGCGCTTCATCTTTCTCTTCCATATTATGATAAAAGATATAATAAAAACTTCCA
  F  H  N  L  R  *  R  K  V  F  C  S  C  A  S  S  F  S  S  I  L  *  *  K  I  *  *  K  L  P
   F  I  I  S  D  E  E  K  Y  F  A  V  A  L  H  L  S  L  P  Y  Y  D  K  R  Y  N  K  N  F  Q
F  S  *  S  Q  M  K  K  S  I  L  Q  L  R  F  I  F  L  F  H  I  M  I  K  D  I  I  K  T  S
```

```
      2530       2540       2550       2560       2570       2580       2590       2600       2610
ATAATAAACGTGGCAGCATTATTTTTTCTAGCTTTCTGCTTATTTTTAATGTTTTTCCAGCTTTCTAGCTATTTCTTGTATTTTCCCAAT
 I  I  N  V  A  A  L  F  F  L  A  F  C  L  F  L  M  F  F  Q  L  S  S  Y  F  L  Y  F  P  N
  *  *  T  W  Q  H  Y  F  F  *  L  S  A  Y  F  *  C  F  S  S  F  L  A  I  S  C  I  F  P  I
N  N  K  R  G  S  I  I  F  S  S  F  L  L  I  F  N  V  F  P  A  F  *  L  F  L  V  F  S  Q

      2620       2630       2640       2650       2660       2670       2680       2690       2700
TTCCTAGTGTTTTTCCAGCTTTTGTAGTTGTTTTTCTAACTCCTTAATACTATCTACGTCAGCCTAACTTCGGGATTTTAATTAACTAAT
 F  L  V  F  F  Q  L  L  *  L  F  F  *  L  L  N  T  I  Y  V  S  L  T  S  G  F  *  L  T  N
  S  *  C  F  S  S  F  C  S  C  F  S  N  S  L  I  L  S  T  S  A  *  L  R  D  F  N  *  L  I
F  P  S  V  F  P  A  F  V  V  V  F  L  T  P  *  Y  Y  L  R  Q  P  N  F  G  I  L  I  N  *

      2710       2720       2730       2740       2750       2760       2770       2780       2790
TTGGTTAATCAAATATCTTCGAATCAAATACTTTGTAGAGCGCTTATTTTTGACACTCTTAATGTTCTTAAAAAAAACCTTACTTCATTG
 L  V  N  Q  I  S  S  N  Q  I  L  C  R  A  L  I  F  D  T  L  N  V  L  K  K  N  L  T  S  L
  W  L  I  K  Y  L  R  I  K  Y  F  V  E  R  L  F  L  T  L  L  M  F  L  K  K  T  L  L  H  C
F  G  *  S  N  I  F  E  S  N  T  L  *  S  A  Y  F  *  H  S  *  C  S  *  K  K  P  Y  F  I

      2800       2810       2820       2830       2840       2850       2860       2870       2880
TTATGTTTTTATGCTTAACATTTATTAATAACTGCAGAAATTATTTTTTTTTTTTATTTTGGATGTGTTCGCTGAGAACTTTGAAAAA
 L  C  F  Y  A  *  H  L  L  I  T  A  E  I  I  F  F  F  Y  F  W  M  C  S  L  R  T  F  E  K
  Y  V  F  M  L  N  I  Y  *  *  L  Q  K  L  F  F  F  F  I  F  G  C  V  R  *  E  L  L  K  K
V  M  F  L  C  L  T  F  I  N  N  C  R  N  Y  F  F  F  L  F  L  D  V  F  A  E  N  F  *  K

      2890       2900       2910       2920       2930       2940       2950       2960       2970
ATTTGTTGTTATGATTTTGATTTTTCACCGTTTTTCATTTTTTTCTCTTATATCTTATTGTGTTTTCCAGCTCTTTCGCTTTACCTAATT
 I  C  C  Y  D  F  D  F  S  P  F  F  I  F  F  S  Y  I  L  L  C  F  P  A  L  S  L  Y  L  I
  F  V  V  M  I  L  I  F  H  R  F  S  F  F  S  L  I  S  Y  C  V  F  Q  L  F  R  F  T  *  L
N  L  L  L  *  F  *  F  F  T  V  F  H  F  F  L  L  Y  L  I  V  F  S  S  S  F  A  L  P  N

      2980       2990       3000       3010       3020       3030       3040       3050       3060
GCAAGATTTTAAAATATCTGACGATTTACAAAATGTGCAACATTATTAATTCTTTATTGTTAATATTGTTTGCATTTTATTTTGCAACTT
 A  R  F  *  N  I  *  R  F  T  K  C  A  T  L  L  I  L  Y  C  *  Y  C  L  H  F  I  L  Q  L
  Q  D  F  K  I  S  D  D  L  Q  N  V  Q  H  Y  *  F  F  I  V  N  I  V  C  I  L  F  C  N  F
C  K  I  L  K  Y  L  T  I  Y  K  M  C  N  I  I  N  S  L  L  L  I  L  F  A  F  Y  F  A  T

      3070       3080       3090       3100       3110       3120       3130       3140       3150
CCTGATTTTATTTCCAAGCTATGCAAACATTTCGTATCTATAAAATCTGTTTCAAACTTAAGCCTACCAATGTTTTTTACAGTAATCCAT
 P  D  F  I  S  K  L  C  K  H  F  V  S  I  K  S  V  S  N  L  S  L  P  M  F  F  T  V  I  H
  L  I  L  F  P  S  Y  A  N  I  S  Y  L  *  N  L  F  Q  T  *  A  Y  Q  C  F  L  Q  *  S  I
S  *  F  Y  F  Q  A  M  Q  T  F  R  I  Y  K  I  C  F  K  L  K  P  T  N  V  F  Y  S  N  P

      3160       3170       3180       3190       3200       3210       3220       3230       3240
TTTTGATTCCCATATAACTCACCCCAATTTCATGTATCTACCGACTTTTATTTATATTTTCATTCATTCGTCCAATTTCACTTTTATTCC
 F  *  F  P  Y  N  S  P  Q  F  H  V  S  T  D  F  Y  L  Y  F  H  S  F  V  Q  F  H  F  Y  S
  F  D  S  H  I  T  H  P  N  F  M  Y  L  P  T  F  I  Y  I  F  I  H  S  S  N  F  T  F  I  P
F  L  I  P  I  *  L  T  P  I  S  C  I  Y  R  L  L  F  I  F  S  F  I  R  P  I  S  L  L  F

      3250       3260       3270       3280       3290       3300       3310       3320       3330
AATTTGAGCATATAAATATGAATAAAATTAAGTATTTTGTATGATGATATTAAGATATGAAAAATTTCCAGCTTATATTTAACTTTTGAT
 N  L  S  I  *  I  *  I  K  L  S  I  L  Y  D  D  I  K  I  *  K  I  S  S  L  Y  L  T  F  D
  I  *  A  Y  K  Y  E  *  N  *  V  F  C  M  M  I  L  R  Y  E  K  F  P  A  Y  I  *  L  L  I
Q  F  E  H  I  N  M  N  K  I  K  Y  F  V  *  *  Y  *  D  M  K  N  F  Q  L  I  F  N  F  *

      3340       3350       3360       3370       3380       3390       3400       3410       3420
ATGCACAGTATAATTCATTTTCTGATTTCCGCACAAAGATTGAAGATGTGAAAATATTTGACCTTTACCAAAATGTGGAGCATTATTATT
 M  H  S  I  I  H  F  L  I  S  A  Q  R  L  K  M  *  K  Y  L  T  F  T  K  M  W  S  I  I  I
  C  T  V  *  F  I  F  *  F  P  H  K  D  *  R  C  E  N  I  *  P  L  P  K  C  G  A  L  L  F
Y  A  Q  Y  N  S  F  S  D  F  R  T  K  I  E  D  V  K  I  F  D  L  Y  Q  N  V  E  H  Y  Y
```

```
      3430       3440       3450       3460       3470       3480       3490       3500       3510
CATATTAATTTTTTCAGTATAATAATGCACTGTCCCATTTATTGTCAACAAAAAAAAAAAAATCCCAAACGCTGAGCTGCTTTTAATAATA
 H  I  N  F  F  S  I  I  M  H  C  P  I  Y  C  Q  Q  K  K  K  I  P  N  A  E  L  L  L  I  I
  I  L  I  F  S  V  *  *  C  T  V  P  F  I  V  N  K  K  K  K  S  Q  T  L  S  C  F  *  *  Y
S  Y  *  F  F  Q  Y  N  N  A  L  S  H  L  L  S  T  K  K  K  N  P  K  R  *  A  A  F  N  N

      3520       3530       3540       3550       3560       3570       3580       3590       3600
TAATTTTAATAATAATTTCATATATTTCCATTAATTGATTATTTTATTATATTCCTTAAACATAGCAGCGTTGTCGTCTTCTTCTCTGAC
 *  F  *  *  *  F  H  I  F  P  L  I  D  Y  F  I  I  F  L  K  H  S  S  V  V  V  F  F  S  D
  N  F  N  N  N  F  I  Y  F  H  *  L  I  I  L  L  Y  S  L  N  I  A  A  L  S  S  S  S  L  T
I  I  L  I  I  I  S  Y  I  S  I  N  *  L  F  Y  Y  I  P  *  T  *  Q  R  C  R  L  L  L  *

      3610       3620       3630       3640       3650       3660       3670       3680       3690
CGTGATAATATTTACGGGTTTGATTGCAATCTTTCAGTTACTGTGAACTTCAAACTTTGGTAATTAAAGCCATAGCAATGATAGTAAAAA
 R  D  N  I  Y  G  F  D  C  N  L  S  V  T  V  N  F  K  L  W  *  L  K  P  *  Q  *  *  *  K
  V  I  I  F  T  G  L  I  A  I  F  Q  L  L  *  T  S  N  F  G  N  *  S  H  S  N  D  S  K  N
P  *  *  Y  L  R  V  *  L  Q  S  F  S  Y  C  E  L  Q  T  L  V  I  K  A  I  A  M  I  V  K

      3700       3710       3720       3730       3740       3750       3760       3770       3780
CACAATATTAAAACAAGAAACANGCGGTGCAAAGAATAAATCAAAACACATAACCTCAACGGAAGAATAAGTAGAAAGAATGAAGAGATG
 H  N  I  K  T  R  N  X  R  C  K  E  *  I  K  T  H  N  L  N  G  R  I  S  R  K  N  E  E  M
  T  I  L  K  Q  E  T  X  G  A  K  N  K  S  K  H  I  T  S  T  E  E  *  V  E  R  M  K  R  *
T  Q  Y  *  N  K  K  X  A  V  Q  R  I  N  Q  N  T  *  P  Q  R  K  N  K  *  K  E  *  R  D

      3790       3800       3810       3820       3830       3840       3850       3860       3870
AATGATTTGATGTCTTATATAGGCAAAGTATACGTGCATATCATAACGCGGGAAAATAAGGAAAGATATTGGAGGATATAAGAGCAAAGG
 N  D  L  M  S  Y  I  G  K  V  Y  V  H  I  I  T  R  E  N  K  E  R  Y  W  R  I  *  E  Q  R
  M  I  *  C  L  I  *  A  K  Y  T  C  I  S  *  R  G  K  I  R  K  D  I  G  G  Y  K  S  K  G
E  *  F  D  V  L  Y  R  Q  S  I  R  A  Y  H  N  A  G  K  *  G  K  I  L  E  D  I  R  A  K

      3880       3890       3900       3910       3920       3930       3940       3950       3960
AGATCATGTAGTACAATAATTGGAGGCAAAGATGGGAACGGATGCGAGTATAGAATGGATCATTTGAATGTGTTATTTAGTTTAGACAGG
 R  S  C  S  T  I  I  G  G  K  D  G  N  G  C  E  Y  R  M  D  H  L  N  V  L  F  S  L  D  R
  D  H  V  V  Q  *  L  E  A  K  M  G  T  D  A  S  I  E  W  I  I  *  M  C  Y  L  V  *  T  G
E  I  M  *  Y  N  N  W  R  Q  R  W  E  R  M  R  V  *  N  G  S  F  E  C  V  I  *  F  R  Q

      3970       3980       3990       4000       4010       4020       4030       4040       4050
AACATCGACATTTGACACGTATTAACATGCAGTAATATTCATTATTACAATTATGTCTATGAATAAAATATATATATATATATATATATC
 N  I  D  I  *  H  V  L  T  C  S  N  I  H  Y  Y  N  Y  V  Y  E  *  N  I  Y  I  Y  I  Y  I
  T  S  T  F  D  T  Y  *  H  A  V  I  F  I  I  T  I  M  S  M  N  K  I  Y  I  Y  I  Y  I  S
E  H  R  H  L  T  R  I  N  M  Q  *  Y  S  L  L  Q  L  C  L  *  I  K  Y  I  Y  I  Y  I  Y

      4060       4070       4080       4090       4100       4110       4120       4130       4140
TTTCGCGACAAAAGGAATTGCTTTTCTCTGAAATTTTAATCCAAAACAAAAAAATTAATAAAAATTAATTCTTCTTTGGTACTTCTCGCT
 F  R  D  K  R  N  C  F  S  L  K  F  *  S  K  T  K  K  L  I  K  I  N  S  S  L  V  L  L  A
  F  A  T  K  G  I  A  F  L  *  N  F  N  P  K  Q  K  N  *  *  K  L  I  L  L  W  Y  F  S  L
L  S  R  Q  K  E  L  L  F  S  E  I  L  I  Q  N  K  K  I  N  K  N  *  F  F  F  G  I  S  R

      4150       4160       4170       4180       4190       4200       4210       4220       4230
TTTCACGTATCTTCCAATTAGTGAAGCCATGAATGAAAACATTACCAACGCAAACGGTAAGTTGTCCAGTAAAGTGACAATTTTGGTTGG
 F  H  V  S  S  N  *  *  S  H  E  *  K  H  Y  Q  R  K  R  *  V  V  Q  *  S  D  N  F  G  W
  F  T  Y  L  P  I  S  E  A  M  N  E  N  I  T  N  A  N  G  K  L  S  S  K  V  T  I  L  V  G
F  S  R  I  F  Q  L  V  K  P  *  M  K  T  L  P  T  Q  T  V  S  C  P  V  K  *  Q  F  W  L

      4240       4250       4260       4270       4280       4290       4300       4310       4320
ATGATATAGGTGGTATATGGAATTTTCACACTCTCAGATAAATGAAAAACGTACTTTGTGTTATACGTCACATTTTCTTTTCACATTATG
 M  I  *  V  V  Y  G  I  F  T  L  S  D  K  *  K  T  Y  F  V  L  Y  V  T  F  S  F  H  I  M
  *  Y  R  W  Y  M  E  F  S  H  S  Q  I  N  E  K  R  T  L  C  Y  T  S  H  F  L  F  T  L  *
D  D  I  G  G  I  W  N  F  H  T  L  R  *  M  K  N  V  L  C  V  I  R  H  I  F  F  S  H  Y
```

```
      4330       4340       4350       4360       4370       4380       4390       4400       4410
ATAAAAGACATAACAAAATTCTCGACAAGTACGGTATTATTTGTGTTTTTCTAACTTTTAGGCTGTTGTTAATTTTTCCAGCTTTCCAGC
 I  K  D  I  T  K  F  S  T  S  T  V  L  F  V  F  F  *  L  L  G  C  C  *  F  F  Q  L  S  S
  *  K  T  *  Q  N  S  R  Q  V  R  Y  Y  L  C  F  S  N  F  *  A  V  V  N  F  S  S  F  P  A
D  K  R  H  N  K  I  L  D  K  Y  G  I  I  C  V  F  L  T  F  R  L  L  L  I  F  P  A  F  Q

      4420       4430       4440       4450       4460       4470       4480       4490       4500
TATTTACGTATATTTTTCCATCTTTTTTGGCCATTTCTTGTGTTTTTTCAGCTTTCTCGTGTTTTTCTAACTCCTTGAATATTTTTCGTA
 Y  L  R  I  F  F  H  L  F  W  P  F  L  V  F  F  Q  L  S  R  V  F  L  T  P  *  I  F  F  V
  I  Y  V  Y  F  S  I  F  F  G  H  F  L  C  F  F  S  F  L  V  F  F  *  L  L  E  Y  F  S  Y
L  F  T  Y  I  F  P  S  F  L  A  I  S  C  V  F  S  A  F  S  C  F  S  N  S  L  N  I  F  R

      4510       4520       4530       4540       4550       4560       4570       4580       4590
TTTTCACAGCTTTTTAGAGGAGATGGTTTACATTTTCCTAAAGAACGGTACGTTTGATAAAGTACTTTATAAAAGGCCCGTTTATGGCGT
 F  S  Q  L  F  R  G  D  G  L  H  F  P  K  E  R  Y  V  *  *  S  T  L  *  K  A  R  L  W  R
  F  H  S  F  L  E  E  M  V  Y  I  F  L  K  N  G  T  F  D  K  V  L  Y  K  R  P  V  Y  G  V
I  F  T  A  F  *  R  R  W  F  T  F  S  *  R  T  V  R  L  I  K  Y  F  I  K  G  P  F  M  A

      4600       4610       4620       4630       4640       4650       4660       4670       4680
TAAAGACTTCACTTTTCATTTTATATTTTGCTATCTTTCAACATTCCTCCCGATTACTTTACAAAAATATCTTTTTTCTATTTTTGCACA
 *  R  L  H  F  S  F  Y  I  L  L  S  F  N  I  P  P  D  Y  F  T  K  I  S  F  F  Y  F  C  T
  K  D  F  T  F  H  F  I  F  C  Y  L  S  T  F  L  P  I  T  L  Q  K  Y  L  F  S  I  F  A  H
L  K  T  S  L  F  I  L  Y  F  A  I  F  Q  H  S  S  R  L  L  Y  K  N  I  F  F  L  F  L  H

      4690       4700       4710       4720       4730       4740       4750       4760       4770
TTTATCTTTTTTTTTGAATTTCTTTCTTTTTTGCTATTTTCTAGATATTTCCAGATTTTTAGTTTTGTATGCAATCTCTGAGAATGTCTT
 F  I  F  F  F  E  F  L  S  F  L  L  F  S  R  Y  F  Q  I  F  S  F  V  C  N  L  *  E  C  L
  L  S  F  F  L  N  F  F  L  F  C  Y  F  L  D  I  S  R  F  L  V  L  Y  A  I  S  E  N  V  L
I  Y  L  F  F  *  I  S  F  F  F  A  I  F  *  I  F  P  D  F  *  F  C  M  Q  S  L  R  M  S

      4780       4790       4800       4810       4820       4830       4840       4850       4860
GCAATTTGCGCAAAAATAGAAAAAGATATTTCATCGATATTTCTTAATTAGTCATTTACATTAGCTGCAGCATAACTCATTGAATTCCAT
 A  I  C  A  K  I  E  K  D  I  S  S  I  F  L  N  *  S  F  T  L  A  A  A  *  L  I  E  F  H
  Q  F  A  Q  K  *  K  K  I  F  H  R  Y  F  L  I  S  H  L  H  *  L  Q  H  N  S  L  N  S  I
C  N  L  R  K  N  R  K  R  Y  F  I  D  I  S  *  L  V  I  Y  I  S  C  S  I  T  H  *  I  P

      4870       4880       4890       4900       4910       4920       4930       4940       4950
ATTTATCCCAAAGATTAGTATATCAACATTATATTAAACGTTGATATACTAATCTCATTTGGAACAAAATGATATCAAAAACATGAAAAA
 I  Y  P  K  D  *  Y  I  N  I  I  L  N  V  D  I  L  I  S  F  G  T  K  *  Y  Q  K  H  E  K
  F  I  P  K  I  S  I  S  T  L  Y  *  T  L  I  Y  *  S  H  L  E  Q  N  D  I  K  N  M  K  N
Y  L  S  Q  R  L  V  Y  Q  H  Y  I  K  R  *  Y  T  N  L  I  W  N  K  M  I  S  K  T  *  K

      4960       4970       4980       4990       5000       5010       5020       5030       5040
TTTGCTACTTATAGAAAATACTGTACATCTAAGCTTACTTCTACTTCTCAATTAATTTTTTCTTGAATCCTTTGTATGCACCGACAGTTA
 F  A  T  Y  R  K  Y  C  T  S  K  L  T  S  T  S  Q  L  I  F  S  *  I  L  C  M  H  R  Q  L
  L  L  L  I  E  N  T  V  H  L  S  L  L  L  L  L  N  *  F  F  L  E  S  F  V  C  T  D  S  Y
I  C  Y  L  *  K  I  L  Y  I  *  A  Y  F  Y  F  S  I  N  F  F  L  N  P  L  Y  A  P  T  V

      5050       5060       5070       5080       5090       5100       5110       5120       5130
TTCTCACCAATCATTTCAATACAATAATCTCTTATTACTCTGTAAACTCAGCAAAAAAAGTATTCCTAATCGCTGAGCTATGTGTAATTA
 F  S  P  I  I  S  I  Q  *  S  L  I  T  L  *  T  Q  Q  K  K  Y  S  *  S  L  S  Y  V  *  L
  S  H  Q  S  F  Q  Y  N  N  L  L  L  L  C  K  L  S  K  K  S  I  P  N  R  *  A  M  C  N  *
I  L  T  N  H  F  N  T  I  I  S  Y  Y  S  V  N  S  A  K  K  V  F  L  I  A  E  L  C  V  I

      5140       5150       5160       5170       5180       5190       5200       5210       5220
GTTACGGTAATTCCATGCATGTATATTAACTGATTGTTTCGCTTATTTCTTACACATACACTTAATATCTTCCATTAATGAAGGCAAATA
 V  T  V  I  P  C  M  Y  I  N  *  L  F  R  L  F  L  T  H  T  L  N  I  F  H  *  *  R  Q  I
  L  R  *  F  H  A  C  I  L  T  D  C  F  A  Y  F  L  H  I  H  L  I  S  S  I  N  E  G  K  Y
S  Y  G  N  S  M  H  V  Y  *  L  I  V  S  L  I  S  Y  T  Y  T  *  Y  L  P  L  M  K  A  N
```

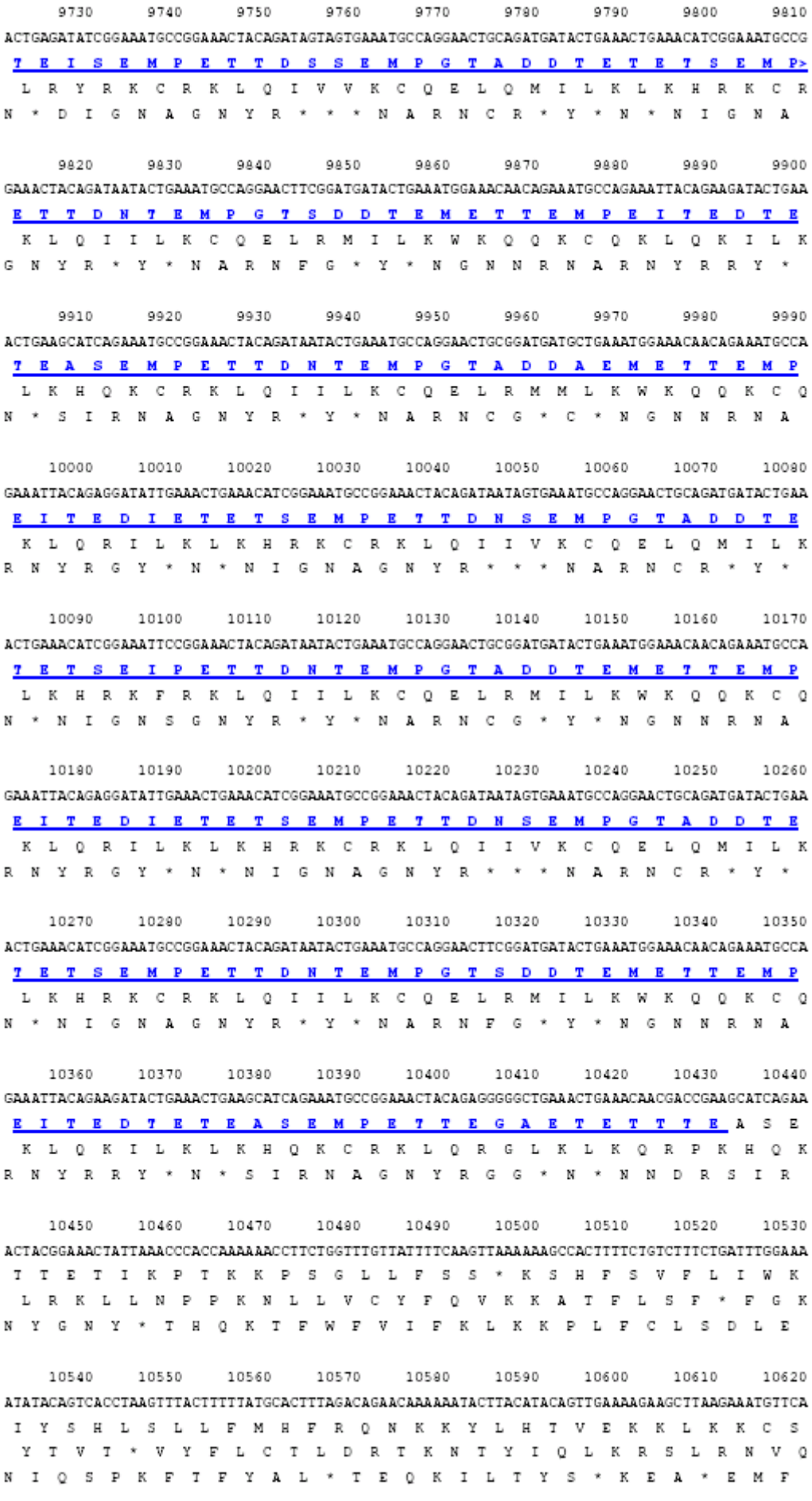
```
      5230       5240       5250       5260       5270       5280       5290       5300       5310
TCAGTTAATCATAGTAATGAATTTTGAAATACAGAAGTGTACAAGTCCCATGAAGTGACATGAATTGGATTATGATAACGCTTTTCATCA
 S  V  N  H  S  N  E  F  *  N  T  E  V  Y  K  S  H  E  V  T  *  I  G  L  *  *  R  F  S  S
  Q  L  I  I  V  M  N  F  E  I  Q  K  C  T  S  P  M  K  *  H  E  L  D  Y  D  N  A  F  H  H
I  S  *  S  *  *  *  I  L  K  Y  R  S  V  Q  V  P  *  S  D  M  N  W  I  M  I  T  L  F  I

      5320       5330       5340       5350       5360       5370       5380       5390       5400
TATTCGCCAATATAATTACTGTTGCAAATGGTAAGTTGCTCAATAAAATGACACGTCAGTTGCGTAAGCACGAAATATCATCGCTGAAGA
 Y  S  P  I  *  L  L  L  Q  M  V  S  C  S  I  K  *  H  V  S  C  V  S  T  K  Y  H  R  *  R
  I  R  Q  Y  N  Y  C  C  K  W  *  V  A  Q  *  N  D  T  S  V  A  *  A  R  N  I  I  A  E  D
I  F  A  N  I  I  T  V  A  N  G  K  L  L  N  K  M  T  R  Q  L  R  K  H  E  I  S  S  L  K

      5410       5420       5430       5440       5450       5460       5470       5480       5490
TATAATGAACATGACGGTTGACACTTCAGATCATACTACATCTTAATCATTATATGTCATTTCAATCCTAAACTGTCGAGTAGCATCTAT
 Y  N  E  H  D  G  *  H  F  R  S  Y  Y  I  L  I  I  I  C  H  F  N  P  K  L  S  S  S  I  Y
  I  M  N  M  T  V  D  T  S  D  H  T  T  S  *  S  L  Y  V  I  S  I  L  N  C  R  V  A  S  I
I  *  *  T  *  R  L  T  L  Q  I  I  L  H  L  N  H  Y  M  S  F  Q  S  *  T  V  E  *  H  L

      5500       5510       5520       5530       5540       5550       5560       5570       5580
TGAAATTCAATAGTGGCATATATTGAATTTCACAGAAGCTCTCACAAGAAATAAGCAAATGAAGAGACTAAATTGAAGATAATACAAATT
 *  N  S  I  V  A  Y  I  E  F  H  R  S  S  H  K  K  *  A  N  E  E  T  K  L  K  I  I  Q  I
  E  I  Q  *  W  H  I  L  N  F  T  E  A  L  T  R  N  K  Q  M  K  R  L  N  *  R  *  Y  K  F
L  K  F  N  S  G  I  Y  *  I  S  Q  K  L  S  Q  E  I  S  K  *  R  D  *  I  E  D  N  T  N

      5590       5600       5610       5620       5630       5640       5650       5660       5670
TTTACTTGAAGATTTAGTAACATTGTATTTGAATAAATTGCAAAGAAATAACCACCAAATCGATGTTGAACATATCTTCAGGTTGTAGTC
 F  T  *  R  F  S  N  I  V  F  E  *  I  A  K  K  *  P  P  N  R  C  *  T  Y  L  Q  V  V  V
  L  L  E  D  L  V  T  L  Y  L  N  K  L  Q  R  N  N  H  Q  I  D  V  E  H  I  F  R  L  *  S
F  Y  L  K  I  *  *  H  C  I  *  I  N  C  K  E  I  T  T  K  S  M  L  N  I  S  S  G  C  S

      5680       5690       5700       5710       5720       5730       5740       5750       5760
TATTTCAGGTTACGTTCGTGGATGTTACTATATTAGTTGGGCTGAACGTCAACAAGGTTCAGTGACGTTTGAAATAATTTCATTATTAAA
 Y  F  R  L  R  S  W  M  L  L  Y  *  L  G  *  T  S  T  R  F  S  D  V  *  N  N  F  I  I  K
  I  S  G  Y  V  R  G  C  Y  Y  I  S  W  A  E  R  Q  Q  G  S  V  T  F  E  I  I  S  L  L  K
L  F  Q  V  T  F  V  D  V  T  I  L  V  G  L  N  V  N  K  V  Q  *  R  L  K  *  F  H  Y  *

      5770       5780       5790       5800       5810       5820       5830       5840       5850
GAGCTAAATAGTTTCAAGGGGATTGTGTACAAATTTTCGAATAAAAAGTAGAAAACCAATGAAATTATTTCAAGATTTCCATCTCAAAAA
 E  L  N  S  F  K  G  I  V  Y  K  F  S  N  K  K  *  K  T  N  E  I  I  S  R  F  P  S  Q  K
  S  *  I  V  S  R  G  L  C  T  N  F  R  I  K  S  R  K  P  M  K  L  F  Q  D  F  H  L  K  N
R  A  K  *  F  Q  G  D  C  V  Q  I  F  E  *  K  V  E  N  Q  *  N  Y  F  K  I  S  I  S  K

      5860       5870       5880       5890       5900       5910       5920       5930       5940
TGAATTTTAAGCCCACGAAATATTAGTGGAAAATAAACATCAATAATTAGCAACTTCAACTTCCAACATAAACTCAATATTCCCTCATTT
 *  I  L  S  P  R  N  I  S  G  K  *  T  S  I  I  S  N  F  N  F  Q  H  K  L  N  I  P  S  F
  E  F  *  A  H  E  I  L  V  E  N  K  H  Q  *  L  A  T  S  T  S  N  I  N  S  I  F  P  H  F
M  N  F  K  P  T  K  Y  *  W  K  I  N  I  N  N  *  Q  L  Q  L  P  T  *  T  Q  Y  S  L  I

      5950       5960       5970       5980       5990       6000       6010       6020       6030
CCTGAATTCTCATCGTTTCGACGTGTGCCTTAGCTACCGTAAACTAATATGGTGAAAATAATAACTAAACAAATTAAAATAAAACAAGCA
 P  E  F  S  S  F  R  R  V  P  *  L  P  *  T  N  M  V  K  I  I  T  K  Q  I  K  I  K  Q  A
  L  N  S  H  R  F  D  V  C  L  S  Y  R  K  L  I  W  *  K  *  *  L  N  K  L  K  *  N  K  H
S  *  I  L  I  V  S  T  C  A  L  A  T  V  N  *  Y  G  E  N  N  N  *  T  N  *  N  K  T  S

      6040       6050       6060       6070       6080       6090       6100       6110       6120
TTGTATTTCTGCCGGAAATGTAATATACATTTTCATCTTAAAGGAGAAGGCAAATTTTTGCCAGAAGATATCCCAAAACACCTATGCACT
 L  Y  F  C  R  K  C  N  I  H  F  H  L  K  G  E  G  K  F  L  P  E  D  I  P  K  H  L  C  T
  C  I  S  A  G  N  V  I  Y  I  F  I  L  K  E  K  A  N  F  C  Q  K  I  S  Q  N  T  Y  A  L
I  V  F  L  P  E  M  *  Y  T  F  S  S  *  R  R  R  Q  I  F  A  R  R  Y  P  K  T  P  M  H
```


6130 6140 6150 6160 6170 6180 6190 6200 6210
CACATTCTTTATGCATTCACCAACGTTGAAGGGAAACGCGCGTAAGTTACAAATGATATTTGCAATGAGAAATGTAAATTAAATTTTCGG
H I L Y A F T N V E G K R A * V T N D I C N E K C K L N F R
T F F M H S P T L K G N A R K L Q M I F A M R N V N * I F G
S H S L C I H Q R * R E T R V S Y K * Y L Q * E M * I K F S
6220 6230 6240 6250 6260 6270 6280 6290 6300
AGTTTCTTTTACAATGCACATTAAAATTATTCTAATTTTTATTGCCTGCTCAAGTGATTATAAACCTGAAACGTTAATATGCGATATTTA
S F F Y N A H * N Y S N F Y C L L K * L * T * N V N M R Y L
V S F T M H I K I I L I F I A C S S D Y K P E T L I C D I *
E F L L Q C T L K L F * F L L P A Q V I I N L K R * Y A I F
6310 6320 6330 6340 6350 6360 6370 6380 6390
GCTGGTTTTGCATAAAATATTCTTCTGTAACAATTGCTATCAGCAAATGTTATTTATTTCTGTAAATTTCTGATTTTAATAATTTTTAAA
A G F A * N I L L * Q L L S A N V I Y F C K F L I L I I F K
L V L H K I F F C N N C Y Q Q M L F I S V N F * F * * F L N
S W F C I K Y S S V T I A I S K C Y L F L * I S D F N N F *
6400 6410 6420 6430 6440 6450 6460 6470 6480
TCTTAGAACATTTGAATGGGACGATGAAGATACCGAATGGACGAAAGGGATATATCCACATATGATAAAAGTAAAAGAGTACGATCCAAC
S * N I * M G R * R Y R M D E R D I S T Y D K S K R V R S N
L R T F E W D D E D T E W T K G I Y P H M I K V K E Y D P T
I L E H L N G T M K I P N G R K G Y I H I * * K * K S T I Q>
6490 6500 6510 6520 6530 6540 6550 6560 6570
CCTGAAAATTCTTCTTTCCTATAGCGACTATAACTCTCACTCATCAGTTTTTGCTGTTCGTTTTTTTTTTAAACTCTCAGCTATCATTTT
P E N S S F L * R L * L S L I S F C C S F F F * T L S Y H F
L K I L L S Y S D Y N S H S S V F A V R F F F K L S A I I F
P * K F F F P I A T I T L T H Q F L L F V F F L N S Q L S F
6580 6590 6600 6610 6620 6630 6640 6650 6660
TTTTTTTAAAGATTGACACTCAGTTAACGTATTGACTAATAATTTCAAAAGAGACTGCTCATNCGTCATTTNTTANACATNAAGATAATG
F F * R L T L S * R I D * * F Q K R L L X R H X L X X K I M
F F K D * H S V N V L T N N F K R D C S X V I X X T X R * C
F F L K I D T Q L T Y * L I I S K E T A H X S F X X H X D N
6670 6680 6690 6700 6710 6720 6730 6740 6750
CAGGGTTACCTCANACGTTTTCCGGATTTACGNNTCATNTTTTTNTTGCTAAAATATCAACGTAAATTTGATGATNTAAAAATNTTTGAA
Q G Y L X R F P D L R X X F X L L K Y Q R K F D D X K X F E
R V T S X V F R I Y X S X F X C * N I N V N L M X * K X L K
A G L P X T F S G F T X H X F X A K I S T * I * * X K N X *
6760 6770 6780 6790 6800 6810 6820 6830 6840
GGCAATAGCAAAATCGACAAAAAACCGGAAAAATTTCATTCAGNCAACTGTAGCATTTCTCCGTAAGCATAAATTTGANGGATTCGATCT
G N S K I D K K P E K F H S X N C S I S P * A * I * X I R S
A I A K S T K N R K N F I Q X T V A F L R K H K F X G F D L
R Q * Q N R Q K T G K I S F X Q L * H F S V S I N L X D S I
6850 6860 6870 6880 6890 6900 6910 6920 6930
CGACTGGGAATATCCAACCGGNGTAGCAAAGGAACACGCTAAACTTCTTAAGGCGAGTGTAATTCCATATTCAACTTTTTTCAGTCCACC
R L G I S N R X S K G T R * T S * G E C N S I F N F F Q S T
D W E Y P T G V A K E H A K L L K A S V I P Y S T F F S P P
S T G N I Q P X * Q R N T L N F L R R V * F H I Q L F S V H
6940 6950 6960 6970 6980 6990 7000 7010 7020
GGAGCACGCTCATTTCGAATTGAGCAATACTTGGTGAGCACTGTTTATAGCAATTAATTAATTAATCACTTGGTGAGCACTGTTTATAGC
G A R S F R I E Q Y L V S T V Y S N * L I N H L V S T V Y S
E H A H F E L S N T W * A L F I A I N * L I T W * A L F I A
R S T L I S N * A I L G E H C L * Q L I N * S L G E H C L *

```
      7030       7040       7050       7060       7070       7080       7090       7100       7110
AATTAATTAATTAATCACTTGGTGAGCACTGTTTATAGCAATTAATTAATTAATCACTTTGTGTATGAAATTAATAGAGAACAAGGAATC
 N  *  L  I  N  H  L  V  S  T  V  Y  S  N  *  L  I  N  H  F  V  Y  E  I  N  R  E  Q  G  I
  I  N  *  L  I  T  W  *  A  L  F  I  A  I  N  *  L  I  T  L  C  M  K  L  I  E  N  K  E  S
Q  L  I  N  *  S  L  G  E  H  C  L  *  Q  L  I  N  *  S  L  C  V  *  N  *  *  R  T  R  N

      7120       7130       7140       7150       7160       7170       7180       7190       7200
TTCTGGCATTGTTTATAGTGGTTAGATAAGGAGCTGTTGTTGAATATCCGGTAGAGGAAAATAATAATAAACAAAGAATTACCTGGTCAT
 F  W  H  C  L  *  W  L  D  K  E  L  L  L  N  I  R  *  R  K  I  I  I  N  K  E  L  P  G  H
  S  G  I  V  Y  S  G  *  I  R  S  C  C  *  I  S  G  R  G  K  *  *  *  T  K  N  Y  L  V  I
L  L  A  L  F  I  V  V  R  *  G  A  V  V  E  Y  P  V  E  E  N  N  N  K  Q  R  I  T  W  S

      7210       7220       7230       7240       7250       7260       7270       7280       7290
CTGATAGGCGGAACTGATAGAAAATAAAAGATATTTTCTGTCGAATCACATATCGAAAGAACAAATAAAGCATATTTTATCACCTGACTG
 L  I  G  G  T  D  R  K  *  K  I  F  S  V  E  S  H  I  E  R  T  N  K  A  Y  F  I  T  *  L
  *  *  A  E  L  I  E  N  K  R  Y  F  L  S  N  H  I  S  K  E  Q  I  K  H  I  L  S  P  D  C
S  D  R  R  N  *  *  K  I  K  D  I  F  C  R  I  T  Y  R  K  N  K  *  S  I  F  Y  H  L  T

      7300       7310       7320       7330       7340       7350       7360       7370       7380
TTCTATAGCATCGATATCAAGAATCAACTATATATTAATTATACAAATATTATTATATTATAAAATTATTTCTCATTTCCGACACTCGTC
 F  Y  S  I  D  I  K  N  Q  L  Y  I  N  Y  T  N  I  I  I  L  *  N  Y  F  S  F  P  T  L  V
  S  I  A  S  I  S  R  I  N  Y  I  L  I  I  Q  I  L  L  Y  Y  K  I  I  S  H  F  R  H  S  S
V  L  *  H  R  Y  Q  E  S  T  I  Y  *  L  Y  K  Y  Y  Y  I  I  K  L  F  L  I  S  D  T  R

      7390       7400       7410       7420       7430       7440       7450       7460       7470
AAATTTTAATTCACATTCCAAAAGAATACTCCACGTGATTTCTGTAATTTTCTCTTCAAAAATGAATTAGTTTCAAAAGCAAAATTGAAA
 K  F  *  F  T  F  Q  K  N  T  P  R  D  F  C  N  F  L  F  K  N  E  L  V  S  K  A  K  L  K
  N  F  N  S  H  S  K  R  I  L  H  V  I  S  V  I  F  S  S  K  M  N  *  F  Q  K  Q  N  *  N
Q  I  L  I  H  I  P  K  E  Y  S  T  *  F  L  *  F  S  L  Q  K  *  I  S  F  K  S  K  I  E

      7480       7490       7500       7510       7520       7530       7540       7550       7560
TATATCAGATCGTACGTGAGAGGATGAAAAATTCAAAACGAACTTTGTAGGAAATGAAAGCGGCGTTCGTGGAAGAAGCCAAGGAATCCG
 Y  I  R  S  Y  V  R  G  *  K  I  Q  N  E  L  C  R  K  *  K  R  R  S  W  K  K  P  R  N  P
  I  S  D  R  T  *  E  D  E  K  F  K  T  N  F  V  G  N  E  S  G  V  R  G  R  S  Q  G  I  R
I  Y  Q  I  V  R  E  R  M  K  N  S  K  R  T  L  *  E  M  K  A  A  F  V  E  E  A  K  E  S

      7570       7580       7590       7600       7610       7620       7630       7640       7650
GCAAGCAACAATTACTTCTTACCGCTGCTGTATCAGCAGTGAAAGAAACTATTGATGAAAGTTACAGTGTTCGATCCCTTGGAAAGTAAA
 A  S  N  N  Y  F  L  P  L  L  Y  Q  Q  *  K  K  L  L  M  K  V  T  V  F  D  P  L  E  S  K
  Q  A  T  I  T  S  Y  R  C  C  I  S  S  E  R  N  Y  *  *  K  L  Q  C  S  I  P  W  K  V  K
G  K  Q  Q  L  L  L  T  A  A  V  S  A  V  K  E  T  I  D  E  S  Y  S  V  R  S  L  G  K  *

      7660       7670       7680       7690       7700       7710       7720       7730       7740
GAAATTGCTTTTCATACATTGACATTAAATTTAACACGCAATAAAATTTACAATACTTCAAATATATTTTTGTAGCGTTGAAAATTTTCA
 E  I  A  F  H  T  L  T  L  N  L  T  R  N  K  I  Y  N  T  S  N  I  F  L  *  R  *  K  F  S
  K  L  L  F  I  H  *  H  *  I  *  H  A  I  K  F  T  I  L  Q  I  Y  F  C  S  V  E  N  F  Q
R  N  C  F  S  Y  I  D  I  K  F  N  T  Q  *  N  L  Q  Y  F  K  Y  I  F  V  A  L  K  I  F

      7750       7760       7770       7780       7790       7800       7810       7820       7830
GAAGTTTGGATCTTCTATTTCTGATGTCGTATGATTTACACGGCAGTTGGGAGAAGAACGTCGACTTACACGCAAAGTTGCGTCCGACGG
 E  V  W  I  F  Y  F  *  C  R  M  I  Y  T  A  V  G  R  R  T  S  T  Y  T  Q  S  C  V  R  R
  K  F  G  S  S  I  S  D  V  V  *  F  T  R  Q  L  G  E  E  R  R  L  T  R  K  V  A  S  D  G
R  S  L  D  L  L  F  L  M  S  Y  D  L  H  G  S  W  E  K  N  V  D  L  H  A  K  L  R  P  T

      7840       7850       7860       7870       7880       7890       7900       7910       7920
AGAGAGAAACATCCGGAACTGGCGTCTTTAATACCGTTAGTTCTTGAAATATATAAAAATTTCTAATTTTTTCTTGCTTGCATGAGACGA
 R  E  K  H  P  E  L  A  S  L  I  P  L  V  L  E  I  Y  K  N  F  *  F  F  L  A  C  M  R  R
  E  R  N  I  R  N  W  R  L  *  Y  R  *  F  L  K  Y  I  K  I  S  N  F  F  L  L  A  *  D  E
E  R  E  T  S  G  T  G  V  F  N  T  V  S  S  *  N  I  *  K  F  L  I  F  S  C  L  H  E  T
```

```
      7930       7940       7950       7960       7970       7980       7990       8000       8010
ATTCATGATTATTTCAGAAATTTGCAGCGAACTATTGGGCAAAGAGAGGTATGCCGAAACAGAAAATTATCATTGGAATTCCGACTTACA
 I  H  D  Y  F  R  N  L  Q  R  T  I  G  Q  R  E  V  C  R  N  R  K  L  S  L  E  F  R  L  T
  F  M  I  I  S  E  I  C  S  E  L  L  G  K  E  R  Y  A  E  T  E  N  Y  H  W  N  S  D  L  Q
N  S  *  L  F  Q  K  F  A  A  N  Y  W  A  K  R  G  M  P  K  Q  K  I  I  I  G  I  P  T  Y

      8020       8030       8040       8050       8060       8070       8080       8090       8100
GTCGGGGATGGACATTAAGTAATTCTTCGGAAACGGCAATTGGAGCGGAGGGTGGTCGTCCATCATCGCCATCGACAACGAATCCCGTGG
 V  G  D  G  H  *  V  I  L  R  K  R  Q  L  E  R  R  V  V  V  H  H  R  H  R  Q  R  I  P  W
  S  G  M  D  I  K  *  F  F  G  N  G  N  W  S  G  G  W  S  S  I  I  A  I  D  N  E  S  R  G
S  R  G  W  T  L  S  N  S  S  E  T  A  I  G  A  E  G  G  R  P  S  S  P  S  T  T  N  P  V

      8110       8120       8130       8140       8150       8160       8170       8180       8190
CGGTACTGCGGCTTATTGGGAGGTTTAATTTAAAGTTCTCATTTACTTTTAAATACATTTCTTGCTGCTGATGAATTTCTCAAATTCCAT
 R  Y  C  G  L  L  G  G  L  I  *  S  S  H  L  L  L  N  T  F  L  A  A  D  E  F  L  K  F  H
  G  T  A  A  Y  W  E  V  *  F  K  V  L  I  Y  F  *  I  H  F  L  L  L  M  N  F  S  N  S  I
A  V  L  R  L  I  G  R  F  N  L  K  F  S  F  T  F  K  Y  I  S  C  C  *  *  I  S  Q  I  P

      8200       8210       8220       8230       8240       8250       8260       8270       8280
TATTAATGAATGTAACCATTTTAAGATATGCAAATATGTGAAAGAAGGTGGCAAGGAAACGGTAGATAAAAAAGGTGTTGGAGCGTATAT
 Y  *  *  M  *  P  F  *  D  M  Q  I  C  E  R  R  W  Q  G  N  G  R  *  K  R  C  W  S  V  Y
  I  N  E  C  N  H  F  K  I  C  K  Y  V  K  E  G  G  K  E  T  V  D  K  K  G  V  G  A  Y  M
L  L  M  N  V  T  I  L  R  Y  A  N  M  *  K  K  V  A  R  K  R  *  I  K  K  V  L  E  R  I>

      8290       8300       8310       8320       8330       8340       8350       8360       8370
GGTAAAAGGAGATCAGTGGCACGGTTACGATAATGAGGAAACTATCAAAATTAAAGTAATTTTTATTGTTCTTGATGTTATTGATAGTGA
 G  K  R  R  S  V  A  R  L  R  *  *  G  N  Y  Q  N  *  S  N  F  Y  C  S  *  C  Y  *  *  *
  V  K  G  D  Q  W  H  G  Y  D  N  E  E  T  I  K  I  K  V  I  F  I  V  L  D  V  I  D  S  E
W  *  K  E  I  S  G  T  V  T  I  M  R  K  L  S  K  L  K  *  F  L  L  F  L  M  L  L  I  V

      8380       8390       8400       8410       8420       8430       8440       8450       8460
AAAAAAAAATAATTTAAGAAATAACGTTGCAAAATATATTTTTTTTTATTTTTAATCTAAATAGTTCAATAAATTTCATAAAGTTTAACGA
 K  K  N  N  L  R  N  N  V  A  K  Y  I  F  F  Y  F  *  S  K  *  F  N  K  F  H  K  V  *  R
  K  K  I  I  *  E  I  T  L  Q  N  I  F  F  F  I  F  N  L  N  S  S  I  N  F  I  K  F  N  E
K  K  K  *  F  K  K  *  R  C  K  I  Y  F  F  L  F  L  I  *  I  V  Q  *  I  S  *  S  L  T

      8470       8480       8490       8500       8510       8520       8530       8540       8550
GAAGAAATTAATTTTAGAATTATTGGCGTAATTACGGTTTATTAATTAAATTCAGTTTGTGTTTTATGTTTGAAACAACATCTAAAAAGT
 E  E  I  N  F  R  I  I  G  V  I  T  V  Y  *  L  N  S  V  C  V  L  C  L  K  Q  H  L  K  S
  K  K  L  I  L  E  L  L  A  *  L  R  F  I  N  *  I  Q  F  V  F  Y  V  *  N  N  I  *  K  V
R  R  N  *  F  *  N  Y  W  R  N  Y  G  L  L  I  K  F  S  L  C  F  M  F  E  T  T  S  K  K

      8560       8570       8580       8590       8600       8610       8620       8630       8640
GATTGGTTCGCATGATTGAAACTTTCAGATGAGATGGTTAAAAAAGAACGGTTATGGCGGTGCCTTTATGTGGACTCTTGATTTTGATGA
 D  W  F  A  *  L  K  L  S  D  E  M  V  K  K  E  R  L  W  R  C  L  Y  V  D  S  *  F  *  *
  I  G  S  H  D  *  N  F  Q  M  R  W  L  K  K  N  G  Y  G  G  A  F  M  W  T  L  D  F  D  D
*  L  V  R  M  I  E  T  F  R  *  D  G  *  K  R  T  V  M  A  V  P  L  C  G  L  L  I  L  M

      8650       8660       8670       8680       8690       8700       8710       8720       8730
TTTCAAAGGTACAAGTTGTGGCAAAGGTCCATATCCACTGTTAAATGCTATCAATAATGGGCTTGAAAGTGAGGTAAGACATGTTTCAAA
 F  Q  R  Y  K  L  W  Q  R  S  I  S  T  V  K  C  Y  Q  *  W  A  *  K  *  G  K  T  C  F  K
  F  K  G  T  S  C  G  K  G  P  Y  P  L  L  N  A  I  N  N  G  L  E  S  E  V  R  H  V  S  K
I  S  K  V  Q  V  V  A  K  V  H  I  H  C  *  M  L  S  I  M  G  L  K  V  R  *  D  M  F  Q

      8740       8750       8760       8770       8780       8790       8800       8810       8820
GAAATATTTCTAACAATTCTGTTTTTCTTTCAATAAGTTTGGATTTAAATTATGCGTGATATTTGTAAATAAATTAATAAAAAAAAAAGG
 E  I  F  L  T  I  L  F  F  F  Q  *  V  W  I  *  I  M  R  D  I  C  K  *  I  N  K  K  K  R
  K  Y  F  *  Q  F  C  F  S  F  N  K  F  G  F  K  L  C  V  I  F  V  N  K  L  I  K  K  K  G
R  N  I  S  N  N  S  V  F  L  S  I  S  L  D  L  N  Y  A  *  Y  L  *  I  N  *  *  K  K  K
```

```
      8830       8840       8850       8860       8870       8880       8890       8900       8910
CGGATGGATGAATCANAATAAACTTTCTATGGGACCATACGGAAAAATGATTTTATGAATTGTAGGATAAATAGAATTTCTGTCCGAAGA
 R  M  D  E  S  X  *  T  F  Y  G  T  I  R  K  N  D  F  M  N  C  R  I  N  R  I  S  V  R  R
  G  W  M  N  X  N  K  L  S  M  G  P  Y  G  K  M  I  L  *  I  V  G  *  I  E  F  L  S  E  D
A  D  G  *  I  X  I  N  F  L  W  D  H  T  E  K  *  F  Y  E  L  *  D  K  *  N  F  C  P  K

      8920       8930       8940       8950       8960       8970       8980       8990       9000
TTATGACTCACTAACTTTTGATTAGCTGGATGGGACGTCAATTTCTATATTGATCATGACGTTGATCAAACAAAATCCTTAAATTTTAAA
 L  *  L  T  N  F  *  L  A  G  W  D  V  N  F  Y  I  D  H  D  V  D  Q  T  K  S  L  N  F  K
  Y  D  S  L  T  F  D  *  L  D  G  T  S  I  S  I  L  I  M  T  L  I  K  Q  N  P  *  I  L  N
I  M  T  H  *  L  L  I  S  W  M  G  R  Q  F  L  Y  *  S  *  R  *  S  N  K  I  L  K  F  *

      9010       9020       9030       9040       9050       9060       9070       9080       9090
TTATCCCTTGAAAGTTAATTTATAAAATTTAACACACATTTTTAGCAAACACCATCAACAACAGCAGCACCGGAAACCACAGAAGACACT
 L  S  L  E  S  *  F  I  K  F  N  T  H  F  *  Q  T  P  S  T  T  A  A  P  E  T  T  E  D  T
  Y  P  L  K  V  N  L  *  N  L  T  H  I  F  S  K  H  H  Q  Q  Q  Q  H  R  K  P  Q  K  T  L
I  I  P  *  K  L  I  Y  K  I  *  H  T  F  L  A  N  T  I  N  N  S  S  T  G  N  H  R  R  H

      9100       9110       9120       9130       9140       9150       9160       9170       9180
GAAGTTGAAACAATGACGGAAGCACCGGGAACTGCAAATGACACTGAAATGGAAACAACAAATATACCAGAAATTACAGAAGATACTGAA
 E  V  E  T  M  T  E  A  P  G  T  A  N  D  T  E  M  E  T  T  N  I  P  E  I  T  E  D  T  E
  K  L  K  Q  *  R  K  H  R  E  L  Q  M  T  L  K  W  K  Q  Q  I  Y  Q  K  L  Q  K  I  L  K
*  S  *  N  N  D  G  S  T  G  N  C  K  *  H  *  N  G  N  N  K  Y  T  R  N  Y  R  R  Y  *

      9190       9200       9210       9220       9230       9240       9250       9260       9270
ACTGAAACATCGGAAATGCCGGAAACTACAGATAATACTGAAATGCCAGGAACTGCAGATGATACTGAAACTGAAACATCGGAAATGCCG
 T  E  T  S  E  M  P  E  T  T  D  N  T  E  M  P  G  T  A  D  D  T  E  T  E  T  S  E  M  P
  L  K  H  R  K  C  R  K  L  Q  I  I  L  K  C  Q  E  L  Q  M  I  L  K  L  K  H  R  K  C  R
N  *  N  I  G  N  A  G  N  Y  R  *  Y  *  N  A  R  N  C  R  *  Y  *  N  *  N  I  G  N  A

      9280       9290       9300       9310       9320       9330       9340       9350       9360
GAAACTACAGATAATACTGAAATGCCGGGAACTGCAGATGATACTGACGAAACATCGGAAATGCCGGAAACTACAGATAATACTGAAATG
 E  T  T  D  N  T  E  M  P  G  T  A  D  D  T  D  E  T  S  E  M  P  E  T  T  D  N  T  E  M
  K  L  Q  I  I  L  K  C  R  E  L  Q  M  I  L  T  K  H  R  K  C  R  K  L  Q  I  I  L  K  C
G  N  Y  R  *  Y  *  N  A  G  N  C  R  *  Y  *  R  N  I  G  N  A  G  N  Y  R  *  Y  *  N

      9370       9380       9390       9400       9410       9420       9430       9440       9450
CCAGGAACTGCGGATGATACTGAAATGGAAACAACAGAAATGCCAGAAATTACAGAGGATACTGAAACTGAAACATGGGAAATGCCGGAA
 P  G  T  A  D  D  T  E  M  E  T  T  E  M  P  E  I  T  E  D  T  E  T  E  T  W  E  M  P  E
  Q  E  L  R  M  I  L  K  W  K  Q  Q  K  C  Q  K  L  Q  R  I  L  K  L  K  H  G  K  C  R  K
A  R  N  C  G  *  Y  *  N  G  N  N  R  N  A  R  N  Y  R  G  Y  *  N  *  N  M  G  N  A  G

      9460       9470       9480       9490       9500       9510       9520       9530       9540
ACTACAGATGATATTGAAATGCCAGGAACTGCAGATGATACTGAAACTGAAACATCGGAAATGCCGGAAATTACAGATAATACTGAAATG
 T  T  D  D  I  E  M  P  G  T  A  D  D  T  E  T  E  T  S  E  M  P  E  I  T  D  N  T  E  M
  L  Q  M  I  L  K  C  Q  E  L  Q  M  I  L  K  L  K  H  R  K  C  R  K  L  Q  I  I  L  K  C
N  Y  R  *  Y  *  N  A  R  N  C  R  *  Y  *  N  *  N  I  G  N  A  G  N  Y  R  *  Y  *  N

      9550       9560       9570       9580       9590       9600       9610       9620       9630
CCGGGAACTGCGGATGATACTGAAATGGAAACAACAGAAATGCCAGAAATTACAGAAGATACTGACACTAAAACATCGACCGATGCATCT
 P  G  T  A  D  D  T  E  M  E  T  T  E  M  P  E  I  T  E  D  T  D  T  K  T  S  T  D  A  S
  R  E  L  R  M  I  L  K  W  K  Q  Q  K  C  Q  K  L  Q  K  I  L  T  L  K  H  R  P  M  H  L
A  G  N  C  G  *  Y  *  N  G  N  N  R  N  A  R  N  Y  R  R  Y  *  H  *  N  I  D  R  C  I

      9640       9650       9660       9670       9680       9690       9700       9710       9720
GAAACTACAGAAGATACTGACAATGAAACATCGGAAATGCCGGAAACTACAGATAATACTGCAATGCCAGGAACTGCAGATGATACTGAA
 E  T  T  E  D  T  D  N  E  T  S  E  M  P  E  T  T  D  N  T  A  M  P  G  T  A  D  D  T  E
  K  L  Q  K  I  L  T  M  K  H  R  K  C  R  K  L  Q  I  I  L  Q  C  Q  E  L  Q  M  I  L  K
*  N  Y  R  R  Y  *  Q  *  N  I  G  N  A  G  N  Y  R  *  Y  C  N  A  R  N  C  R  *  Y  *
```


9730 9740 9750 9760 9770 9780 9790 9800 9810
ACTGAGATATCGGAAATGCCGGAAACTACAGATAGTAGTGAAATGCCAGGAACTGCAGATGATACTGAAACTGAAACATCGGAAATGCCG
T E I S E M P E T T D S S E M P G T A D D T E T E T S E M P>
L R Y R K C R K L Q I V V K C Q E L Q M I L K L K H R K C R
N * D I G N A G N Y R * * * N A R N C R * Y * N * N I G N A
9820 9830 9840 9850 9860 9870 9880 9890 9900
GAAACTACAGATAATACTGAAATGCCAGGAACTTCGGATGATACTGAAATGGAAACAACAGAAATGCCAGAAATTACAGAAGATACTGAA
E T T D N T E M P G T S D D T E M E T T E M P E I T E D T E
K L Q I I L K C Q E L R M I L K W K Q Q K C Q K L Q K I L K
G N Y R * Y * N A R N F G * Y * N G N N R N A R N Y R R Y *
9910 9920 9930 9940 9950 9960 9970 9980 9990
ACTGAAGCATCAGAAATGCCGGAAACTACAGATAATACTGAAATGCCAGGAACTGCGGATGATGCTGAAATGGAAACAACAGAAATGCCA
T E A S E M P E T T D N T E M P G T A D D A E M E T T E M P
L K H Q K C R K L Q I I L K C Q E L R M M L K W K Q Q K C Q
N * S I R N A G N Y R * Y * N A R N C G * C * N G N N R N A
10000 10010 10020 10030 10040 10050 10060 10070 10080
GAAATTACAGAGGATATTGAAACTGAAACATCGGAAATGCCGGAAACTACAGATAATAGTGAAATGCCAGGAACTGCAGATGATACTGAA
E I T E D I E T E T S E M P E T T D N S E M P G T A D D T E
K L Q R I L K L K H R K C R K L Q I I V K C Q E L Q M I L K
R N Y R G Y * N * N I G N A G N Y R * * * N A R N C R * Y *
10090 10100 10110 10120 10130 10140 10150 10160 10170
ACTGAAACATCGGAAATTCCGGAAACTACAGATAATACTGAAATGCCAGGAACTGCGGATGATACTGAAATGGAAACAACAGAAATGCCA
T E T S E I P E T T D N T E M P G T A D D T E M E T T E M P
L K H R K F R K L Q I I L K C Q E L R M I L K W K Q Q K C Q
N * N I G N S G N Y R * Y * N A R N C G * Y * N G N N R N A
10180 10190 10200 10210 10220 10230 10240 10250 10260
GAAATTACAGAGGATATTGAAACTGAAACATCGGAAATGCCGGAAACTACAGATAATAGTGAAATGCCAGGAACTGCAGATGATACTGAA
E I T E D I E T E T S E M P E T T D N S E M P G T A D D T E
K L Q R I L K L K H R K C R K L Q I I V K C Q E L Q M I L K
R N Y R G Y * N * N I G N A G N Y R * * * N A R N C R * Y *
10270 10280 10290 10300 10310 10320 10330 10340 10350
ACTGAAACATCGGAAATGCCGGAAACTACAGATAATACTGAAATGCCAGGAACTTCGGATGATACTGAAATGGAAACAACAGAAATGCCA
T E T S E M P E T T D N T E M P G T S D D T E M E T T E M P
L K H R K C R K L Q I I L K C Q E L R M I L K W K Q Q K C Q
N * N I G N A G N Y R * Y * N A R N F G * Y * N G N N R N A
10360 10370 10380 10390 10400 10410 10420 10430 10440
GAAATTACAGAAGATACTGAAACTGAAGCATCAGAAATGCCGGAAACTACAGAGGGGGCTGAAACTGAAACAACGACCGAAGCATCAGAA
E I T E D T E T E A S E M P E T T E G A E T E T T T E A S E
K L Q K I L K L K H Q K C R K L Q R G L K L K Q R P K H Q K
R N Y R R Y * N * S I R N A G N Y R G G * N * N N D R S I R
10450 10460 10470 10480 10490 10500 10510 10520 10530
ACTACGGAAACTATTAAACCCCACCAAAAAACCTTCTGGTTTGTTATTTTCAAGTTAAAAAAGCCACTTTTCTGTCTTTCTGATTTGGAAA
T T E T I K P T K K P S G L L F S S * K S H F S V F L I W K
L R K L L N P P K N L L V C Y F Q V K K A T F L S F * F G K
N Y G N Y * T H Q K T F W F V I F K L K K P L F C L S D L E
10540 10550 10560 10570 10580 10590 10600 10610 10620
ATATACAGTCACCTAAGTTTACTTTTTATGCACTTTAGACAGAACAAAAAATACTTACATACAGTTGAAAAGAAGCTTAAGAAATGTTCA
I Y S H L S L L F M H F R Q N K K Y L H T V E K K L K K C S
Y T V T * V Y F L C T L D R T K N T Y I Q L K R S L R N V Q
N I Q S P K F T F Y A L * T E Q K I L T Y S * K E A * E M F

```
    10630     10640     10650     10660     10670     10680     10690     10700     10710
GTACTCTGAAAGCTAAATGACTCTTCCCAAAGTAATGTTTACTAAAATTTCGCAAGTACATCTGATATTAATCGTAGAAAAATTATTTCG
 V  L  *  K  L  N  D  S  S  Q  S  N  V  Y  *  N  F  A  S  T  S  D  I  N  R  R  K  I  I  S
  Y  S  E  S  *  M  T  L  P  K  V  M  F  T  K  I  S  Q  V  H  L  I  L  I  V  E  K  L  F  R
S  T  L  K  A  K  *  L  F  P  K  *  C  L  L  K  F  R  K  Y  I  *  Y  *  S  *  K  N  Y  F

    10720     10730     10740     10750     10760     10770     10780     10790     10800
AATATTTGATATCAGAAAACCCGAAATGGTCCATAATGAGTCCGTTGAAATTGAAGAATACTTAAATGCATTTTTTTTTTAGTATTGAGC
 N  I  *  Y  Q  K  T  R  N  G  P  *  *  V  R  *  N  *  R  I  L  K  C  I  F  F  L  V  L  S
  I  F  D  I  R  K  P  E  M  V  H  N  E  S  V  E  I  E  E  Y  L  N  A  F  F  F  *  Y  *  A
E  Y  L  I  S  E  N  P  K  W  S  I  M  S  P  L  K  L  K  N  T  *  M  H  F  F  F  S  I  E

    10810     10820     10830     10840     10850     10860     10870     10880     10890
AACTGAATCATGAAAACATGTCAGTAACTGCATAGTTAAATTTGGGATTTGTACTGTATTGAACATAACATCGTTTTATGCTTTGTGTAA
 N  *  I  M  K  T  C  Q  *  L  H  S  *  I  W  D  L  Y  C  I  E  H  N  I  V  L  C  F  V  *
  T  E  S  *  K  H  V  S  N  C  I  V  K  F  G  I  C  T  V  L  N  I  T  S  F  Y  A  L  C  K
Q  L  N  H  E  N  M  S  V  T  A  *  L  N  L  G  F  V  L  Y  *  T  *  H  R  F  M  L  C  V

    10900     10910     10920     10930     10940     10950     10960     10970     10980
AACCGAATCCAAAAAAATACGAAAGGAAAAATTGTTGTTCTATTTCAAATTTCGAGCATAGTCTATATGGCAATGTGGCAAAAAAGACGT
 N  R  I  Q  K  N  T  K  G  K  I  V  V  L  F  Q  I  S  S  I  V  Y  M  A  M  W  Q  K  R  R
  T  E  S  K  K  I  R  K  E  K  L  L  F  Y  F  K  F  R  A  *  S  I  W  Q  C  G  K  K  D  V
K  P  N  P  K  K  Y  E  R  K  N  C  C  S  I  S  N  F  E  H  S  L  Y  G  N  V  A  K  K  T

    10990     11000     11010     11020     11030     11040     11050     11060     11070
CATGGTCTTTTTAAGAAAAAACTAATTTACCGAAATCAGAATTCAAATTGGTGTAAGATTTTGAAAATGAAAAAAATGAGGAGAAAATTA
 H  G  L  F  K  K  K  L  I  Y  R  N  Q  N  S  N  W  C  K  I  L  K  M  K  K  M  R  R  K  L
  M  V  F  L  R  K  N  *  F  T  E  I  R  I  Q  I  G  V  R  F  *  K  *  K  K  *  G  E  N  *
S  W  S  F  *  E  K  T  N  L  P  K  S  E  F  K  L  V  *  D  F  E  N  E  K  N  E  E  K  I

    11080     11090     11100     11110     11120     11130     11140     11150     11160
GTCCCAGTATTATTACAGAAATCTTTAAAAAGAAAGTTTTGGAATTACGGATGGAAAATATTTTAGCAAATAGTCAAAATTGCTGAGAAC
 V  P  V  L  L  Q  K  S  L  K  R  K  F  W  N  Y  G  W  K  I  F  *  Q  I  V  K  I  A  E  N
  S  Q  Y  Y  Y  R  N  L  *  K  E  S  F  G  I  T  D  G  K  Y  F  S  K  *  S  K  L  L  R  T
S  P  S  I  I  T  E  I  F  K  K  K  V  L  E  L  R  M  E  N  I  L  A  N  S  Q  N  C  *  E

    11170     11180     11190     11200     11210     11220     11230     11240     11250
AATTATCTGGACTTTCACATGCATAAAAGAGAAATTATCACGGTTACAATCACTGTTAATTGAGTAAGATTGATCGGAAATGACGATTAT
 N  Y  L  D  F  H  M  H  K  R  E  I  I  T  V  T  I  T  V  N  *  V  R  L  I  G  N  D  D  Y
  I  I  W  T  F  T  C  I  K  E  K  L  S  R  L  Q  S  L  L  I  E  *  D  *  S  E  M  T  I  I
Q  L  S  G  L  S  H  A  *  K  R  N  Y  H  G  Y  N  H  C  *  L  S  K  I  D  R  K  *  R  L

    11260     11270     11280     11290     11300     11310     11320     11330     11340
AAATTAGCCGAATGGATATTAATTGAACAAACCTAAACAAAAGGAATAACTAAGCAGGAATAAGCAAATGGATGTATAAGCAACAATGGG
 K  L  A  E  W  I  L  I  E  Q  T  *  T  K  G  I  T  K  Q  E  *  A  N  G  C  I  S  N  N  G
  N  *  P  N  G  Y  *  L  N  K  P  K  Q  K  E  *  L  S  R  N  K  Q  M  D  V  *  A  T  M  G
*  I  S  R  M  D  I  N  *  T  N  L  N  K  R  N  N  *  A  G  I  S  K  W  M  Y  K  Q  Q  W

    11350     11360     11370     11380     11390     11400     11410     11420     11430
AAATGAACAATTACAGAGTTGAAGAAATCGACTGAAAAAGTCAGAATTAGACGATGGCAATAAATGCTCAATCCAATAACCATGCTTAAT
 K  *  T  I  T  E  L  K  K  S  T  E  K  V  R  I  R  R  W  Q  *  M  L  N  P  I  T  M  L  N
  N  E  Q  L  Q  S  *  R  N  R  L  K  K  S  E  L  D  D  G  N  K  C  S  I  Q  *  P  C  L  I
E  M  N  N  Y  R  V  E  E  I  D  *  K  S  Q  N  *  T  M  A  I  N  A  Q  S  N  N  H  A  *

    11440     11450     11460     11470     11480     11490     11500     11510     11520
AAAGCCTAATTCAATAATTATGCTTCTTATGCTGAAATAATACGAAAAATTATGCCATACCGACAAAATTACTCATAATCTTTCAGAAAT
 K  A  *  F  N  N  Y  A  S  Y  A  E  I  I  R  K  I  M  P  Y  R  Q  N  Y  S  *  S  F  R  N
  K  P  N  S  I  I  M  L  L  M  L  K  *  Y  E  K  L  C  H  T  D  K  I  T  H  N  L  S  E  M
*  S  L  I  Q  *  L  C  F  L  C  *  N  N  T  K  N  Y  A  I  P  T  K  L  L  I  I  F  Q  K
```

```
    11530      11540      11550      11560      11570      11580      11590      11600      11610
GAGACAGAAATTTTGATTGCCAAAAAATAATATCGAAGTTAAATGAGCCTTAAAAGCCGATATTGTTTATCTCTCTTTACTACATATAAT
 E T E I L I A K K * Y R S * M S L K S R Y C L S L F T T Y N
  R Q K F * L P K N N I E V K * A L K A D I V Y L S L L H I I
* D R N F D C Q K I I S K L N E P * K P I L F I S L Y Y I *

    11620      11630      11640      11650      11660      11670      11680      11690      11700
TTATTCTTCTTTTTAGTCCAAAAAACTTTCTAATCTTACGATAGATAATCTAGAAACAATGAATTATCTTGTTATGATATTGAAATCTAT
 L F F F L V Q K T F * S Y D R * S R N N E L S C Y D I E I Y
  Y S S F * S K K L S N L T I D N L E T M N Y L V M I L K S I
F I L L F S P K N F L I L R * I I * K Q * I I L L * Y * N L

    11710      11720      11730      11740      11750      11760      11770      11780      11790
TTCAGAGTTGAAATGTCCCCGGCCTTCAGGTCTGTTTCCGCATCCCGACGATTGCCATTTGTTCCTTCATTGTGCACACAACTATCCACA
 F R V E M S P A F R S V S A S R R L P F V P S L C T Q L S T
  S E L K C P R P S G L F P H P D D C H L F L H C A H N Y P H
F Q S * N V P G L Q V C F R I P T I A I C S F I V H T T I H

    11800      11810      11820      11830      11840      11850      11860      11870      11880
TGTGATGGAATGTCCGGCAGCAACTTTCTTCAATGAGAAATACAAAGTTTGCGATCATCAGAGAAATGCTCCAGAAGGATGCGTATGATT
 C D G M S G S N F L Q * E I Q S L R S S E K C S R R M R M I
  V M E C P A A T F F N E K Y K V C D H Q R N A P E G C V * F
M * W N V R Q Q L S S M R N T K F A I I R E M L Q K D A Y D

    11890      11900      11910      11920      11930      11940      11950      11960      11970
TGTATTATTTCCATGATAATATATGCAATTGTACATGTTGTTGTAAAAAAAAATAAGAATTTGAAGTAAAATTTTTTGGCAGCACAATGAA
 C I I S M I I Y A I V H V V V K K I R I * S K I F W Q H N E
  V L F P * * Y M Q L Y M L L * K K * E F E V K F F G S T M N
L Y Y F H D N I C N C T C C C K K N K N L K * N F L A A Q *

    11980      11990      12000      12010      12020      12030      12040      12050      12060
TTGGTACTTACTCGTATGAAAATTTGTGATATGTTCAAAAAATTATAAAGATAGAAGTAAAGAAATGAGTCAACTGAATCATTTAAAATG
 L V L T R M K I C D M F K K L * R * K * R N E S T E S F K M
  W Y L L V * K F V I C S K N Y K D R S K E M S Q L N H L K *
I G T Y S Y E N L * Y V Q K I I K I E V K K * V N * I I * N

    12070      12080      12090      12100      12110      12120      12130      12140      12150
AAGATATTTGTTATTTGCATTTCTTCACATCTCAGCAAATCCATAAATAACTTCAGATAGACATTAATGCATTCAACGAACGGTAATTGA
 K I F V I C I S S H L S K S I N N F R * T L M H S T N G N *
  R Y L L F A F L H I S A N P * I T S D R H * C I Q R T V I D
E D I C Y L H F F T S Q Q I H K * L Q I D I N A F N E R * L

    12160      12170      12180      12190      12200      12210      12220      12230      12240
TAATCAAAAAAAAAAAAAGTAGAGATAAATGCTTAAAAATACCACTAGCTTTGCACCATAAGTTAAAATTTTACTCAAGAATCAATCCATT
 * S K K K K * R * M L K N T T S F A P * V K I L L K N Q S I
  N Q K K K S R D K C L K I P L A L H H K L K F Y S R I N P L
I I K K K K V E I N A * K Y H * L C T I S * N F T Q E S I H

    12250      12260      12270      12280      12290      12300      12310      12320      12330
GAAAAATTTATTCCATTTCCGATACTTTTAGCTGTGACATCACGCTAATACATAATGGTTTGAAGCCACCTTCCATGTACATTTTGAAAT
 E K F I P F P I L L A V T S R * Y I M V * S H L P C T F * N
  K N L F H F R Y F * L * H H A N T * W F E A T F H V H F E I
* K I Y S I S D T F S C D I T L I H N G L K P P S M Y I L K

    12340      12350      12360      12370      12380      12390      12400      12410      12420
TTGCTGTTTATGAAATAATTTCTACTCTCTTTTCTTAGCGCCTGAATTGCAAAAAAATAATTTCTCCATATTATTCACTCATATTTTCTT
 L L F M K * F L L S F L S A * I A K K * F L H I I H S Y F L
  C C L * N N F Y S L F L A P E L Q K N N F S I L F T H I F F
F A V Y E I I S T L F S * R L N C K K I I S P Y Y S L I F S
```

```
    12430     12440     12450     12460     12470     12480     12490     12500     12510
CGGCAGCTTCCTGGAATGTAAAACATTCCTTTTCTCTTCTACGACTATTCGGCCTAATTTTTTTCCTTTTTTCCTCTATGACTATTCGGC
 R  Q  L  P  G  M  *  N  I  P  F  L  F  Y  D  Y  S  A  *  F  F  S  F  F  P  L  *  L  F  G
  G  S  F  L  E  C  K  T  F  L  F  S  S  T  T  I  R  P  N  F  F  P  F  F  L  Y  D  Y  S  A
S  A  A  S  W  N  V  K  H  S  F  S  L  L  R  L  F  G  L  I  F  F  L  F  S  S  M  T  I  R

    12520     12530     12540     12550     12560     12570     12580     12590     12600
TTAAAAATTATAATAGAAATGATGTGAGCACATATGTCAAGTCAACGGATAGTGAAATTCTGCTCGTTTTCTGTTTCACGTATAAATGAA
 L  K  I  I  I  E  M  M  *  A  H  M  S  S  Q  R  I  V  K  F  C  S  F  S  V  S  R  I  N  E
  *  K  L  *  *  K  *  C  E  H  I  C  Q  V  N  G  *  *  N  S  A  R  F  L  F  H  V  *  M  K
L  K  N  Y  N  R  N  D  V  S  T  Y  V  K  S  T  D  S  E  I  L  L  V  F  C  F  T  Y  K  *

    12610     12620     12630     12640     12650     12660     12670     12680     12690
GAAATGTTACGCCACATCTTTTTATTATCCATTTGACGATCAATATTACCGAAAATAAAGTTATTGGTAAATTGAGAATTGAAGCATTAT
 E  M  L  R  H  I  F  L  L  S  I  *  R  S  I  L  P  K  I  K  L  L  V  N  *  E  L  K  H  Y
  K  C  Y  A  T  S  F  Y  Y  P  F  D  D  Q  Y  Y  R  K  *  S  Y  W  *  I  E  N  *  S  I  I
R  N  V  T  P  H  L  F  I  I  H  L  T  I  N  I  T  E  N  K  V  I  G  K  L  R  I  E  A  L

    12700     12710     12720     12730     12740     12750     12760     12770     12780
TAATTAGCTCATTCATCGTTCATCAACAAGGATTACGGGACAAATTTCAAGAAATCAAAAAAGTGAAATTCTATTCGTTTTCTATTTCACG
 *  L  A  H  S  S  F  I  N  K  D  Y  G  T  N  F  K  K  S  K  S  E  I  L  F  V  F  Y  F  T
  N  *  L  I  H  R  S  S  T  R  I  T  G  Q  I  S  R  N  Q  K  V  K  F  Y  S  F  S  I  S  R
L  I  S  S  F  I  V  H  Q  Q  G  L  R  D  K  F  Q  E  I  K  K  *  N  S  I  R  F  L  F  H

    12790     12800     12810     12820     12830     12840     12850     12860     12870
TATAAAATATCACACAGAAATGAATTCTATCCAAACATGAGGCTCGAAATTTGAAATGTATATGAAATTGAGCACAATTTGGTGCGAACA
 Y  K  I  S  H  R  N  E  F  Y  P  N  M  R  L  E  I  *  N  V  Y  E  I  E  H  N  L  V  R  T
  I  K  Y  H  T  E  M  N  S  I  Q  T  *  G  S  K  F  E  M  Y  M  K  L  S  T  I  W  C  E  Q
V  *  N  I  T  Q  K  *  I  L  S  K  H  E  A  R  N  L  K  C  I  *  N  *  A  Q  F  G  A  N

    12880     12890     12900     12910     12920     12930     12940     12950     12960
ATATATTCTGATGTCGATAACTCCTGCCTTTTAGCATGCAAAATTCATTTCCGATATTAATTAGCATTTTTTTTTTATAATATGTATTAA
 I  Y  S  D  V  D  N  S  C  L  L  A  C  K  I  H  F  R  Y  *  L  A  F  F  F  Y  N  M  Y  *
  Y  I  L  M  S  I  T  P  A  F  *  H  A  K  F  I  S  D  I  N  *  H  F  F  F  I  I  C  I  N
N  I  F  *  C  R  *  L  L  P  F  S  M  Q  N  S  F  P  I  L  I  S  I  F  F  L  *  Y  V  L

    12970     12980     12990     13000     13010     13020     13030     13040     13050
TGGAAGAAGTCTATCTTAAAAATTTTTTTACTTCTTATTTTTTACTTACTGTAAAGTTTTTTAGTTTGTGGGGGGGGGGGGGGGAAAACGT
 W  K  K  S  I  L  K  I  F  Y  F  L  F  F  T  Y  C  K  V  F  *  F  V  G  G  G  G  G  K  R
  G  R  S  L  S  *  K  F  F  T  S  Y  F  L  L  T  V  K  F  F  S  L  W  G  G  G  G  E  N  V
M  E  E  V  Y  L  K  N  F  L  L  L  I  F  Y  L  L  *  S  F  L  V  C  G  G  G  G  G  K  T

    13060     13070     13080     13090     13100     13110     13120     13130     13140
TTATACTGAACAATCAAATATTTGATAAAGATTGATTTGCTTATGGAGAAGTTTTTGTTTTAGGACGTTACACACCTGTCAAAATAAAAG
 L  Y  *  T  I  K  Y  L  I  K  I  D  L  L  M  E  K  F  L  F  *  D  V  T  H  L  S  K  *  K
  Y  T  E  Q  S  N  I  *  *  R  L  I  C  L  W  R  S  F  C  F  R  T  L  H  T  C  Q  N  K  S
F  I  L  N  N  Q  I  F  D  K  D  *  F  A  Y  G  E  V  F  V  L  G  R  Y  T  P  V  K  I  K

    13150     13160     13170     13180     13190     13200     13210     13220     13230
TAATCAACTAATAGCTAACAAGAATTATGTGTACTTAATATTTTTATTATTATTCTCCAAAAATTATTCTCCACACATATGCAAATGAAT
 *  S  T  N  S  *  Q  E  L  C  V  L  N  I  F  I  I  I  L  Q  K  L  F  S  T  H  M  Q  M  N
  N  Q  L  I  A  N  K  N  Y  V  Y  L  I  F  L  L  L  F  S  K  N  Y  S  P  H  I  C  K  *  I
V  I  N  *  *  L  T  R  I  M  C  T  *  Y  F  Y  Y  Y  S  P  K  I  I  L  H  T  Y  A  N  E

    13240     13250     13260     13270     13280     13290     13300     13310     13320
TTATTTATGATCAAAGGCGAGCTTCAATTATTTCTAACGATTCCTATCACTGGATGGTGTGTACAATGTACGGAATTTTCAGCATTATGA
 L  F  M  I  K  G  E  L  Q  L  F  L  T  I  P  I  T  G  W  C  V  Q  C  T  E  F  S  A  L  *
  Y  L  *  S  K  A  S  F  N  Y  F  *  R  F  L  S  L  D  G  V  Y  N  V  R  N  F  Q  H  Y  D
F  I  Y  D  Q  R  R  A  S  I  I  S  N  D  S  Y  H  W  M  V  C  T  M  Y  G  I  F  S  I  M
```

```
    13330     13340     13350     13360     13370     13380     13390     13400     13410
TAAAAGATATGATGAAAACTTTTAATAATAAAAACTCTCGACAAATATAATAGCATTATTACTTATATTTTTCCAGCTTTCCGTCTAATT
 *  K  I  *  *  K  L  L  I  I  K  T  L  D  K  Y  N  S  I  I  T  Y  I  F  P  A  F  R  L  I
  K  R  Y  D  E  N  F  *  *  *  K  L  S  T  N  I  I  A  L  L  L  I  F  F  Q  L  S  V  *  F
I  K  D  M  M  K  T  F  N  N  K  N  S  R  Q  I  *  *  H  Y  Y  L  Y  F  S  S  F  P  S  N

    13420     13430     13440     13450     13460     13470     13480     13490     13500
TTAATGTTTTTTCAGCTTTCGAGTTATTTGTTTTTTTCCAGCTTTCCAGCTATTTCTTGTTTTTTCTAACTTCTTAATATTATTTATCTC
 L  M  F  F  Q  L  S  S  Y  L  F  F  S  S  F  P  A  I  S  C  F  F  *  L  L  N  I  I  Y  L
  *  C  F  F  S  F  R  V  I  C  F  F  P  A  F  Q  L  F  L  V  F  S  N  F  L  I  L  F  I  S
F  N  V  F  S  A  F  E  L  F  V  F  F  Q  L  S  S  Y  F  L  F  F  L  T  S  *  Y  Y  L  S

    13510     13520     13530     13540     13550     13560     13570     13580     13590
ATTCCAACTTCGGATTTTAATTAATCAATTTAATTAATCTAATATCTTCTAATCAAAAATATTTTTCATTACATGACTTGTAGCTGAAAG
 I  P  T  S  D  F  N  *  S  I  *  L  I  *  Y  L  L  I  K  N  I  F  H  Y  M  T  C  S  *  K
  F  Q  L  R  I  L  I  N  Q  F  N  *  S  N  I  F  *  S  K  I  F  F  I  T  *  L  V  A  E  R
H  S  N  F  G  F  *  L  I  N  L  I  N  L  I  S  S  N  Q  K  Y  F  S  L  H  D  L  *  L  K

    13600     13610     13620     13630     13640     13650     13660     13670     13680
GCCCTTGCATATGAAAAAGTTTTGTTTGGGGTTTATAGGACAATTATCAGAGTCTTATCAATAATACAGTAGATTACTTTCTCGATACTC
 A  L  A  Y  E  K  V  L  F  G  V  Y  R  T  I  I  R  V  L  S  I  I  Q  *  I  T  F  S  I  L
  P  L  H  M  K  K  F  C  L  G  F  I  G  Q  L  S  E  S  Y  Q  *  Y  S  R  L  L  S  R  Y  S
G  P  C  I  *  K  S  F  V  W  G  L  *  D  N  Y  Q  S  L  I  N  N  T  V  D  Y  F  L  D  T

    13690     13700     13710     13720     13730     13740     13750     13760     13770
GAAAAATGTTGTTTTTTTCCCATTTTCTCATCGGAGTGAAATTTTTTTGAAAGCATATGATTGTTTGTTTGATTGACAAATACTTTGCAG
 E  K  C  C  F  F  P  I  F  S  S  E  *  N  F  F  E  S  I  *  L  F  V  *  L  T  N  T  L  Q
  K  N  V  V  F  F  P  F  S  H  R  S  E  I  F  L  K  A  Y  D  C  L  F  D  *  Q  I  L  C  R
R  K  M  L  F  F  S  H  F  L  I  G  V  K  F  F  *  K  H  M  I  V  C  L  I  D  K  Y  F  A

    13780     13790     13800     13810     13820     13830     13840     13850     13860
AATGCTTATTTGAATATTCTGGAAAAAAGCCTGATCCGAATTCGTTCCCTTTAGTGAGGGTTAATTCCGCGGCCGCTCTAGAACTAGTGG
 N  A  Y  L  N  I  L  E  K  S  L  I  R  I  R  S  L  *  *  G  L  I  P  R  P  L  *  N  *  W
  M  L  I  *  I  F  W  K  K  A  *  S  E  F  V  P  F  S  E  G  *  F  R  G  R  S  R  T  S  G
E  C  L  F  E  Y  S  G  K  K  P  D  P  N  S  F  P  L  V  R  V  N  S  A  A  A  L  E  L  V

    13870     13880     13890     13900     13910     13920     13930     13940     13950
ATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGT
 I  P  R  A  A  G  I  R  Y  Q  A  Y  R  Y  R  R  P  R  G  G  A  R  Y  P  I  R  P  I  V  S
  S  P  G  L  Q  E  F  D  I  K  L  I  D  T  V  D  L  E  G  G  P  G  T  Q  F  A  L  *  *  V
D  P  P  G  C  R  N  S  I  S  S  L  S  I  P  S  T  S  R  G  G  P  V  P  N  S  P  Y  S  E

    13960
CGTATTACAATTC
 R  I  T  I  X
  V  L  Q  F
S  Y  Y  N  S
```

Appendix 2: Insert of Clone 9; Sequence Range: 1 to 13753

Gene II: 3952…8592

Gene III: 12096…13750

```
       10        20        30        40        50        60        70        80        90
CAGTGAATTGTAATACGACTCACTATAGGGCGAATTGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATT
 Q  *  I  V  I  R  L  T  I  G  R  I  G  Y  R  A  P  P  R  G  R  R  Y  R  *  A  *  Y  R  I
  S  E  L  *  Y  D  S  L  *  G  E  L  G  T  G  P  P  L  E  V  D  G  I  D  K  L  D  I  E  F
   V  N  C  N  T  T  H  Y  R  A  N  W  V  P  G  P  P  S  R  S  T  V  S  I  S  L  I  S  N

      100       110       120       130       140       150       160       170       180
CCTGCAGCCCGGGGGATCCACTAGTTCTAGAGCGGCCGCGGAATTAACCCTCACTAAAGGGAACGAATTCGGATCAAAAAAATCGACGAA
 P  A  A  R  G  I  H  *  F  *  S  G  R  G  I  N  P  H  *  R  E  R  I  R  I  K  K  I  D  E
  L  Q  P  G  G  S  T  S  S  R  A  A  A  E  L  T  L  T  K  G  N  E  F  G  S  K  K  S  T  K
S  C  S  P  G  D  P  L  V  L  E  R  P  R  N  *  P  S  L  K  G  T  N  S  D  Q  K  N  R  R

      190       200       210       220       230       240       250       260       270
ATTAATTGATTATGATATCAGTACAACTAACGTAGATAATCATACAATATAATCATATAATATAATCACATAAATATAATCACACAATAT
 I  N  *  L  *  Y  Q  Y  N  *  R  R  *  S  Y  N  I  I  I  *  Y  N  H  I  N  I  I  T  Q  Y
  L  I  D  Y  D  I  S  T  T  N  V  D  N  H  T  I  *  S  Y  N  I  I  T  *  I  *  S  H  N  I
N  *  L  I  M  I  S  V  Q  L  T  *  I  I  I  Q  Y  N  H  I  I  *  S  H  K  Y  N  H  T  I

      280       290       300       310       320       330       340       350       360
CTATAATATCTCTACATTCCCATTGTCTTTGATGCGCAGGTTCTCAAAAAGGACAAAAAAAAAAGCTTTGAATCACTAAATGATGCCGAAA
 L  *  Y  L  Y  I  P  I  V  F  D  A  Q  V  L  K  K  D  K  K  K  A  L  N  H  *  M  M  P  K
  Y  N  I  S  T  F  P  L  S  L  M  R  R  F  S  K  R  T  K  K  K  L  *  I  T  K  *  C  R  K
S  I  I  S  L  H  S  H  C  L  *  C  A  G  S  Q  K  G  Q  K  K  S  F  E  S  L  N  D  A  E

      370       380       390       400       410       420       430       440       450
AACATTAAATTAATAAGATTTAAAAGTATACCTCTTTTTGGGACAGTAATAAAAAAATGACAATGTTCCTGTAGGTAAGAATAGTCAACTG
 N  I  K  L  I  R  F  K  S  I  P  L  F  G  T  V  I  K  N  D  N  V  P  V  G  K  N  S  Q  L
  T  L  N  *  *  D  L  K  V  Y  L  F  L  G  Q  *  *  K  M  T  M  F  L  *  V  R  I  V  N  *
K  H  *  I  N  K  I  *  K  Y  T  S  F  W  D  S  N  K  K  *  Q  C  S  C  R  *  E  *  S  T

      460       470       480       490       500       510       520       530       540
ATTCATGTTCATCGTCATAACTATCTCGATCATCGTAGATTGTATTTCGATCATCGTTGTCACCTCTCCATAACGCTCTCAACATCTTTC
 I  H  V  H  R  H  N  Y  L  D  H  R  R  L  Y  F  D  H  R  C  H  L  S  I  T  L  S  T  S  F
  F  M  F  I  V  I  T  I  S  I  I  V  D  C  I  S  I  I  V  V  T  S  P  *  R  S  Q  H  L  S
D  S  C  S  S  S  *  L  S  R  S  S  *  I  V  F  R  S  S  L  S  P  L  H  N  A  L  N  I  F

      550       560       570       580       590       600       610       620       630
TCTAACGAATGTATGGTTACAATGTTATTATAATAATAAATAATAAAAGCATAAATAAGTTATACTATAAAAATGTTTATGTTTTGATCG
 S  N  E  C  M  V  T  M  L  L  *  *  *  I  I  K  A  *  I  S  Y  T  I  K  M  F  M  F  *  S
  L  T  N  V  W  L  Q  C  Y  Y  N  N  K  *  *  K  H  K  *  V  I  L  *  K  C  L  C  F  D  R
L  *  R  M  Y  G  Y  N  V  I  I  I  I  N  N  K  S  I  N  K  L  Y  Y  K  N  V  Y  V  L  I

      640       650       660       670       680       690       700       710       720
ACTTATTAGGAAAAATCCCGGATTAGAACCCTCAATCGTGTAAGAAAAACATCCACTCCTGACGTGCTAATAGGCAATATTATTTTCTAC
 T  Y  *  E  K  S  R  I  R  T  L  N  R  V  R  K  T  S  T  P  D  V  L  I  G  N  I  I  F  Y
  L  I  R  K  N  P  G  L  E  P  S  I  V  *  E  K  H  P  L  L  T  C  *  *  A  I  L  F  S  T
D  L  L  G  K  I  P  D  *  N  P  Q  S  C  K  K  N  I  H  S  *  R  A  N  R  Q  Y  Y  F  L

      730       740       750       760       770       780       790       800       810
AATAATAATATGGCCCTATCATGAAAACTATCATCCTAAATTGAAAAGCAACGTTTGTTGGAAATGCAAATACAAACGGCAATATAAGTA
 N  N  N  M  A  L  S  *  K  L  S  S  *  I  E  K  Q  R  L  L  E  M  Q  I  Q  T  A  I  *  V
  I  I  I  W  P  Y  H  E  N  Y  H  P  K  L  K  S  N  V  C  W  K  C  K  Y  K  R  Q  Y  K  *
```

```
Q * * Y G P I M K T I I L N * K A T F V G N A N T N G N I S

      820       830       840       850       860       870       880       890       900
GAATATATTCATTAATAATTGAGAATTCTTAAGAGAATACTAATTAAGATATGTGAAACTTTTTTTGTTTACTTTTCATTTAAATCATAT
 E Y I H * * L R I L K R I L I K I C E T F F V Y F S F K S Y
  N I F I N N * E F L R E Y * L R Y V K L F L F T F H L N H I
R I Y S L I I E N S * E N T N * D M * N F F C L L F I * I I

      910       920       930       940       950       960       970       980       990
TAAATTTCATTATGAAATCAGCCGTAATTTGATTATGCAATGTTTCTCCATATCTAAAAAGTTAAATTTAAATTTAAAAAGAAAAATGCT
 * I S L * N Q P * F D Y A M F L H I * K V K F K F K K K N A
  K F H Y E I S R N L I M Q C F S I S K K L N L N L K R K M L
L N F I M K S A V I * L C N V S P Y L K S * I * I * K E K C>

     1000      1010      1020      1030      1040      1050      1060      1070      1080
CCACAACTAAATGTTAGAAAAAGAATGGTGAAAAGTTATTTCATTATTTCGTAACACAGTATAGTGCTTTGCACCAAAACTAAAATAATT
 P Q L N V R K R M V K S Y F I I S * H S I V L C T K T K I I
  H N * M L E K E W * K V I S L F R N T V * C F A P K L K * F
S T T K C * K K N G E K L F H Y F V T Q Y S A L H Q N * N N

     1090      1100      1110      1120      1130      1140      1150      1160      1170
TTTCTTCTTCCTCTTCATTCCATAAATTTTCGTATTATTAATATATTATTATTATTATTAAAAACGCTTCAAATTTTGATCGAATTGAAT
 F L L P L H S I N F R I I N I L L L L L K T L Q I L I E L N
  F F F L F I P * I F V L L I Y Y Y Y Y * K R F K F * S N * I
F S S S S S F H K F S Y Y * Y I I I I I K N A S N F D R I E

     1180      1190      1200      1210      1220      1230      1240      1250      1260
TCCTAGTTTTTGCAAAAAATGATATTGATAACTATATTGAAAATGTTTTTTTAAGTGAATTTCTTTGAAGAATGTTTTATTGTTTATTGA
 S * F L Q K M I L I T I L K M F F * V N F F E E C F I V Y *
  P S F C K K * Y * * L Y * K C F F K * I S L K N V L L F I E
F L V F A K N D I D N Y I E N V F L S E F L * R M F Y C L L

     1270      1280      1290      1300      1310      1320      1330      1340      1350
ATTCGTTTTGAAGAATGCTTTATGTTTTTTAAGTGAATTTCTTTGAAGGATGTTTAACTGAAAGAGAAGAAAATCTAATGAAAGGTGTTA
 I R F E E C F M F F K * I S L K D V * L K E K K I * * K V L
  F V L K N A L C F L S E F L * R M F N * K R R K S N E R C *
N S F * R M L Y V F * V N F F E G C L T E R E E N L M K G V

     1360      1370      1380      1390      1400      1410      1420      1430      1440
AATTAATAATTATACAGAAATAAATGAAAAATCATTGCTCCAAAATCCGTTTTTTTTTAGTCTTTTTTTTTGCTCCTTTTGCTTCAATTT
 N * * L Y R N K * K I I A P K S V F F * S F F L L L L L Q F
  I N N Y T E I N E K S L L Q N P F F F S L F F C S F C F N F
K L I I I Q K * M K N H C S K I R F F L V F F F A P F A S I

     1450      1460      1470      1480      1490      1500      1510      1520      1530
TCGGAATTAAGATTAGAAACTCATTGCTATTACGATCAATATTTCTATTATGTTCGGGGTCAAATATAATTAATAATGGTCGCATTTATC
 S E L R L E T H C Y Y D Q Y F Y Y V R G Q I * L I M V A F I
  R N * D * K L I A I T I N I S I M F G V K Y N * * W S H L S
F G I K I R N S L L L R S I F L L C S G S N I I N N G R I Y

     1540      1550      1560      1570      1580      1590      1600      1610      1620
TATCGATTTCTATAATCGATAGCATAGAAATATTCAAGAAATGCGTATAAGATCAGAGTTTTTTCGAAATGATTTGAATTTTTTTCGGGA
 Y R F L * S I A * K Y S R N A Y K I R V F S K * F E F F S G
  I D F Y N R * H R N I Q E M R I R S E F F R N D L N F F R E
L S I S I I D S I E I F K K C V * D Q S F F E M I * I F F G

     1630      1640      1650      1660      1670      1680      1690      1700      1710
AATAATCTTTAAAATTTGAATTGTCATCATGTACTTACAGCATGACTTATAACGCTCTAATTTCTTGGTAGTGAATCATCAAAAAAAAGC
 N N L * N L N C H H V L T A * L I T L * F L G S E S S K K S
  I I F K I * I V I M Y L Q H D L * R S N F L V V N H Q K K A
```

```
K * S L K F E L S S C T Y S M T Y N A L I S W * * I I K K K

      1720       1730       1740       1750       1760       1770       1780       1790       1800
ACCACTGTTAGTTGGCAAGGCAAATTAATTTATCACATGAAAGCAAAAAGCGATTGAGGAATTCAGAAAAAAAGTCAAAATTTTCAGTTC
 T T V S W Q G K L I Y H M K A K S D * G I Q K K S Q N F Q F
  P L L V G K A N * F I T * K Q K A I E E F R K K V K I F S S
H H C * L A R Q I N L S H E S K K R L R N S E K K S K F S V

      1810       1820       1830       1840       1850       1860       1870       1880       1890
CAATTATTGAATGGTGTTGAAGGTTGACAGTTATCACTGCCATAGCAATGATAGATAATTGTATGATAGCATATGATAGATCATGATATG
 Q L L N G V E G * Q L S L P * Q * * I I V * * H M I D H D M
  N Y * M V L K V D S Y H C H S N D R * L Y D S I * * I M I *
P I I E W C * R L T V I T A I A M I D N C M I A Y D R S * Y

      1900       1910       1920       1930       1940       1950       1960       1970       1980
ATAATGTGAGTTGAACAGCTAATCAACTATAATGGTAACAGATAAGCGTAAATGATAAGGTAGAATAACATTTCATGAGCAGCAAATAAA
 I M * V E Q L I N Y N G N R * A * M I R * N N I S * A A N K
  * C E L N S * S T I M V T D K R K * * G R I T F H E Q Q I K
D N V S * T A N Q L * W * Q I S V N D K V E * H F M S S K *

      1990       2000       2010       2020       2030       2040       2050       2060       2070
ACAATACCGACAGTATTAACATAAAATAATAAATAAAATTATACTAACGTTATTATTGATAAATTAATAAAAGTTAATACGTAATTCTCG
 T I P T V L T * N N K * N Y T N V I I D K L I K V N T * F S
  Q Y R Q Y * H K I I N K I I L T L L L I N * * K L I R N S R
N N T D S I N I K * * I K L Y * R Y Y * * I N K S * Y V I L

      2080       2090       2100       2110       2120       2130       2140       2150       2160
ATAATCATTATTTATAATAAATAACGTCATTTAAGCAATCAATTAAAAATATTAGATGTATTTCATATCAAGTAAATGATTTTCACATTT
 I I I I Y N K * R H L S N Q L K I L D V F H I K * M I F T F
  * S L F I I N N V I * A I N * K Y * M Y F I S S K * F S H L
D N H Y L * * I T S F K Q S I K N I R C I S Y Q V N D F H I

      2170       2180       2190       2200       2210       2220       2230       2240       2250
ATAGATTTGATTGCTTATTAATATTTCCTATTATCATTATTTGTGACAACACTATCGGTTTTATCTCCGAATATTCGTTTATCCATTGTT
 I D L I A Y * Y F L L S L F V T T L S V L S P N I R L S I V
  * I * L L I N I S Y Y H Y L * Q H Y R F Y L R I F V Y P L L
Y R F D C L L I F P I I I I C D N T I G F I S E Y S F I H C

      2260       2270       2280       2290       2300       2310       2320       2330       2340
AAATAAAATTAAAATAACATTAATACCCATTGTCAATGCTATTTTAATTGTTTATGCTCGTTAATCGTTAATTTTAGTTAATTCATCTTT
 K * N * N N I N T H C Q C Y F N C L C S L I V N F S * F I F
  N K I K I T L I P I V N A I L I V Y A R * S L I L V N S S F
* I K L K * H * Y P L S M L F * L F M L V N R * F * L I H L

      2350       2360       2370       2380       2390       2400       2410       2420       2430
CGTTCTTTCTAATTTTTGGATATAATTTGTGAATCATAATTAATCAAATAATTGCTCGCTTATAACTTTAAAGAACAAAAAAATTGAGCT
 R S F * F L D I I C E S * L I K * L L A Y N F K E Q K N * A
  V L S N F W I * F V N H N * S N N C S L I T L K N K K I E L
S F F L I F G Y N L * I I I N Q I I A R L * L * R T K K L S

      2440       2450       2460       2470       2480       2490       2500       2510       2520
AAAAGTACTCTTCAACAATTTTTTTATTCATCGTTTTTAAAGTACTGCAACCTTTTATATTTGAACCGTAATTGTATTCCTCCTTTATATT
 K S T L Q Q F F I H R F * S T A T F Y I * T V I V F L L Y I
  K V L F N N F L F I V F K V L Q P F I F E P * L Y S S F I F
* K Y S S T I F Y S S F L K Y C N L L Y L N R N C I P P L Y

      2530       2540       2550       2560       2570       2580       2590       2600       2610
TGCAGTGAAATTTCAAATTATACTTTAAGCCAATTCTCAAAATTTGTCATATACGCAAAATGTCAATCAATAAACTTTCAACGTTATCTT
 C S E I S N Y T L S Q F S K F V I Y A K C Q S I N F Q R Y L
  A V K F Q I I L * A N S Q N L S Y T Q N V N Q * T F N V I F
```

```
L  Q  *  N  F  K  L  Y  F  K  P  I  L  K  I  C  H  I  R  K  M  S  I  N  K  L  S  T  L  S

      2620       2630       2640       2650       2660       2670       2680       2690       2700
TCTTCTTATCTAATATTTACTTTCAACTATTATTCACTACTTTTTTTTTCCAATTCTTTTTTGTTTATCGTTTATATCATTTAAATAATTT
 S  S  Y  L  I  F  T  F  N  Y  Y  S  L  L  F  F  P  I  L  F  C  L  S  F  I  S  F  K  *  F
  L  L  I  *  Y  L  L  S  T  I  I  H  Y  F  F  F  Q  F  F  F  V  Y  R  L  Y  H  L  N  N  F
F  F  L  S  N  I  Y  F  Q  L  L  F  T  T  F  F  S  N  S  F  L  F  I  V  Y  I  I  *  I  I

      2710       2720       2730       2740       2750       2760       2770       2780       2790
TATATTCAAGTTGAATGTTTCTAAAATTTTCAACTGTTTTTTAAAAAATTTAATAATGAAAATGATCGTACTTAGGTATGACGATAATCA
 Y  I  Q  V  E  C  F  *  N  F  Q  L  F  F  K  K  F  N  N  E  N  D  R  T  *  V  *  R  *  S
  I  F  K  L  N  V  S  K  I  F  N  C  F  L  K  N  L  I  M  K  M  I  V  L  R  Y  D  D  N  Q
L  Y  S  S  *  M  F  L  K  F  S  T  V  F  *  K  I  *  *  *  K  *  S  Y  L  G  M  T  I  I

      2800       2810       2820       2830       2840       2850       2860       2870       2880
ATTTCATTTAGCATATTTCCGTTAACTAATATTTCAGATTTTTTTTTCCTTATATTTATATAAAGCGCAGGTAAAAAAAAACTGGTTAACC
 I  S  F  S  I  F  P  L  T  N  I  S  D  F  F  F  L  I  F  I  *  S  A  G  K  K  N  W  L  T
  F  H  L  A  Y  F  R  *  L  I  F  Q  I  F  F  S  L  Y  L  Y  K  A  Q  V  K  K  T  G  *  P
N  F  I  *  H  I  S  V  N  *  Y  F  R  F  F  F  P  Y  I  Y  I  K  R  R  *  K  K  L  V  N

      2890       2900       2910       2920       2930       2940       2950       2960       2970
TTGTTAATAAATAAAATATATTTGGGAAATTTTGATATATTTGAGTTAACAAAAAAGAAGCTACATTATTTAAGGAAGACAGATGTCTAA
 L  L  I  N  K  I  Y  L  G  N  F  D  I  F  E  L  T  K  K  K  L  H  Y  L  R  K  T  D  V  *
  C  *  *  I  K  Y  I  W  E  I  L  I  Y  L  S  *  Q  K  R  S  Y  I  I  *  G  R  Q  M  S  N
L  V  N  K  *  N  I  F  G  K  F  *  Y  I  *  V  N  K  K  E  A  T  L  F  K  E  D  R  C  L

      2980       2990       3000       3010       3020       3030       3040       3050       3060
CAAATATTTTATAGTAAATATTTAATGATGCAGAAGTGGACAATAATTAACGGTTAAGTGGTATATATTTATACATATATGTATGAATAA
 Q  I  F  Y  S  K  Y  L  M  M  Q  K  W  T  I  I  N  G  *  V  V  Y  I  Y  T  Y  M  Y  E  *
  K  Y  F  I  V  N  I  *  *  C  R  S  G  Q  *  L  T  V  K  W  Y  I  F  I  H  I  C  M  N  N
T  N  I  L  *  *  I  F  N  D  A  E  V  D  N  N  *  R  L  S  G  I  Y  L  Y  I  Y  V  *  I

      3070       3080       3090       3100       3110       3120       3130       3140       3150
TTACACGATGGTACTATTAACTTTTATTACATTTGGTGTTACAATTTTTCCTATACAAGGTAAGTTAATTTTTTAATTTCTAGCACTTGT
 L  H  D  G  T  I  N  F  Y  Y  I  W  C  Y  N  F  S  Y  T  R  *  V  N  F  L  I  S  S  T  C
  Y  T  M  V  L  L  T  F  I  T  F  G  V  T  I  F  P  I  Q  G  K  L  I  F  *  F  L  A  L  V
I  T  R  W  Y  Y  *  L  L  L  H  L  V  L  Q  F  F  L  Y  K  V  S  *  F  F  N  F  *  H  L

      3160       3170       3180       3190       3200       3210       3220       3230       3240
TTAACATTAAATTTTACTAAGTTGCAAAAAAAAGGCGATAATTTGAAGTAATGTAAGAAATGTTAGCTAGACTTTTCTATGTAAAATATTA
 L  T  L  N  F  T  K  L  Q  K  K  A  I  I  *  S  N  V  R  N  V  S  *  T  F  L  C  K  I  L
  *  H  *  I  L  L  S  C  K  K  R  R  *  F  E  V  M  *  E  M  L  A  R  L  F  Y  V  K  Y  *
F  N  I  K  F  Y  *  V  A  K  K  G  D  N  L  K  *  C  K  K  C  *  L  D  F  S  M  *  N  I

      3250       3260       3270       3280       3290       3300       3310       3320       3330
GAAAGTGTAATACTAAATTGTAGTATAAAATACAAGCAAATTTTTCTCATTCATCGTTTATCTATCATACCCAGTAATATCAAATGAACG
 E  S  V  I  L  N  C  S  I  K  Y  K  Q  I  F  L  I  H  R  L  S  I  I  P  S  N  I  K  *  T
  K  V  *  Y  *  I  V  V  *  N  T  S  K  F  F  S  F  I  V  Y  L  S  Y  P  V  I  S  N  E  R
R  K  C  N  T  K  L  *  Y  K  I  Q  A  N  F  S  H  S  S  F  I  Y  H  T  Q  *  Y  Q  M  N

      3340       3350       3360       3370       3380       3390       3400       3410       3420
AAATCTTGAAACCCACTTCTCTCATTATCAAAATAGTGATAATTCACTTCAATTATTCCAACATCTTGATACCTCTTTTTAATCATTATA
 K  S  *  N  P  L  L  S  L  S  K  *  *  *  F  T  S  I  I  P  T  S  *  Y  L  F  L  I  I  I
  N  L  E  T  H  F  S  H  Y  Q  N  S  D  N  S  L  Q  L  F  Q  H  L  D  T  S  F  *  S  L  *
E  I  L  K  P  T  S  L  I  I  K  I  V  I  I  H  F  N  Y  S  N  I  L  I  P  L  F  N  H  Y

      3430       3440       3450       3460       3470       3480       3490       3500       3510
AAAATTTTGATAACTCGATTTAATCTTTATTGAAATCTTAGTAATCGCTTTAATCGCTATCAAAATCTTGATAACTCGTTTCAATCAATT
 K  I  L  I  T  R  F  N  L  Y  *  N  L  S  N  R  F  N  R  Y  Q  N  L  D  N  S  F  Q  S  I
  K  F  *  *  L  D  L  I  F  I  E  I  L  V  I  A  L  I  A  I  K  I  L  I  T  R  F  N  Q  F
```

```
K  N  F  D  N  S  I  *  S  L  L  K  S  *  *  S  L  *  S  L  S  K  S  *  *  L  V  S  I  N

      3520       3530       3540       3550       3560       3570       3580       3590       3600
TTGGTAATTTAATCATTGCTAACATATTTTTGATAATGATTGAAGAAGATGATATATAATAGCAGTTGCATGTGATTGCTATTATCGTAA
 L  V  I  *  S  L  L  T  Y  F  *  *  *  L  K  K  M  I  Y  N  S  S  C  M  *  L  L  L  S  *
  W  *  F  N  H  C  *  H  I  F  D  N  D  *  R  R  *  Y  I  I  A  V  A  C  D  C  Y  Y  R  N
F  G  N  L  I  I  A  N  I  F  L  I  M  I  E  E  D  D  I  *  *  Q  L  H  V  I  A  I  I  V

      3610       3620       3630       3640       3650       3660       3670       3680       3690
TTATATACACTTAGTGAATAAACGAAAAAATGGGATTTATGATGGAAGATTTGAAAAAAGATAGCGAAAAATTACTGTCGCATTTCGATT
 L  Y  T  L  S  E  *  T  K  K  W  D  L  *  W  K  I  *  K  K  I  A  K  N  Y  C  R  I  S  I
  Y  I  H  L  V  N  K  R  K  N  G  I  Y  D  G  R  F  E  K  R  *  R  K  I  T  V  A  F  R  F
I  I  Y  T  *  *  I  N  E  K  M  G  F  M  M  E  D  L  K  K  D  S  E  K  L  L  S  H  F  D

      3700       3710       3720       3730       3740       3750       3760       3770       3780
CCTTTTATGATAACTTCGTTAATTATCAATGCATCTTATCTTGCACCTCATATACTTCAATCACTATAGTATAACATAATATGTAGGAAA
 P  F  M  I  T  S  L  I  I  N  A  S  Y  L  A  P  H  I  L  Q  S  L  *  Y  N  I  I  C  R  K
  L  L  *  *  L  R  *  L  S  M  H  L  I  L  H  L  I  Y  F  N  H  Y  S  I  T  *  Y  V  G  N
S  F  Y  D  N  F  V  N  Y  Q  C  I  L  S  C  T  S  Y  T  S  I  T  I  V  *  H  N  M  *  E

      3790       3800       3810       3820       3830       3840       3850       3860       3870
TGTCACATAACATCAAAAGACAAATGCATTTTTTAAATTAAAAAAAACTGTGTTGAAAAAGTAATGAAAAATTAAATCATAAATAAATTT
 C  H  I  T  S  K  D  K  C  I  F  *  I  K  K  N  C  V  E  K  V  M  K  N  *  I  I  N  K  F
  V  T  *  H  Q  K  T  N  A  F  F  K  L  K  K  T  V  L  K  K  *  *  K  I  K  S  *  I  N  L
M  S  H  N  I  K  R  Q  M  H  F  L  N  *  K  K  L  C  *  K  S  N  E  K  L  N  H  K  *  I

      3880       3890       3900       3910       3920       3930       3940       3950       3960
GAATAAAAATATCAATTAAAATAAAATTTTGGAGTAAACAAAAAAATTTGGGAAAATGCGCAAATTTAATGACACTGTTTAGCAGTGAAT
 E  *  K  Y  Q  L  K  *  N  F  G  V  N  K  K  I  W  E  N  A  Q  I  *  *  H  C  L  A  V  N
  N  K  N  I  N  *  N  K  I  L  E  *  T  K  K  F  G  K  M  R  K  F  N  D  T  V  *  Q  *  I
*  I  K  I  S  I  K  I  K  F  W  S  K  Q  K  N  L  G  K  C  A  N  L  M  T  L  F  S  S  E

      3970       3980       3990       4000       4010       4020       4030       4040       4050
TATGCCGCAATTTTTGCACATTTGTCGAAATACGTTCGCGGATGTTATTACACCAATTGGGCTCAATATCGATATGGTTCAGTGATGTTT
 Y  A  A  I  F  A  H  L  S  K  Y  V  R  G  C  Y  Y  T  N  W  A  Q  Y  R  Y  G  S  V  M  F
  M  P  Q  F  L  H  I  C  R  N  T  F  A  D  V  I  T  P  I  G  L  N  I  D  M  V  Q  *  C  L
L  C  R  N  F  C  T  F  V  E  I  R  S  R  M  L  L  H  Q  L  G  S  I  S  I  W  F  S  D  V

      4060       4070       4080       4090       4100       4110       4120       4130       4140
GAAATAATTTCATTTTTAAAATTAAATAGTTGAAATAGACTCTAAGGAATGTTTACATTTTTTGAATAAAAATTAGAAAGCTTATGTAAA
 E  I  I  S  F  L  K  L  N  S  *  N  R  L  *  G  M  F  T  F  F  E  *  K  L  E  S  L  C  K
  K  *  F  H  F  *  N  *  I  V  E  I  D  S  K  E  C  L  H  F  L  N  K  N  *  K  A  Y  V  K
*  N  N  F  I  F  K  I  K  *  L  K  *  T  L  R  N  V  Y  I  F  *  I  K  I  R  K  L  M  *

      4150       4160       4170       4180       4190       4200       4210       4220       4230
GAACTTAAATCGAAAAAAAAGCAAGCTCTAAGCCATATAAAATATGAGTAGAAAATAAAAATTTTCAGAAAATTAAATAATTGACAGTTTG
 E  L  K  S  K  K  S  K  L  *  A  I  *  N  M  S  R  K  *  K  F  S  E  N  *  I  I  D  S  L
  N  L  N  R  K  K  A  S  S  K  P  Y  K  I  *  V  E  N  K  N  F  Q  K  I  K  *  L  T  V  *
R  T  *  I  E  K  K  Q  A  L  S  H  I  K  Y  E  *  K  I  K  I  F  R  K  L  N  N  *  Q  F

      4240       4250       4260       4270       4280       4290       4300       4310       4320
ACTTCCATTTTTTAGTTCTTGATTAGTTCTTACATTAATTCTCTTTCGTAATGATATATCTTGAATATACATCATTTCGACGTGTTTCTT
 T  S  I  F  *  F  L  I  S  S  Y  I  N  S  L  S  *  *  Y  I  L  N  I  H  H  F  D  V  F  L
  L  P  F  F  S  S  *  L  V  L  T  L  I  L  F  R  N  D  I  S  *  I  Y  I  I  S  T  C  F  L
D  F  H  F  L  V  L  D  *  F  L  H  *  F  S  F  V  M  I  Y  L  E  Y  T  S  F  R  R  V  S

      4330       4340       4350       4360       4370       4380       4390       4400       4410
AGCACCGTAAACTAATGTGGCGCAAGAAAAAAGTAACCAAAGTTAAGTAAAATAAGGATTGTGCTTCTACTAGTAAGTGAGTTGCAACAC
 S  T  V  N  *  C  G  A  R  K  K  *  P  K  L  S  K  I  R  I  V  L  L  L  V  S  E  L  Q  H
  A  P  *  T  N  V  A  Q  E  K  S  N  Q  S  *  V  K  *  G  L  C  F  Y  *  *  V  S  C  N  T
```

```
* H R K L M W R K K K V T K V K * N K D C A S T S K * V A T

      4420       4430       4440       4450       4460       4470       4480       4490       4500
TGATCGATGTAATATGCATTTTCATGTTAAAGGAGAAGGCAAATTCTTACCGGAGGATATTCCAAAAGGCCTGTGCACTCACATTCTTTA
 * S M * Y A F S C * R R R Q I L T G G Y S K R P V H S H S L
  D R C N M H F H V K G E G K F L P E D I P K G L C T H I L Y
L I D V I C I F M L K E K A N S Y R R I F Q K A C A L T F F

      4510       4520       4530       4540       4550       4560       4570       4580       4590
TGCATTCGCCAAAGTGAACCAGAAAGGCACGTCATTGTAAGTTGTAAATAATATTTGCAGTGAGAAATATGCTCGAAATTTTTATTGGGC
 C I R Q S E P E R H V I V S C K * Y L Q * E I C S K F L L G
  A F A K V N Q K G T S L * V V N N I C S E K Y A R N F Y W A
M H S P K * T R K A R H C K L * I I F A V R N M L E I F I G

      4600       4610       4620       4630       4640       4650       4660       4670       4680
TATTCATATGTATCGTGAAAAATTTTTGATTGAGTAGTTTTTGTTTCATAGACAAAAATTGTTGTGAAGAGTTAATCTTTGTGGAAGATT
 Y S Y V S * K I F D * V V F V S * T K I V V K S * S L W K I
  I H M Y R E K F L I E * F L F H R Q K L L * R V N L C G R L
L F I C I V K N F * L S S F C F I D K N C C E E L I F V E D

      4690       4700       4710       4720       4730       4740       4750       4760       4770
AATGGTAATACAGTATGACAATTTTAGTGACGAAAATTTCTTGATTATTGAGATTGAAATAATCATTTCTAAGTTGGACGGAATTTGAAC
 N G N T V * Q F * * R K F L D Y * D * N N H F * V G R N L N
  M V I Q Y D N F S D E N F L I I E I E I I I S K L D G I * T
* W * Y S M T I L V T K I S * L L R L K * S F L S W T E F E

      4790       4790       4800       4810       4820       4830       4840       4850       4860
GGAATTGATTGAATCTCCGACCAGTTCCGAACAGTAATTAATCGCGCAACGGAAGAGTGAGCTGTTGAATGTGACCGTTGATGTAGACAA
 G I D * I S D Q F R T V I N R A T E E * A V E C D R * C R Q
  E L I E S P T S S E Q * L I A Q R K S E L L N V T V D V D N
R N * L N L R P V P N S N * S R N G R V S C * M * P L M * T

      4870       4880       4890       4900       4910       4920       4930       4940       4950
TAGGGGTGACCACTAAGAGCAATGGACATTATATATAGCCTGGTTGGACAGTACTCGGTGTGCACTTAATGTGACAGACAGTTTTGTATT
 * G * P L R A M D I I Y S L V G Q Y S V C T * C D R Q F C I
  R G D H * E Q W T L Y I A W L D S T R C A L N V T D S F V F
I G V T T K S N G H Y I * P G W T V L G V H L M * Q T V L Y

      4960       4970       4980       4990       5000       5010       5020       5030       5040
TGAACAAAGAAATGACAGTCATTCGATGGATGAACAAAGGATTATGCAAATTCAGTTGAGAGATGAAACGAATTGCAACAAAGAATCGTC
 * T K K * Q S F D G * T K D Y A N S V E R * N E L Q Q R I V
  E Q R N D S H S M D E Q R I M Q I Q L R D E T N C N K E S S
L N K E M T V I R W M N K G L C K F S * E M K R I A T K N R

      5050       5060       5070       5080       5090       5100       5110       5120       5130
TACACATGCAACAGACAAAAAATTCCTTCTCAACTATTTAGACATAGAGCAGATAAAGCCAATGGCTAACTAGTGATTACATAGTTCTTA
 Y T C N R Q K I P S Q L F R H R A D K A N G * L V I T * F L
  T H A T D K K F L L N Y L D I E Q I K P M A N * * L H S S *
L H M Q Q T K N S F S T I * T * S R * S Q W L T S D Y I V L

      5140       5150       5160       5170       5180       5190       5200       5210       5220
ACACAATTTATCACATTTTTCGGCTTTCGCCGACAAAGTAGTCACAGATTTGAGGCATCCATATGCGGCATTTAAGCCGATTTTGTGTAA
 T Q F I T F F G F R R Q S S H R F E A S I C G I * A D F V *
  H N L S H F S A F A D K V V T D L R H P Y A A F K P I L C N
N T I Y H I F R L S P T K * S Q I * G I H M R H L S R F C V

      5230       5240       5250       5260       5270       5280       5290       5300       5310
TATATTCTGTGACGATCACTACCAGCAAATATTATTTCTGTAAATTTAACATTTAAATAACTTTTAATCTTAGGGCATTTGAATGGAATG
 Y I L * R S L P A N I I S V N L T F K * L L I L G H L N G M
  I F C D D H Y Q Q I L F L * I * H L N N F * S * G I * M E *
```

```
I  Y  S  V  T  I  T  T  S  K  Y  Y  F  C  K  F  N  I  *  I  T  F  N  L  R  A  F  E  W  N

      5320       5330       5340       5350       5360       5370       5380       5390       5400
ATGAAGATACCGAATGGTCAAAAGGGATGTACTCACGTGTAATAAAACTGAGAGAGAACGATCCAACACTGAAAATCCTTCTTTCTTATG
 M  K  I  P  N  G  Q  K  G  C  T  H  V  *  *  N  *  E  R  T  I  Q  H  *  K  S  F  F  L  M
  *  R  Y  R  M  V  K  R  D  V  L  T  C  N  K  T  E  R  E  R  S  N  T  E  N  P  S  F  L  W
D  E  D  T  E  W  S  K  G  M  Y  S  R  V  I  K  L  R  E  N  D  P  T  L  K  I  L  L  S  Y

      5410       5420       5430       5440       5450       5460       5470       5480       5490
GTGGCTATAACTTCGGATCGTCAACTTTTACTGTTTGTTTCATTTTTTTAAAGACTTACACTCGCTCAACTCATGAATTTTCTTTTGATT
 V  A  I  T  S  D  R  Q  L  L  L  F  V  S  F  F  *  R  L  T  L  A  Q  L  M  N  F  L  L  I
  W  L  *  L  R  I  V  N  F  Y  C  L  F  H  F  F  K  D  L  H  S  L  N  S  *  I  F  F  *  F
G  G  Y  N  F  G  S  S  T  F  T  V  C  F  I  F  L  K  T  Y  T  R  S  T  H  E  F  S  F  D

      5500       5510       5520       5530       5540       5550       5560       5570       5580
CGGGAAAGCAGATTGGTATTCCTCGATAATTATATTGAACAGGTTATTCGGAATAGAGTGGAAATATTACGGGAGAACAGTAATGAATAG
 R  E  S  R  L  V  F  L  D  N  Y  I  E  Q  V  I  R  N  R  V  E  I  L  R  E  N  S  N  E  *
  G  K  A  D  W  Y  S  S  I  I  I  L  N  R  L  F  G  I  E  W  K  Y  Y  G  R  T  V  M  N  S
S  G  K  Q  I  G  I  P  R  *  L  Y  *  T  G  Y  S  E  *  S  G  N  I  T  G  E  Q  *  *  I

      5590       5600       5610       5620       5630       5640       5650       5660       5670
TACTATTTCTGACGATTGATTTGTGAAACTTTGACGAGTCACTTTTCCCTTTTGATCTGTATATATATATAGAGTGTCCAACGTAACTAT
 Y  Y  F  *  R  L  I  C  E  T  L  T  S  H  F  S  L  L  I  C  I  Y  I  *  S  V  Q  R  N  Y
  T  I  S  D  D  *  F  V  K  L  *  R  V  T  F  P  F  *  S  V  Y  I  Y  R  V  S  N  V  T  I
V  L  F  L  T  I  D  L  *  N  F  D  E  S  L  F  P  F  D  L  Y  I  Y  I  E  C  P  T  *  L

      5680       5690       5700       5710       5720       5730       5740       5750       5760
ACATATACCAAATATTGGGCGATAACAGGATAGTTATTTAGATACATAAAATCAACATAAAGCACGTAGAGATACGTGTGATAAGTCACA
 T  Y  T  K  Y  W  A  I  T  G  *  L  F  R  Y  I  K  S  T  *  S  T  *  R  Y  V  *  *  V  T
  H  I  P  N  I  G  R  *  Q  D  S  Y  L  D  T  *  N  Q  H  K  A  R  R  D  T  C  D  K  S  Q
Y  I  Y  Q  I  L  G  D  N  R  I  V  I  *  I  H  K  I  N  I  K  H  V  E  I  R  V  I  S  H

      5770       5780       5790       5800       5810       5820       5830       5840       5850
AAGGATCATACAGATATCAGTCATATAATAAATCAATAGAAATCACGTAGAGATACGCGTGATAAGTTACAAGAATGACGCAGATTTGCG
 K  D  H  T  D  I  S  H  I  I  N  Q  *  K  S  R  R  D  T  R  D  K  L  Q  E  *  R  R  F  A
  R  I  I  Q  I  S  V  I  *  *  I  N  R  N  H  V  E  I  R  V  I  S  Y  K  N  D  A  D  L  R
K  G  S  Y  R  Y  Q  S  Y  N  K  S  I  E  I  T  *  R  Y  A  *  *  V  T  R  M  T  Q  I  C

      5860       5870       5880       5890       5900       5910       5920       5930       5940
CAGGATAAATGGGTAGCCGACATGCAGACATATTACAGATAATAGAATAGTTAGGCAATGTAAATGAAAAGAAAGAAAGAGCGATGGGGA
 Q  D  K  W  V  A  D  M  Q  T  Y  Y  R  *  *  N  S  *  A  M  *  M  K  R  K  K  E  R  W  G
  R  I  N  G  *  P  T  C  R  H  I  T  D  N  R  I  V  R  Q  C  K  *  K  E  R  K  S  D  G  E
A  G  *  M  G  S  R  H  A  D  I  L  Q  I  I  E  *  L  G  N  V  N  E  K  K  E  R  A  M  G

      5950       5960       5970       5980       5990       6000       6010       6020       6030
AGAAGGAGACTGTTTGATTACTCATTGGCTAGATTTGTTGTTATAGATAAAAAGAAAGATAGAGAACATGTAATAGCGGGAAATGTTGGC
 R  R  R  L  F  D  Y  S  L  A  R  F  V  V  I  D  K  K  K  D  R  E  H  V  I  A  G  N  V  G
  E  G  D  C  L  I  T  H  W  L  D  L  L  L  *  I  K  R  K  I  E  N  M  *  *  R  E  M  L  A
K  K  E  T  V  *  L  L  I  G  *  I  C  C  Y  R  *  K  E  R  *  R  T  C  N  S  G  K  C  W

      6040       6050       6060       6070       6080       6090       6100       6110       6120
GGGTAATTTACTGTGAATGACATTCAAATAGCACAAATATAAAATAACGGATGGCAGGAAATTAAATTTCACATTGAAGAAAGATTGTGA
 G  *  F  T  V  N  D  I  Q  I  A  Q  I  *  N  N  G  W  Q  E  I  K  F  H  I  E  E  R  L  *
  G  N  L  L  *  M  T  F  K  *  H  K  Y  K  I  T  D  G  R  K  L  N  F  T  L  K  K  D  C  E
R  V  I  Y  C  E  *  H  S  N  S  T  N  I  K  *  R  M  A  G  N  *  I  S  H  *  R  K  I  V

      6130       6140       6150       6160       6170       6180       6190       6200       6210
ATAGCAACTTCTTTAGCTTGGCTGTCAATTATACAAGTCATAAAACTTCGGGTTTCCCGACACTAACACATTGACAAATGATTTCAAAAA
 I  A  T  S  L  A  W  L  S  I  I  Q  V  I  K  L  R  V  S  R  H  *  H  I  D  K  *  F  Q  K
  *  Q  L  L  *  L  G  C  Q  L  Y  K  S  *  N  F  G  F  P  D  T  N  T  L  T  N  D  F  K  K
```

```
N  S  N  F  F  S  L  A  V  N  Y  T  S  H  K  T  S  G  F  P  T  L  T  H  *  Q  M  I  S  K

      6220       6230       6240       6250       6260       6270       6280       6290       6300
GAGCTGCAAACACGTGATGCATTGTATCTCAAATGTTCTCCAACTTTACGTCTCACTTTTTTCTTGCTGAAATATTTGAAGGCAATAGCA
 E  L  Q  T  R  D  A  L  Y  L  K  C  S  P  T  L  R  L  T  F  F  L  L  K  Y  L  K  A  I  A
  S  C  K  H  V  M  H  C  I  S  N  V  L  Q  L  Y  V  S  L  F  S  C  *  N  I  *  R  Q  *  Q
R  A  A  N  T  *  C  I  V  S  Q  M  F  S  N  F  T  S  H  F  F  L  A  E  I  F  E  G  N  S

      6310       6320       6330       6340       6350       6360       6370       6380       6390
AAGTCGGCGGAAAAACGGAAGCATTTCATTCAGTCAGCTACAACATTTCTTCGCAAGCATAAATTTGATGGATTCGATCTCGACTGGGAA
 K  S  A  E  K  R  K  H  F  I  Q  S  A  T  T  F  L  R  K  H  K  F  D  G  F  D  L  D  W  E
  S  R  R  K  N  G  S  I  S  F  S  Q  L  Q  H  F  F  A  S  I  N  L  M  D  S  I  S  T  G  N
K  V  G  G  K  T  E  A  F  H  S  V  S  Y  N  I  S  S  Q  A  *  I  *  W  I  R  S  R  L  G

      6400       6410       6420       6430       6440       6450       6460       6470       6480
TATCCAACCGGTGTAGCAGAGGAACATGCCAAACTTGTTGAGGTGAGTGTAATTCCATCACCGACTAACGTTTAATGTTTTGTTTACTTG
 Y  P  T  G  V  A  E  E  H  A  K  L  V  E  V  S  V  I  P  S  P  T  N  V  *  C  F  V  Y  L
  I  Q  P  V  *  Q  R  N  M  P  N  L  L  R  *  V  *  F  H  H  R  L  T  F  N  V  L  F  T  *
I  S  N  R  C  S  R  G  T  C  Q  T  C  *  G  E  C  N  S  I  T  D  *  R  L  M  F  C  L  L

      6490       6500       6510       6520       6530       6540       6550       6560       6570
AAATCTCTGGCACTATTGTAGCAATCAGATATATGCCATCAAATTTTAATTGACATTCCAAAAGAGACAATTCTCGCTTCAAAAATGAGT
 K  S  L  A  L  L  *  Q  S  D  I  C  H  Q  I  L  I  D  I  P  K  E  T  I  L  A  S  K  M  S
  N  L  W  H  Y  C  S  N  Q  I  Y  A  I  K  F  *  L  T  F  Q  K  R  Q  F  S  L  Q  K  *  V
E  I  S  G  T  I  V  A  I  R  Y  M  P  S  N  F  N  *  H  S  K  R  D  N  S  R  F  K  N  E

      6580       6590       6600       6610       6620       6630       6640       6650       6660
TGGTTTCAAAAATAAAACTGAAATATATTACATCCATCTGAAAGAGGGAGAAAAATTCAAAACGAATTTTGTAGGAAATGAAAGCGGCGT
 W  F  Q  K  *  N  *  N  I  L  H  P  S  E  R  G  R  K  I  Q  N  E  F  C  R  K  *  K  R  R
  G  F  K  N  K  T  E  I  Y  Y  I  H  L  K  E  G  E  K  F  K  T  N  F  V  G  N  E  S  G  V
L  V  S  K  I  K  L  K  Y  I  T  S  I  *  K  R  E  K  N  S  K  R  I  L  *  E  M  K  A  A

      6670       6680       6690       6700       6710       6720       6730       6740       6750
TCGTGGAAGAAGCCAAGGAATCCGGCAAGCAACAATTACTTCTTACCGCTGCTGTATCAGCCGGGAAAATGACTATTGATGAAAGTTACA
 S  W  K  K  P  R  N  P  A  S  N  N  Y  F  L  P  L  L  Y  Q  P  G  K  *  L  L  M  K  V  T
  R  G  R  S  Q  G  I  R  Q  A  T  I  T  S  Y  R  C  C  I  S  R  E  N  D  Y  *  *  K  L  Q
F  V  E  E  A  K  E  S  G  K  Q  Q  L  L  L  T  A  A  V  S  A  G  K  M  T  I  D  E  S  Y

      6760       6770       6780       6790       6800       6810       6820       6830       6840
ATGTTCAATCCCTTGGAAAGTAAAAAAAAATTGCTTTTCATATACAGACATTAAATTTGACACGCAATAAAATTTACAATACTTCAAATAT
 M  F  N  P  L  E  S  K  K  N  C  F  S  Y  T  D  I  K  F  D  T  Q  *  N  L  Q  Y  F  K  Y
  C  S  I  P  W  K  V  K  K  I  A  F  H  I  Q  T  L  N  L  T  R  N  K  I  Y  N  T  S  N  I
N  V  Q  S  L  G  K  *  K  K  L  L  F  I  Y  R  H  *  I  *  H  A  I  K  F  T  I  L  Q  I

      6850       6860       6870       6880       6890       6900       6910       6920       6930
AATTTTGTAGCATTGAAATTTTCAGAAGTTTGGATCTTCTATTTCTGATGTCGTATGATTTACACGGTAGTTGGGAGAAGAACGTCGACT
 N  F  V  A  L  K  F  S  E  V  W  I  F  Y  F  *  C  R  M  I  Y  T  V  V  G  R  R  T  S  T
  I  L  *  H  *  N  F  Q  K  F  G  S  S  I  S  D  V  V  *  F  T  R  *  L  G  E  E  R  R  L
*  F  C  S  I  E  I  F  R  S  L  D  L  L  F  L  M  S  Y  D  L  H  G  S  W  E  K  N  V  D

      6940       6950       6960       6970       6980       6990       7000       7010       7020
TACACGGAGAGTTGCGTCCGACGGAGAGAGAAACATCCGGAATTGGCGTCTGTAATACCGTTAGTTCTTGAAATAAAAATTTCTAATTTT
 Y  T  E  S  C  V  R  R  R  E  K  H  P  E  L  A  S  V  I  P  L  V  L  E  I  K  I  S  N  F
  T  R  R  V  A  S  D  G  E  R  N  I  R  N  W  R  L  *  Y  R  *  F  L  K  *  K  F  L  I  F
L  H  G  E  L  R  P  T  E  R  E  T  S  G  I  G  V  C  N  T  V  S  S  *  N  K  N  F  *  F

      7030       7040       7050       7060       7070       7080       7090       7100       7110
TTCTTGTTTGCATGAAGACGAATTCTTGATGATTTCAGGAATTTGCAGCAAATTATTGGGCAGAGAAAGGTATGCCGAAACAGAAAATTA
 F  L  F  A  *  R  R  I  L  D  D  F  R  N  L  Q  Q  I  I  G  Q  R  K  V  C  R  N  R  K  L
  S  C  L  H  E  D  E  F  L  M  I  S  G  I  C  S  K  L  L  G  R  E  R  Y  A  E  T  E  N  Y
```

```
F  L  V  C  M  K  T  N  S  *  *  F  Q  E  F  A  A  N  Y  W  A  E  K  G  M  P  K  Q  K  I

      7120       7130       7140       7150       7160       7170       7180       7190       7200
TCATTGGGATTCCGGCTTACAGTCGGGGATGGACATTAAGTAATCCTTCGGAAACGGCAATTGGAGCGGAGGGTGGTCGTCCATCATCGC
 S  L  G  F  R  L  T  V  G  D  G  H  *  V  I  L  R  K  R  Q  L  E  R  R  V  V  V  H  H  R
  H  W  D  S  G  L  Q  S  G  M  D  I  K  *  S  F  G  N  G  N  W  S  G  G  W  S  S  I  I  A
I  I  G  I  P  A  Y  S  R  G  W  T  L  S  N  P  S  E  T  A  I  G  A  E  G  G  R  P  S  S

      7210       7220       7230       7240       7250       7260       7270       7280       7290
CATCGACAACAAATCCCGCTGGCGGTACTGCGGCTTATTGGGAGGTTTAATTTAAAGTTCTCATTTACTTTTGAATATATTTCTTGCTGC
 H  R  Q  Q  I  P  L  A  V  L  R  L  I  G  R  F  N  L  K  F  S  F  T  F  E  Y  I  S  C  C
  I  D  N  K  S  R  W  R  Y  C  G  L  L  G  G  L  I  *  S  S  H  L  L  L  N  I  F  L  A  A
P  S  7  T  N  P  A  G  G  T  A  A  Y  W  E  V  *  F  K  V  L  I  Y  F  *  I  Y  F  L  L

      7300       7310       7320       7330       7340       7350       7360       7370       7380
TGATGAATTTCTCAAATTCCATTATTAATTAATGTAACCATTTTAAGATATGCAAATATGTGAAAGAAGGTGGCAAGGAAACGGTGGATA
 *  *  I  S  Q  I  P  L  L  I  N  V  T  I  L  R  Y  A  N  M  *  K  K  V  A  R  K  R  W  I
  D  E  F  L  K  F  H  Y  *  L  M  *  P  F  *  D  M  Q  I  C  E  R  R  W  Q  G  N  G  G  *
L  M  N  F  S  N  S  I  I  N  *  C  N  H  F  K  I  C  K  Y  V  K  E  G  G  K  E  7  V  D

      7390       7400       7410       7420       7430       7440       7450       7460       7470
AAAAAAGTGTTGGAGCGTATATGGTAAAAGGAGATCAGTGGCACGGTTACGATAATGAGGAAACTATTAAAATTAAAGTAATTTTTATTT
 K  K  V  L  E  R  I  W  *  K  E  I  S  G  T  V  T  I  M  R  K  L  L  K  L  K  *  F  L  F
  K  K  C  W  S  V  Y  G  K  R  R  S  V  A  R  L  R  *  *  G  N  Y  *  N  *  S  N  F  Y  F
K  K  S  V  G  A  Y  M  V  K  G  D  Q  W  H  G  Y  D  N  E  E  T  I  K  I  K  V  I  F  I

      7480       7490       7500       7510       7520       7530       7540       7550       7560
CGCTTATTGTTGTTGATAGTCAAAAATAAAAAAGTGATTTGATAGTCAAAAAATAAGAAAAGATAATAAGATAGTCGAAAAATAAAAAAT
 R  L  L  L  L  I  V  K  N  K  K  V  I  *  *  S  K  N  K  K  R  *  *  D  S  R  K  I  K  N
  A  Y  C  C  *  *  S  K  I  K  K  *  F  D  S  Q  K  I  R  K  D  N  K  I  V  E  K  *  K  I
S  L  I  V  V  D  S  Q  K  *  K  S  D  L  I  V  K  K  *  E  K  I  I  R  *  S  K  N  K  K

      7570       7580       7590       7600       7610       7620       7630       7640       7650
TAAAAAATAACGTTGCCAATATATCTTGCCTTAACTTTAATTTAAAGAGCACCATAAATTTTATAATGTTTAATAAGAAGAAATTAATTT
 *  K  I  T  L  P  I  Y  L  A  L  T  L  I  *  R  A  P  *  I  L  *  C  L  I  R  R  N  *  F
  K  K  *  R  C  Q  Y  I  L  P  *  L  *  F  K  E  H  H  K  F  Y  N  V  *  *  E  E  I  N  F
L  K  N  N  V  A  N  I  S  C  L  N  F  N  L  K  S  T  I  N  F  I  M  F  N  K  K  K  L  I

      7660       7670       7680       7690       7700       7710       7720       7730       7740
CGGAGTTTACTCGTGTAATTGTTTTACTTTAGAATTAAAATTTTCAGATGAAATGGCTGAAAGAGAAAGGTTATGGCGGTGCCTTCATAT
 R  S  L  L  V  *  L  F  Y  F  R  I  K  I  F  R  *  N  G  *  K  R  K  V  M  A  V  P  S  Y
  G  V  Y  S  C  N  C  F  T  L  E  L  K  F  S  D  E  M  A  E  R  E  R  L  W  R  C  L  H  M
S  E  F  T  R  V  I  V  L  L  *  N  *  N  F  Q  M  K  W  L  K  E  K  G  Y  G  G  A  F  I

      7750       7760       7770       7780       7790       7800       7810       7820       7830
GGGCTCTTGATTTTGATGATTTTAAGGGTACAAGTTGTGGCAAAGGTCCCTATCCACTGTTAAATGCTCTTAATAATGAGCTTCGAGACG
 G  L  L  I  L  M  I  L  R  V  Q  V  V  A  K  V  P  I  H  C  *  M  L  L  I  M  S  F  E  T
  G  S  *  F  *  *  F  *  G  Y  K  L  W  Q  R  S  L  S  T  V  K  C  S  *  *  *  A  S  R  R
W  A  L  D  F  D  D  F  K  G  T  S  C  G  K  G  P  Y  P  L  L  N  A  L  N  N  E  L  R  D

      7840       7850       7860       7870       7880       7890       7900       7910       7920
GCTTCGTATTTTGGGTGAAGGATTTTACAACTTCGTAATGATTTGTTGCATATCATAAAAATGTTATAATCGGTCGGAACTTAAAGGATT
 A  S  Y  F  G  *  R  I  L  Q  L  R  N  D  L  L  H  I  I  K  M  L  *  S  V  G  T  *  R  I
  L  R  I  L  G  E  G  F  Y  N  F  V  M  I  C  C  I  S  *  K  C  Y  N  R  S  E  L  K  G  L
G  F  V  F  W  V  K  D  F  T  T  S  *  *  F  V  A  Y  H  K  N  V  I  I  G  R  N  L  K  D

      7930       7940       7950       7960       7970       7980       7990       8000       8010
GAAATTTAATCCGGAAATCCTAAAATTCATAATGTAATGCTGTATTTTAAACAGTATTGCAAACGCAACTGGTGCATTTTTGGATATTCT
 E  I  *  S  G  N  P  K  I  H  N  V  M  L  Y  F  K  Q  Y  C  K  R  N  W  C  I  F  G  Y  S
  K  F  N  P  E  I  L  K  F  I  M  *  C  C  I  L  N  S  I  A  N  A  T  G  A  F  L  D  I  L
```

```
* N L I R K S * N S * C N A V F * T V L Q T Q L V H F W I F

      8020      8030      8040      8050      8060      8070      8080      8090      8100
TTGCTTTCTTACAAATGGTCCGCAAAAAATGAATCACGACAAGCCGAAATGTCTATTCAATTAAATTTTTATATTAATTTCGTTCTATAC
 L L S Y K W S A K N E S R Q A E M S I Q L N F Y I N F V L Y
  C F L T N G P Q K M N H D K P K C L F N * I F I L I S F Y T
F A F L Q M V R K K * I T T S R N V Y S I K F L Y * F R S I

      8110      8120      8130      8140      8150      8160      8170      8180      8190
AGATTTTAGATTTAAAAATTAATAAAGCAATCATTTTTAGTCATCAATGTTGGCAAAATCAAATAATGAAATATTGAGGAAAACAGGGAG
 R F * I * K L I K Q S F L V I N V G K I K * * N I E E N R E
  D F R F K N * * S N H F * S S M L A K S N N E I L R K T G R
Q I L D L K I N K A I I F S H Q C W Q N Q I M K Y * G K Q G

      8200      8210      8220      8230      8240      8250      8260      8270      8280
AGCCTCAGGAAGTAGTGAAATTGAAACGATCAAAAGTGGTTTTGAAACAGTACGAAGAACTTCTGGTTTGTTATCCGCAAGTTAAAATGT
 S L R K * * N * N D Q K W F * N S T K N F W F V I R K L K C
  A S G S S E I E T I K S G F E T V R R T S G L L S A S * N V
E P Q E V V K L K R S K V V L K Q Y E E L L V C Y P Q V K M

      8290      8300      8310      8320      8330      8340      8350      8360      8370
CGTTCTTGATCTGTTCTTTTGCTTTAAACACAGCGCAACTTTTCGAGTTTGCTCTTCATGCATTTTGGATAGAATGGAAATGATTGTGGA
 R S * S V L L L * T Q R N F S S L L F M H F G * N G N D C G
  V L D L F F C F K H S A T F R V C S S C I L D R M E M I V D
S F L I C S F A L N T A Q L F E F A L H A F W I E W K * L W

      8380      8390      8400      8410      8420      8430      8440      8450      8460
TAGATGAAAAGAAATTTAGGAAAATCAATGTTCTGCTGTAACATTTCAGATATAAAATGTTTGGAGTCATATGGTCTGTTCCCGCATCCT
 * M K R N L G K S M F C C N I S D I K C L E S Y G L F P H P
  R * K E I * E N Q C S A V T F Q I * N V W S H M V C S R I L
I D E K K F R K I N V L L * H F R Y K M F G V I W S V P A S

      8470      8480      8490      8500      8510      8520      8530      8540      8550
GACGATTGCCATTTATTTATTCATTGCGCAAATGACTATCCATACGTGAAACAATGTCCGCTACATACCTTTTTTAATGATGATATCAAA
 D D C H L F I H C A N D Y P Y V K Q C P L H T F F N D D I K
  T I A I Y L F I A Q M T I H T * N N V R Y I P F L M M I S K
* R L P F I Y S L R K * L S I R E T M S A T Y L F * * * Y Q

      8560      8570      8580      8590      8600      8610      8620      8630      8640
ATTTGCGATCACTTAGTAAATGCGCCGATCACATGCAAATGATTTTCATATTACCAAGAATAAAAGTTTAATGTAAAAGTTGTGGCAGTC
 I C D H L V N A P I T C K * F S Y Y Q E * K F N V K V V A V
  F A I T * * M R R S H A N D F H I T K N K S L M * K L W Q S
N L R S L S K C A D H M Q M I F I L P R I K V * C K S C G S

      8650      8660      8670      8680      8690      8700      8710      8720      8730
AAATGAATTGACCATACTCCTGTAATAATTCATATTATATAGGAGAATAAATATGACTATATGAGGACAAATGTCAAAAATGAGTCTAAC
 K * I D H T P V I I H I I * E N K Y D Y M R T N V K N E S N
  N E L T I L L * * F I L Y R R I N M T I * G Q M S K M S L T
Q M N * P Y S C N N S Y Y I G E * I * L Y E D K C Q K * V *

      8740      8750      8760      8770      8780      8790      8800      8810      8820
GGAATTATTAAAAATGAAAAAATTTGTTATTATCTGTATTTTATTTTGCGTAGCAAAAATCTTTCAACAGTTATCACCCATAACTTAGTA
 G I I K N E K I C Y Y L Y F I L R S K N L S T V I T H N L V
  E L L K M K K F V I I C I L F C V A K I F Q Q L S P I T * *
R N Y * K * K N L L L S V F Y F A * Q K S F N S Y H P * L S

      8830      8840      8850      8860      8870      8880      8890      8900      8910
GAAAACATATGAATGATTCGGAAGATGAATAAATCACATGAAAAGGATGAAATTCTGATAAAAAATAAGTACTCAATAAAAAAGGATGAA
 E N I * M I R K M N K S H E K D E I L I K N K Y S I K K D E
  K T Y E * F G R * I N H M K R M K F * * K I S T Q * K R M K
```

```
R  K  H  M  N  D  S  E  D  E  *  I  T  *  K  G  *  N  S  D  K  K  *  V  L  N  K  K  G  *

      8920       8930       8940       8950       8960       8970       8980       8990       9000
ATCATTATAAAAAGGAAATTCCGATAAAAAATAAATACTCAAAGATAAAAAATACTGTTCCATTGCATCAGGACTGGAGTATTTCCATCAA
 I  I  I  K  R  K  F  R  *  K  I  N  T  Q  R  *  K  Y  C  S  I  A  S  G  L  E  Y  F  H  Q
  S  L  *  K  G  N  S  D  K  K  *  I  L  K  D  K  N  T  V  P  L  H  Q  D  W  S  I  S  I  K
N  H  Y  K  K  E  I  P  I  K  N  K  Y  S  K  I  K  I  L  F  H  C  I  R  T  G  V  F  P  S

      9010       9020       9030       9040       9050       9060       9070       9080       9090
GAATCTTTTATTCCGTTTCCGCATTTTCACTGCTGCATAATTCCTTCCTTTTCTTTTGGCATTCCTTTTGTTTCCTTTACGACAATTCAG
 E  S  F  I  P  F  P  H  F  H  C  C  I  I  P  S  F  S  F  G  I  P  F  V  S  F  T  T  I  Q
  N  L  L  F  R  F  R  I  F  T  A  A  *  F  L  P  F  L  L  A  F  L  L  F  P  L  R  Q  F  S
R  I  F  Y  S  V  S  A  F  S  L  L  H  N  S  F  L  F  F  W  H  S  F  C  F  L  Y  D  N  S

      9100       9110       9120       9130       9140       9150       9160       9170       9180
CTTAATTTTGATAAGAGAGTTTGTAGTTGGCTAATGTGATGACCCTCAATTCGAAAAAGGAAACAGTCAAAGGCGAATTTCAATTATTTT
 L  N  F  D  K  R  V  C  S  W  L  M  *  *  P  S  I  R  K  R  K  Q  S  K  A  N  F  N  Y  F
  L  I  L  I  R  E  F  V  V  G  *  C  D  D  P  Q  F  E  K  G  N  S  Q  R  R  I  S  I  I  F
A  *  F  *  *  E  S  L  *  L  A  N  V  M  T  L  N  S  K  K  E  T  V  K  G  E  F  Q  L  F

      9190       9200       9210       9220       9230       9240       9250       9260       9270
CAACAATTCTCATCATTGAATGATATGTGCGATAGAAGGAATTTTCATAATCTCAGATGAAGAAAAGTATTTTGCAGTCTCAGATGAAGA
 Q  Q  F  S  S  L  N  D  M  C  D  R  R  N  F  H  N  L  R  *  R  K  V  F  C  S  L  R  *  R
  N  N  S  H  H  *  M  I  C  A  I  E  G  I  F  I  I  S  D  E  E  K  Y  F  A  V  S  D  E  E
S  T  I  L  I  I  E  *  Y  V  R  *  K  E  F  S  *  S  Q  M  K  K  S  I  L  Q  S  Q  M  K

      9280       9290       9300       9310       9320       9330       9340       9350       9360
AAAGTATTTTGCAGTTGCGCTTCATCTTTCTCTTCCATATTATGATAAAAGATATAATAAAAACTTCCAATAATAAACGTGGCAGCATTA
 K  V  F  C  S  C  A  S  S  F  S  S  I  L  *  *  K  I  *  *  K  L  P  I  I  N  V  A  A  L
  K  Y  F  A  V  A  L  H  L  S  L  P  Y  Y  D  K  R  Y  N  K  N  F  Q  *  *  T  W  Q  H  Y
K  S  I  L  Q  L  R  F  I  F  L  F  H  I  M  I  K  D  I  I  K  T  S  N  N  K  R  G  S  I

      9370       9380       9390       9400       9410       9420       9430       9440       9450
TTTTTTCTAGCTTTCTGCTTATTTTTAATGTTTTTCCAGCTTTCTAGCTATTTCTTGTATTTTCCCAATTTCCTAGTGTTTTTCCAGCTT
 F  F  L  A  F  C  L  F  L  M  F  F  Q  L  S  S  Y  F  L  Y  F  P  N  F  L  V  F  F  Q  L
  F  F  *  L  S  A  Y  F  *  C  F  S  S  F  L  A  I  S  C  I  F  P  I  S  *  C  F  S  S  F
I  F  S  S  F  L  L  I  F  N  V  F  P  A  F  *  L  F  L  V  F  S  Q  F  P  S  V  F  P  A

      9460       9470       9480       9490       9500       9510       9520       9530       9540
TTGTAGTTGTTTTTCTAACTCCTTAATACTATCTACGTCAGCCTAACTTCGGGATTTTAATTAACTAATTTGGTTAATCAAATATCTTCG
 L  *  L  F  F  *  L  L  N  T  I  Y  V  S  L  T  S  G  F  *  L  T  N  L  V  N  Q  I  S  S
  C  S  C  F  S  N  S  L  I  L  S  T  S  A  *  L  R  D  F  N  *  L  I  W  L  I  K  Y  L  R
F  V  V  V  F  L  T  P  *  Y  Y  L  R  Q  P  N  F  G  I  L  I  N  *  F  G  *  S  N  I  F

      9550       9560       9570       9580       9590       9600       9610       9620       9630
AATCAAATACTTTGTAGAGCGCTTATTTTTGACACTCTTAATGTTCTTAAAAAAAACCTTACTTCATTGTTATGTTTTTATGCTTAACAT
 N  Q  I  L  C  R  A  L  I  F  D  T  L  N  V  L  K  K  N  L  T  S  L  L  C  F  Y  A  *  H
  I  K  Y  F  V  E  R  L  F  L  T  L  L  M  F  L  K  K  T  L  L  H  C  Y  V  F  M  L  N  I
E  S  N  T  L  *  S  A  Y  F  *  H  S  *  C  S  *  K  K  P  Y  F  I  V  M  F  L  C  L  T

      9640       9650       9660       9670       9680       9690       9700       9710       9720
TTATTAATAACTGCAGAAATTATTTTTTTTTTTTTATTTTTGGATGTGTTCGCTGAGAACTTTTGAAAAAATTTGTTGTTATGATTTTGAT
 L  L  I  T  A  E  I  I  F  F  F  Y  F  W  M  C  S  L  R  T  F  E  K  I  C  C  Y  D  F  D
  Y  *  *  L  Q  K  L  F  F  F  F  I  F  G  C  V  R  *  E  L  L  K  K  F  V  V  M  I  L  I
F  I  N  N  C  R  N  Y  F  F  F  L  F  L  D  V  F  A  E  N  F  *  K  N  L  L  L  *  F  *

      9730       9740       9750       9760       9770       9780       9790       9800       9810
TTTTCACCGTTTTTCATTTTTTTCTCTTATATCTTATTGTGTTTTCCAGCTCTTTCGCTTTACCTAATTGCAAGATTTTAAAATATCTGA
 F  S  P  F  F  I  F  F  S  Y  I  L  L  C  F  P  A  L  S  L  Y  L  I  A  R  F  *  N  I  *
  F  H  R  F  S  F  F  S  L  I  S  Y  C  V  F  Q  L  F  R  F  T  *  L  Q  D  F  K  I  S  D
```

```
F  F  T  V  F  H  F  F  L  L  Y  L  I  V  F  S  S  S  F  A  L  P  N  C  K  I  L  K  Y  L

      9820       9830       9840       9850       9860       9870       9880       9890       9900
CGATTTACAAAATGTGCAACATTATTAATTCTTTATTGTTAATATTGTTTGCATTTTATTTTGCAACTTCCTGATTTTATTTCCAAGCTA
 R  F  T  K  C  A  T  L  L  I  L  Y  C  *  Y  C  L  H  F  I  L  Q  L  P  D  F  I  S  K  L
  D  L  Q  N  V  Q  H  Y  *  F  F  I  V  N  I  V  C  I  L  F  C  N  F  L  I  L  F  P  S  Y
T  I  Y  K  M  C  N  I  I  N  S  L  L  L  I  L  F  A  F  Y  F  A  T  S  *  F  Y  F  Q  A

      9910       9920       9930       9940       9950       9960       9970       9980       9990
TGCAAACATTTCGTATCTATAAAATCTGTTTCAAACTTAAGCCTACCAATGTTTTTTACAGTAATCCATTTTTGATTCCCATATAACTCA
 C  K  H  F  V  S  I  K  S  V  S  N  L  S  L  P  M  F  F  T  V  I  H  F  *  F  P  Y  N  S
  A  N  I  S  Y  L  *  N  L  F  Q  T  *  A  Y  Q  C  F  L  Q  *  S  I  F  D  S  H  I  T  H
M  Q  T  F  R  I  Y  K  I  C  F  K  L  K  P  T  N  V  F  Y  S  N  P  F  L  I  P  I  *  L

     10000      10010      10020      10030      10040      10050      10060      10070      10080
CCCCAATTTCATGTATCTACCGACTTTTATTTATATTTTCATTCATTCGTCCAATTTCACTTTTATTCCAATTTGAGCATATAAATATGA
 P  Q  F  H  V  S  T  D  F  Y  L  Y  F  H  S  F  V  Q  F  H  F  Y  S  N  L  S  I  *  I  *
  P  N  F  M  Y  L  P  T  F  I  Y  I  F  I  H  S  S  N  F  T  F  I  P  I  *  A  Y  K  Y  E
T  P  I  S  C  I  Y  R  L  L  F  I  F  S  F  I  R  P  I  S  L  L  F  Q  F  E  H  I  N  M

     10090      10100      10110      10120      10130      10140      10150      10160      10170
ATAAAATTAAGTATTTTGTATGATGATATTAAGATATGAAAAATTTCCAGCTTATATTTAACTTGTGATATGCACAGTATAATTCATTTT
 I  K  L  S  I  L  Y  D  D  I  K  I  *  K  I  S  S  L  Y  L  T  C  D  M  H  S  I  I  H  F
  *  N  *  V  F  C  M  M  I  L  R  Y  E  K  F  P  A  Y  I  *  L  V  I  C  T  V  *  F  I  F
N  K  I  K  Y  F  V  *  *  Y  *  D  M  K  N  F  Q  L  I  F  N  L  *  Y  A  Q  Y  N  S  F

     10180      10190      10200      10210      10220      10230      10240      10250      10260
CTGATTTCCGCACAAAGATTGAAGATGTGAAAATATTTGACCTTTACCAAAATGTGGAGCATTATTATTCATATTAATTTTTTCAGTATA
 L  I  S  A  Q  R  L  K  M  *  K  Y  L  T  F  T  K  M  W  S  I  I  I  H  I  N  F  F  S  I
  *  F  P  H  K  D  *  R  C  E  N  I  *  P  L  P  K  C  G  A  L  L  F  I  L  I  F  S  V  *
S  D  F  R  T  K  I  E  D  V  K  I  F  D  L  Y  Q  N  V  E  H  Y  Y  S  Y  *  F  F  Q  Y

     10270      10280      10290      10300      10310      10320      10330      10340      10350
ATAATGCACTGTCCCATTTATTGTCAACAAAAAAAAAAAATCCCAAACGCTGAGCTGCTTCTTAATAATATAATGTTAATAATAATTTCA
 I  M  H  C  P  I  Y  C  Q  Q  K  K  K  I  P  N  A  E  L  L  L  N  N  I  M  L  I  I  I  S
  *  C  T  V  P  F  I  V  N  K  K  K  K  S  Q  T  L  S  C  F  L  I  I  *  C  *  *  *  F  H
N  N  A  L  S  H  L  L  S  T  K  K  K  N  P  K  R  *  A  A  S  *  *  Y  N  V  N  N  N  F

     10360      10370      10380      10390      10400      10410      10420      10430      10440
TATATTTCCAGTAATTGATTATTTTATTATATTCCTTAAACATACACGTTGTCGTCTTCTTCTCTGACCGTGATAATATTTAGGGTTTGA
 Y  I  S  S  N  *  L  F  Y  Y  I  P  *  T  Y  T  L  S  S  S  S  L  T  V  I  I  F  R  V  *
  I  F  P  V  I  D  Y  F  I  I  F  L  K  H  T  R  C  R  L  L  L  *  P  *  *  Y  L  G  F  D
I  Y  F  Q  *  L  I  I  L  L  Y  S  L  N  I  H  V  V  V  F  F  S  D  R  D  N  I  *  G  L

     10450      10460      10470      10480      10490      10500      10510      10520      10530
TTGCAATCTTTCAGTTACTGTGAACTTCAAACTGGTAATTAAACCATAGCAATGATAGTAAAAACACAATATTAAAACAAGAAACAAGCG
 L  Q  S  F  S  Y  C  E  L  Q  T  G  N  *  T  I  A  M  I  V  K  T  Q  Y  *  N  K  K  Q  A
  C  N  L  S  V  T  V  N  F  K  L  V  I  K  P  *  Q  *  *  *  K  H  N  I  K  T  R  N  K  R
I  A  I  F  Q  L  L  *  T  S  N  W  *  L  N  H  S  N  D  S  K  N  T  I  L  K  Q  E  T  S

     10540      10550      10560      10570      10580      10590      10600      10610      10620
GTACAAAGAATAAATCAAAACACATAACCTCAACGGAAGAATAAGTAGAAAGAATGAAGAGATGAATGATTTGATGTCTTATATAGGCAA
 V  Q  R  I  N  Q  N  T  *  P  Q  R  K  N  K  *  K  E  *  R  D  E  *  F  D  V  L  Y  R  Q
  Y  K  E  *  I  K  T  H  N  L  N  G  R  I  S  R  K  N  E  E  M  N  D  L  M  S  Y  I  G  K
G  T  K  N  K  S  K  H  I  T  S  T  E  E  *  V  E  R  M  K  R  *  M  I  *  C  L  I  *  A

     10630      10640      10650      10660      10670      10680      10690      10700      10710
AGTATACGTGCATATCATAACGCGGGAAAATAAGGAAAGATATTGGAGGATATAAGAGCAAAGGAGATCATGTAGTACAATAATTGGAGG
 S  I  R  A  Y  H  N  A  G  K  *  G  K  I  L  E  D  I  R  A  K  E  I  M  *  Y  N  N  W  R
  V  Y  V  H  I  I  T  R  E  N  K  E  R  Y  W  R  I  *  E  Q  R  R  S  C  S  T  I  I  G  G
```

```
K  Y  T  C  I  S  *  R  G  K  I  R  K  D  I  G  G  Y  K  S  K  G  D  H  V  V  Q  *  L  E

     10720     10730     10740     10750     10760     10770     10780     10790     10800
CAAAGATGGGAACGGATGCGAGTATAGAATGGATCATTTGAATGTGTTATTTAGTTTAGACAGGAACATCGACATTTGACACGTATTAAC
 Q  R  W  E  R  M  R  V  *  N  G  S  F  E  C  V  I  *  F  R  Q  E  H  R  H  L  T  R  I  N
  K  D  G  N  G  C  E  Y  R  M  D  H  L  N  V  L  F  S  L  D  R  N  I  D  I  *  H  V  L  T
A  K  M  G  T  D  A  S  I  E  W  I  I  *  M  C  Y  L  V  *  T  G  T  S  T  F  D  T  Y  *

     10810     10820     10830     10840     10850     10860     10870     10880     10890
ATGCAGTAATATTCATTATTACAATTATGTCTATGAATAAAATATATATATATATATATATCTTTCGCGACAAAAGGAATTGCTTTTC
 M  Q  *  Y  S  L  L  Q  L  C  L  *  I  K  Y  I  Y  I  Y  I  Y  L  S  R  Q  K  E  L  L  F
  C  S  N  I  H  Y  Y  N  Y  V  Y  E  *  N  I  Y  I  Y  I  Y  I  F  R  D  K  R  N  C  F  S
H  A  V  I  F  I  I  T  I  M  S  M  N  K  I  Y  I  Y  I  Y  I  S  F  A  T  K  G  I  A  F

     10900     10910     10920     10930     10940     10950     10960     10970     10980
TCTGAAATTTTAATCCAAAACAAAAAAATTAATAAAAATTAATTCTTCTTTGGTACTTCTCGCTTTTCACGTATCTTCCAATTAGTGAAG
 S  E  I  L  I  Q  N  K  K  I  N  K  N  *  F  F  F  G  T  S  R  F  S  R  I  F  Q  L  V  K
  L  K  F  *  S  K  T  K  K  L  I  K  I  N  S  S  L  V  L  L  A  F  H  V  S  S  N  *  *  S
L  *  N  F  N  P  K  Q  K  N  *  *  K  L  I  L  L  W  Y  F  S  L  F  T  Y  L  P  I  S  E

     10990     11000     11010     11020     11030     11040     11050     11060     11070
CCATGAATGAAAACATTACCAACGCAAACGGTAAGTTGTCCAGTAAAGTGACAATTTTGGTTGGATGATATAGGTGGTATATGGAATTTT
 P  *  M  K  T  L  P  T  Q  T  V  S  C  P  V  K  *  Q  F  W  L  D  D  I  G  G  I  W  N  F
  H  E  *  K  H  Y  Q  R  K  R  *  V  V  Q  *  S  D  N  F  G  W  M  I  *  V  V  Y  G  I  F
A  M  N  E  N  I  T  N  A  N  G  K  L  S  S  K  V  T  I  L  V  G  *  Y  R  W  Y  M  E  F

     11080     11090     11100     11110     11120     11130     11140     11150     11160
CACACTCTCAGATAAATGAAAAACGTACTTTGTGTTATACGTCACATTTTCTTTTCACATTATGATAAAAGACATAACAAAATTCTCGAC
 H  T  L  R  *  M  K  N  V  L  C  V  I  R  H  I  F  F  S  H  Y  D  K  R  H  N  K  I  L  D
  T  L  S  D  K  *  K  T  Y  F  V  L  Y  V  T  F  S  F  H  I  M  I  K  D  I  T  K  F  S  T
S  H  S  Q  I  N  E  K  R  T  L  C  Y  T  S  H  F  L  F  T  L  *  *  K  T  *  Q  N  S  R

     11170     11180     11190     11200     11210     11220     11230     11240     11250
AAGTACGGTATTATTTGTGTTTTTCTAACTTTTAGGCTGTTGTTAATTTTTCCAGCTTTCCAGCTATTTACGTATATTTTTCCATCTTTT
 K  Y  G  I  I  C  V  F  L  T  F  R  L  L  L  I  F  P  A  F  Q  L  F  T  Y  I  F  P  S  F
  S  T  V  L  F  V  F  F  *  L  L  G  C  C  *  F  F  Q  L  S  S  Y  L  R  I  F  F  H  L  F
Q  V  R  Y  Y  L  C  F  S  N  F  *  A  V  V  N  F  S  S  F  P  A  I  Y  V  Y  F  S  I  F

     11260     11270     11280     11290     11300     11310     11320     11330     11340
TTGGCCATTTCTTGTGTTTTTTCAGCTTTCTCGTGTTTTTCTAACTCCTTGAATATTTTTCGTATTTTCACAGCTTTTTAGAGGAGATGG
 L  A  I  S  C  V  F  S  A  F  S  C  F  S  N  S  L  N  I  F  R  I  F  T  A  F  *  R  R  W
  W  P  F  L  V  F  F  Q  L  S  R  V  F  L  T  P  *  I  F  F  V  F  S  Q  L  F  R  G  D  G
F  G  H  F  L  C  F  F  S  F  L  V  F  F  *  L  L  E  Y  F  S  Y  F  H  S  F  L  E  E  M

     11350     11360     11370     11380     11390     11400     11410     11420     11430
TTTACATTTTCCTAAAGAACGGTACGTTTGATAAAGTACTTTATAAAAGGCCCGTTTATGGCGTTAAAGACTTCACTTTTCATTTTATAT
 F  T  F  S  *  R  T  V  R  L  I  K  Y  F  I  K  G  P  F  M  A  L  K  T  S  L  F  I  L  Y
  L  H  F  P  K  E  R  Y  V  *  *  S  T  L  *  K  A  R  L  W  R  *  R  L  H  F  S  F  Y  I
V  Y  I  F  L  K  N  G  T  F  D  K  V  L  Y  K  R  P  V  Y  G  V  K  D  F  T  F  H  F  I

     11440     11450     11460     11470     11480     11490     11500     11510     11520
TTTGCTATCTTTCAACATTCCTCCCGATTACTTTACAAAAATATCTTTTTTCTATTTTTGCACATTTATCTTTTTTTTTGAATTTCTTTC
 F  A  I  F  Q  H  S  S  R  L  L  Y  K  N  I  F  F  L  F  L  H  I  Y  L  F  F  *  I  S  F
  L  L  S  F  N  I  P  P  D  Y  F  T  K  I  S  F  F  Y  F  C  T  F  I  F  F  F  E  F  L  S
F  C  Y  L  S  T  F  L  P  I  T  L  Q  K  Y  L  F  S  I  F  A  H  L  S  F  F  L  N  F  F

     11530     11540     11550     11560     11570     11580     11590     11600     11610
TTTTTTGCTATTTTCTAGATATTTCCAGATTTTTAGTTTTGTATGCAATCTCTGAGAATGTCTTGCAATTTGCGCAAAAATAGAAAAAGA
 F  F  A  I  F  *  I  F  P  D  F  *  F  C  M  Q  S  L  R  M  S  C  N  L  R  K  N  R  K  R
  F  L  L  F  S  R  Y  F  Q  I  F  S  F  V  C  N  L  *  E  C  L  A  I  C  A  K  I  E  K  D
```

```
L  F  C  Y  F  L  D  I  S  R  F  L  V  L  Y  A  I  S  E  N  V  L  Q  F  A  Q  K  *  K  K

     11620      11630      11640      11650      11660      11670      11680      11690      11700
TATTTCATCGATATTTCTTAATTAGTCATTTACATTAGCTGCAGCATAACTCATTGAATTCCATATTTATCCCAAAGATTAGTATATCAA
 Y  F  I  D  I  S  *  L  V  I  Y  I  S  C  S  I  T  H  *  I  P  Y  L  S  Q  R  L  V  Y  Q
  I  S  S  I  F  L  N  *  S  F  T  L  A  A  A  *  L  I  E  F  H  I  Y  P  K  D  *  Y  I  N
I  F  H  R  Y  F  L  I  S  H  L  H  *  L  Q  H  N  S  L  N  S  I  F  I  P  K  I  S  I  S

     11710      11720      11730      11740      11750      11760      11770      11780      11790
CATTATATTAAACGTTGATATACTAATCTCATTTGGAACAAAATGATATCAAAAACATGAAAAATTTGCTACTTATAGAAAATACTGTAC
 H  Y  I  K  R  *  Y  T  N  L  I  W  N  K  M  I  S  K  T  *  K  I  C  Y  L  *  K  I  L  Y
  I  I  L  N  V  D  I  L  I  S  F  G  T  K  *  Y  Q  K  H  E  K  F  A  T  Y  R  K  Y  C  T
T  L  Y  *  T  L  I  Y  *  S  H  L  E  Q  N  D  I  K  N  M  K  N  L  L  L  I  E  N  T  V

     11800      11810      11820      11830      11840      11850      11860      11870      11880
ATCTAAGCTTACTTCTACTTCTCAATTAATTTTTTCTTGAATCCTTTGTATGCACCGACAGTTATTCTCACCAATCATTTCAATACAATA
 I  *  A  Y  F  Y  F  S  I  N  F  F  L  N  P  L  Y  A  P  T  V  I  L  T  N  H  F  N  T  I
  S  K  L  T  S  T  S  Q  L  I  F  S  *  I  L  C  M  H  R  Q  L  F  S  P  I  I  S  I  Q  *
H  L  S  L  L  L  L  L  N  *  F  F  L  E  S  F  V  C  T  D  S  Y  S  H  Q  S  F  Q  Y  N

     11890      11900      11910      11920      11930      11940      11950      11960      11970
ATCTCTTATTACTCTGTAAACTCAGCAAAAAAAGTATTCCTAATCGCTGAGCTATGTGTAATTAGTTACGGTAATTCCATGCATGTATAT
 I  S  Y  Y  S  V  N  S  A  K  K  V  F  L  I  A  E  L  C  V  I  S  Y  G  N  S  M  H  V  Y
  S  L  I  T  L  *  T  Q  Q  K  K  Y  S  *  S  L  S  Y  V  *  L  V  T  V  I  P  C  M  Y  I
N  L  L  L  L  C  K  L  S  K  K  S  I  P  N  R  *  A  M  C  N  *  L  R  *  F  H  A  C  I

     11980      11990      12000      12010      12020      12030      12040      12050      12060
TTAACTGATTGTTTCGCTTATTTTCTTACACATCCACTTAATATCTTCCATTAATGAAGGCAAATATCAGTTAATCATAGTAATGAATTT
 L  T  D  C  F  A  Y  F  L  T  H  P  L  N  I  F  H  *  *  R  Q  I  S  V  N  H  S  N  E  F
  *  L  I  V  S  L  I  F  L  H  I  H  L  I  S  S  I  N  E  G  K  Y  Q  L  I  I  V  M  N  F
F  N  *  L  F  R  L  F  S  Y  T  S  T  *  Y  L  P  L  M  K  A  N  I  S  *  S  *  *  *  I

     12070      12080      12090      12100      12110      12120      12130      12140      12150
TGAAATACAGAAGTGTACAAGTCCCATGAAGTGACATGAATTGGATTATGATAACGCTTTTCATCATATTCGCCAATATAATTACTGTTG
 *  N  T  E  V  Y  K  S  H  E  V  T  *  I  G  L  *  *  R  F  S  S  Y  S  P  I  *  L  L  L
  E  I  Q  K  C  T  S  P  M  K  *  H  E  L  D  Y  D  N  A  F  H  H  I  R  Q  Y  N  Y  C  C
L  K  Y  R  S  V  Q  V  P  *  S  D  M  N  W  I  M  I  T  L  F  I  I  F  A  N  I  I  T  V

     12160      12170      12180      12190      12200      12210      12220      12230      12240
CAAATGGTAAGTTGCTCAATAAAATGACACGTCAGTTGCGTAAGCACGAAATATCATCGCTGAAGATATAATGAACATGACGGTTGACAC
 Q  M  V  S  C  S  I  K  *  H  V  S  C  V  S  T  K  Y  H  R  *  R  Y  N  E  H  D  G  *  H
  K  W  *  V  A  Q  *  N  D  T  S  V  A  *  A  R  N  I  I  A  E  D  I  M  N  M  T  V  D  T
A  N  G  K  L  L  N  K  M  T  R  Q  L  R  K  H  E  I  S  S  L  K  I  *  *  T  *  R  L  T

     12250      12260      12270      12280      12290      12300      12310      12320      12330
TTCAGATCATACTACATCTTAATCATTATATGTCATTTCAATCCTAAACTGTCGAGTAGCATCTATTGAAATTCAATAGTGGCATATATT
 F  R  S  Y  Y  I  L  I  I  I  C  H  F  N  P  K  L  S  S  S  I  Y  *  N  S  I  V  A  Y  I
  S  D  H  T  T  S  *  S  L  Y  V  I  S  I  L  N  C  R  V  A  S  I  E  I  Q  *  W  H  I  L
L  Q  I  I  L  H  L  N  H  Y  M  S  F  Q  S  *  T  V  E  *  H  L  L  K  F  N  S  G  I  Y

     12340      12350      12360      12370      12380      12390      12400      12410      12420
GAATTTCACAGAAGCTCTCACAAGAAATAAGCAAATGAAGAGACTAAATTGAAGATAATACAAATTTTTACTTGAAGATTTAGTAACATT
 E  F  H  R  S  S  H  K  K  *  A  N  E  E  T  K  L  K  I  I  Q  I  F  T  *  R  F  S  N  I
  N  F  T  E  A  L  T  R  N  K  Q  M  K  R  L  N  *  R  *  Y  K  F  L  L  E  D  L  V  T  L
*  I  S  Q  K  L  S  Q  E  I  S  K  *  R  D  *  I  E  D  N  T  N  F  Y  L  K  I  *  *  H

     12430      12440      12450      12460      12470      12480      12490      12500      12510
GTATTTGAATAAATTGCAAAGAAATAACCACCAAATCGATGTTGAACATATCTTCAGGTTGTAGTCTATTTCAGGTTACGTTCGTGGATG
 V  F  E  *  I  A  K  K  *  P  P  N  R  C  *  T  Y  L  Q  V  V  V  Y  F  R  L  R  S  W  M
  Y  L  N  K  L  Q  R  N  N  H  Q  I  D  V  E  H  I  F  R  L  *  S  I  S  G  Y  V  R  G  C
```

```
C  I  *  I  N  C  K  E  I  T  T  K  S  M  L  N  I  S  S  G  C  S  L  F  Q  V  T  F  V  D

     12520      12530      12540      12550      12560      12570      12580      12590      12600
TTACTATATTAGTTGGGCTGAACGTCAACAAGGTTCAGTGACGTTTGAAATAATTTCATTATTAAAGAGCTAAATAGTTTCAAGGGGATT
 L  L  Y  *  L  G  *  T  S  T  R  F  S  D  V  *  N  N  F  I  I  K  E  L  N  S  F  K  G  I
   Y  Y  I  S  W  A  E  R  Q  Q  G  S  V  T  F  E  I  I  S  L  L  K  S  *  I  V  S  R  G  L
V  T  I  L  V  G  L  N  V  N  K  V  Q  *  R  L  K  *  F  H  Y  *  R  A  K  *  F  Q  G  D

     12610      12620      12630      12640      12650      12660      12670      12680      12690
GTGTACAAATTTTCGAATAAAAAGTAGAAAACCAATGAAATTATTTCAAGATTTCCATCTCAAAAATGAATTTTAAGCCCACGAAATATT
 V  Y  K  F  S  N  K  K  *  K  T  N  E  I  I  S  R  F  P  S  Q  K  *  I  L  S  P  R  N  I
   C  T  N  F  R  I  K  S  R  K  P  M  K  L  F  Q  D  F  H  L  K  N  E  F  *  A  H  E  I  L
C  V  Q  I  F  E  *  K  V  E  N  Q  *  N  Y  F  K  I  S  I  S  K  M  N  F  K  P  T  K  Y

     12700      12710      12720      12730      12740      12750      12760      12770      12780
AGTGGAAAATAAACATCAATAATTAGCAACTTCAACTTCCAACATAAACTCAATATTCCCTCATTTCCTGAATTCTCATCGTTTCGACGT
 S  G  K  *  T  S  I  I  S  N  F  N  F  Q  H  K  L  N  I  P  S  F  P  E  F  S  S  F  R  R
   V  E  N  K  H  Q  *  L  A  T  S  T  S  N  I  N  S  I  F  P  H  F  L  N  S  H  R  F  D  V
*  W  K  I  N  I  N  N  *  Q  L  Q  L  P  T  *  T  Q  Y  S  L  I  S  *  I  L  I  V  S  T

     12790      12800      12810      12820      12830      12840      12850      12860      12870
GTGCCTTAGCTACCGTAAAACAATATGGTGAAAATAATAACTAAACAAATTAAAATAAAACAAGCATTGTATTTCTGCCGGAAATGTAAT
 V  P  *  L  P  *  N  N  M  V  K  I  I  T  K  Q  I  K  I  K  Q  A  L  Y  F  C  R  K  C  N
   C  L  S  Y  R  K  T  I  W  *  K  *  *  L  N  K  L  K  *  N  K  H  C  I  S  A  G  N  V  I
C  A  L  A  T  V  K  Q  Y  G  E  N  N  N  *  T  N  *  N  K  T  S  I  V  F  L  P  E  M  *

     12880      12890      12900      12910      12920      12930      12940      12950      12960
ATACATTTTCATCTTAAAGGAGAAGGCAAATTTTTGCCAGAAGATATCCCAAAACACCTATGCACTCACATCCTTTATGCATTCACCAAC
 I  H  F  H  L  K  G  E  G  K  F  L  P  E  D  I  P  K  H  L  C  T  H  I  L  Y  A  F  T  N
   Y  I  F  I  L  K  E  K  A  N  F  C  Q  K  I  S  Q  N  T  Y  A  L  T  S  F  M  H  S  P  T
Y  T  F  S  S  *  R  R  R  Q  I  F  A  R  R  Y  P  K  T  P  M  H  S  H  P  L  C  I  H  Q

     12970      12980      12990      13000      13010      13020      13030      13040      13050
GTTGAAGGGAAACGCGCGTAAGTTACAAATGATATTTGCAATGAGAAATGTAAATTAAATTTTCGGAGTTTCTTTTACAATGCACATTAA
   V  E  G  K  R  A  *  V  T  N  D  I  C  N  E  K  C  K  L  N  F  R  S  F  F  Y  N  A  H  *
 L  K  G  N  A  R  K  L  Q  M  I  F  A  M  R  N  V  N  *  I  F  G  V  S  F  T  M  H  I  K
R  *  R  E  T  R  V  S  Y  K  *  Y  L  Q  *  E  M  *  I  K  F  S  E  F  L  L  Q  C  T  L

     13060      13070      13080      13090      13100      13110      13120      13130      13140
AATTATTCTAATTTTTATTGCCTGCTCAAGTGATTATAAACCTGAAACGTTAATATGCGATATTTAGCTGGTTTTGCATAAAATATTCTT
 N  Y  S  N  F  Y  C  L  L  K  *  L  *  T  *  N  V  N  M  R  Y  L  A  G  F  A  *  N  I  L
   I  I  L  I  F  I  A  C  S  S  D  Y  K  P  E  T  L  I  C  D  I  *  L  V  L  H  K  I  F  F
K  L  F  *  F  L  L  P  A  Q  V  I  I  N  L  K  R  *  Y  A  I  F  S  W  F  C  I  K  Y  S

     13150      13160      13170      13180      13190      13200      13210      13220      13230
CTGTAACAATTGCTATCAGCAAATGTTATTTATTTCTGTAAATTTCTGATTTTAATAATTTTTAAATCTTAGAACATTTGAATGGGACGA
 L  *  Q  L  L  S  A  N  V  I  Y  F  C  K  F  L  I  L  I  I  F  K  S  *  N  I  *  M  G  R
   C  N  N  C  Y  Q  Q  M  L  F  I  S  V  N  F  *  F  *  *  F  L  N  L  R  T  F  E  W  D  D
S  V  T  I  A  I  S  K  C  Y  L  F  L  *  I  S  D  F  N  N  F  *  I  L  E  H  L  N  G  T

     13240      13250      13260      13270      13280      13290      13300      13310      13320
TGAAGATACCGAATGGACGAAAGGGATATATCCACATATGATAAAAGTAAAAGAGTACGATCCAACCCTGAAAATTCTTCTTTCCTATAG
 *  R  Y  R  M  D  E  R  D  I  S  T  Y  D  K  S  K  R  V  R  S  N  P  E  N  S  S  F  L  *
   E  D  T  E  W  T  K  G  I  Y  P  H  M  I  K  V  K  E  Y  D  P  T  L  K  I  L  L  S  Y  S
M  K  I  P  N  G  R  K  G  Y  I  H  I  *  *  K  *  K  S  T  I  Q  P  *  K  F  F  F  P  I

     13330      13340      13350      13360      13370      13380      13390      13400      13410
CGACTATAACTCTCACTCATCAGTTTTTGCTGTTCGTTTTTTTTTTTAAACTCTCAGCTATCATTTTTTTTTTTAAAGATTGACACTCAGC
 R  L  *  L  S  L  I  S  F  C  C  S  F  F  F  *  T  L  S  Y  H  F  F  F  *  R  L  T  L  S
   D  Y  N  S  H  S  S  V  F  A  V  R  F  F  F  K  L  S  A  I  I  F  F  F  K  D  *  H  S  A
```

```
A  T  I  T  L  T  H  Q  F  L  L  F  V  F  F  L  N  S  Q  L  S  F  F  F  L  K  I  D  T  Q

     13420      13430      13440      13450      13460      13470      13480      13490      13500
TAACGTATTGACTAATAATTTCAAAAGAGACTGCTCATACGTCATTTCTTAGACATCAAGATAATGCAGTGTTACCTCAGACGTTTTCCG
 *  R  I  D  *  *  F  Q  K  R  L  L  I  R  H  F  L  D  I  K  I  M  Q  C  Y  L  R  R  F  P
  N  V  L  T  N  N  F  K  R  D  C  S  Y  V  I  S  *  T  S  R  *  C  S  V  T  S  D  V  F  R
L  T  Y  *  L  I  I  S  K  E  T  A  H  T  S  F  L  R  H  Q  D  N  A  V  L  P  Q  T  F  S

     13510      13520      13530      13540      13550      13560      13570      13580      13590
GATTTACGTCTCATCTTTTTCTTGCTAAAATATCAACGTAAATTTGATGATATAAAAATATTTGAAGGCAATAGCAAAATCGACAAAAAA
 D  L  R  L  I  F  F  L  L  K  Y  Q  R  K  F  D  D  I  K  I  F  E  G  N  S  K  I  D  K  K
  I  Y  V  S  S  F  S  C  *  N  I  N  V  N  L  M  I  *  K  Y  L  K  A  I  A  K  S  T  K  N
G  F  T  S  H  L  F  L  A  K  I  S  T  *  I  *  *  Y  K  N  I  *  R  Q  *  Q  N  R  Q  K

     13600      13610      13620      13630      13640      13650      13660      13670      13680
CCGGAAAAATTTCATTCAGTCAACTGTAGCATTTCTCCGTAAGCATAAATTTGATGGATTCGATCCGAATTCTTCGCCCTATAGTGAGTC
 P  E  K  F  H  S  V  N  C  S  I  S  P  *  A  *  I  *  W  I  R  S  E  F  F  A  L  *  *  V
  R  K  N  F  I  Q  S  T  V  A  F  L  R  K  H  K  F  D  G  F  D  P  N  S  S  P  Y  S  E  S
T  G  K  I  S  F  S  Q  L  *  H  F  S  V  S  I  N  L  M  D  S  I  R  I  L  R  P  I  V  S

     13690      13700      13710      13720      13730      13740      13750
GTATTACGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTTCGAGCTGGCG
 V  L  R  G  R  H  R  G  G  A  P  A  F  V  P  F  S  E  G  *  F  R  A  G  X
  Y  Y  A  A  A  T  A  V  E  L  Q  L  L  F  P  L  V  R  V  N  F  E  L  A
R  I  T  R  P  P  P  R  W  S  S  S  F  C  S  L  *  *  G  L  I  S  S  W  R
```

Appendix 3: Partial insert of clone 12. Sequence Range: 1 to 6506

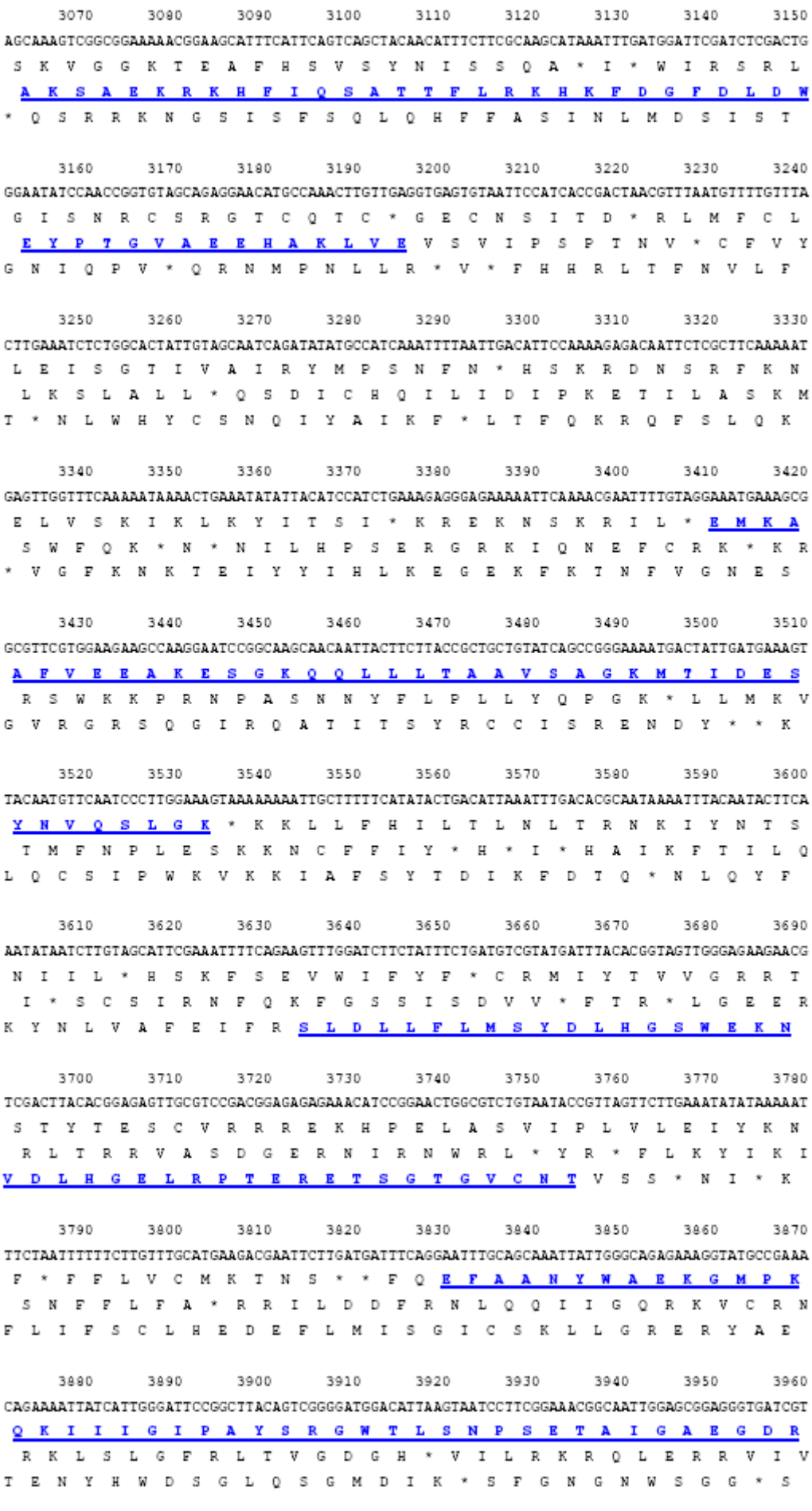
Gene I: 1011...5984

```
          10         20         30         40         50         60         70         80         90
ACGTTTTAACTGGGTAACATGTACAGCATACGGAATTTTCATATTCTCATAAGAAAAATTTTGTTCATTCCTGTTACGGTAAAAGACATA
 T  F  *  L  G  N  M  Y  S  I  R  N  F  H  I  L  I  R  K  I  L  F  I  P  V  T  V  K  D  I
  R  F  N  W  V  T  C  T  A  Y  G  I  F  I  F  S  *  E  K  F  C  S  F  L  L  R  *  K  T  *
   V  L  T  G  *  H  V  Q  H  T  E  F  S  Y  S  H  K  K  N  F  V  H  S  C  Y  G  K  R  H

         100        110        120        130        140        150        160        170        180
ACAAAAACTCCAGACAATAGAATTCTTGGCAAATATGGTAGCAGTGTTATTAAGGATTTTTTAGCTTTCTGCATATTTTTAATGTTTTTC
 T  K  T  P  D  N  R  I  L  G  K  Y  G  S  S  V  I  K  D  F  L  A  F  C  I  F  L  M  F  F
  Q  K  L  Q  T  I  E  F  L  A  N  M  V  A  V  L  L  R  I  F  *  L  S  A  Y  F  *  C  F  S
N  K  N  S  R  Q  *  N  S  W  Q  I  W  *  Q  C  Y  *  G  F  F  S  F  L  H  I  F  N  V  F

         190        200        210        220        230        240        250        260        270
CAGCTTTTTAGCTATTTCCTTGTTTTTCCAGCCTTTTAGCCATTTCTTGTGTTTTTCCAGCTTTCTATCTGTTTTTCGTTTTTTTTTTAA
 Q  L  F  S  Y  F  L  V  F  P  A  F  *  P  F  L  V  F  F  Q  L  S  I  C  F  S  F  F  F  *
  S  F  L  A  I  S  L  F  F  Q  P  F  S  H  F  L  C  F  S  S  F  L  S  V  F  R  F  F  F  N
P  A  F  *  L  F  P  C  F  S  S  L  L  A  I  S  C  V  F  P  A  F  Y  L  F  F  V  F  F  L

         290        290        300        310        320        330        340        350        360
CTCCTTAAAATTATTCATGTCATCCCAACATTGAAATTTCGACTAATTAATTCAGTTAGCCCAGTATCCGTCAATCAAATTACGCGTTTT
 L  L  K  I  I  H  V  I  P  T  L  K  F  R  L  I  N  S  V  S  P  V  S  V  N  Q  I  T  R  F
  S  L  K  L  F  M  S  S  Q  H  *  N  F  D  *  L  I  Q  L  A  Q  Y  P  S  I  K  L  R  V  F
T  P  *  N  Y  S  C  H  P  N  I  E  I  S  T  N  *  F  S  *  P  S  I  R  Q  S  N  Y  A  F
```

```
      370       380       390       400       410       420       430       440       450
CAGCTTTTTGGCTTCACTCTTTGCAAAGTGTGAAAATATATGATTTACAAAGTGTGCAACATTATTAATATTAGATTGTTGATATTGTTT
 Q  L  F  G  F  T  L  C  K  V  *  K  Y  M  I  Y  K  V  C  N  I  I  N  I  R  L  L  I  L  F
  S  F  L  A  S  L  F  A  K  C  E  N  I  *  F  T  K  C  A  T  L  L  I  L  D  C  *  Y  C  F
S  A  F  W  L  H  S  L  Q  S  V  K  I  Y  D  L  Q  S  V  Q  H  Y  *  Y  *  I  V  D  I  V

      460       470       480       490       500       510       520       530       540
TCATTTTATTTCCAACTTTCTTGTTTTATTTGAAAGCTGTGCAAACATTCCATATCTATAAAAATCTGTATAAAATTCAATCCTGATTCG
 S  F  Y  F  Q  L  S  C  F  I  *  K  L  C  K  H  S  I  S  I  K  I  C  I  K  F  N  P  D  S
  H  F  I  S  N  F  L  V  L  F  E  S  C  A  N  I  P  Y  L  *  K  S  V  *  N  S  I  L  I  R
F  I  L  F  P  T  F  L  F  Y  L  K  A  V  Q  T  F  H  I  Y  K  N  L  Y  K  I  Q  S  *  F

      550       560       570       580       590       600       610       620       630
TCTATTTTGGCTTCAGCGTGGTTTACTTAATTCCATATCTATTCCAATTTTAGCATATAAACAAAAATAAAATTAAATATAGGATTAAAA
 S  I  L  A  S  A  W  F  T  *  F  H  I  Y  S  N  F  S  I  *  T  K  I  K  L  N  I  G  L  K
  L  F  W  L  Q  R  G  L  L  N  S  I  S  I  P  I  L  A  Y  K  Q  K  *  N  *  I  *  D  *  N
V  Y  F  G  F  S  V  V  Y  L  I  P  Y  L  F  Q  F  *  H  I  N  K  N  K  I  K  Y  R  I  K

      640       650       660       670       680       690       700       710       720
CATGAAAAACTTTCCATTTTGTTATTAAAACGACATTTTAAAGTTCTCATCTTCAAGTTTCATCTAGTTTTTCTAATGAATGTTTTTCTT
 H  E  K  L  S  I  L  L  L  K  R  H  F  K  V  L  I  F  K  F  H  L  V  F  L  M  N  V  F  L
  M  K  N  F  P  F  C  Y  *  N  D  I  L  K  F  S  S  S  S  F  I  *  F  F  *  *  M  F  F  F
T  *  K  T  F  H  F  V  I  K  T  T  F  *  S  S  H  L  Q  V  S  S  S  F  S  N  E  C  F  S

      730       740       750       760       770       780       790       800       810
CAACCTTTGATATGTAGCAACAATTATTCTTACTAACTCCTTTAATACAATGCCGCTATTGTTCCCCATTTAATTGTTAAAATTCATAAT
 Q  P  L  I  C  S  N  N  Y  S  Y  *  L  L  *  Y  N  A  A  I  V  P  H  L  I  V  K  I  H  N
  N  L  *  Y  V  A  T  I  I  L  T  N  S  F  N  T  M  P  L  L  F  P  I  *  L  L  K  F  I  I
S  T  F  D  M  *  Q  Q  L  F  L  L  T  P  L  I  Q  C  R  Y  C  S  P  F  N  C  *  N  S  *

      820       830       840       850       860       870       880       890       900
CATCAAATTACATATAATTATGATAGTAATTTCATAATTTTCCATTAACTGATTGTCTTATTTTATTCCTAACGCATATGCATATGACGC
 H  Q  I  T  Y  N  Y  D  S  N  F  I  I  F  H  *  L  I  V  L  F  Y  S  *  R  I  C  I  *  R
  I  K  L  H  I  I  M  I  V  I  S  *  F  S  I  N  *  L  S  Y  F  I  P  N  A  Y  A  Y  D  A
S  S  N  Y  I  *  L  *  *  *  F  H  N  F  P  L  T  D  C  L  I  L  F  L  T  H  M  H  M  T

      910       920       930       940       950       960       970       980       990
CGTATCTTCTAATAATGAAATTATTACTGAAAACAAATATCAGTTAGTCACAGTAATAAATGTTGAAATGCAGAAGTGTACAAGTCTTAC
 R  I  F  *  *  *  N  Y  Y  *  K  Q  I  S  V  S  H  S  N  K  C  *  N  A  E  V  Y  K  S  Y
  V  S  S  N  N  E  I  I  T  E  N  K  Y  Q  L  V  T  V  I  N  V  E  M  Q  K  C  T  S  L  T
P  Y  L  L  I  M  K  L  L  L  K  T  N  I  S  *  S  Q  *  *  M  L  K  C  R  S  V  Q  V  L

     1000      1010      1020      1030      1040      1050      1060      1070      1080
AAGGTGACATGAACTGGATTATGATAACGCTTTTCATTATACTTGCCAACGCAATTACTGTTGTAAACGGTAAGTTGCTCAATAAAATGA
 K  V  T  *  T  G  L  *  *  R  F  S  L  Y  L  P  T  Q  L  L  L  *  T  V  S  C  S  I  K  *
  R  *  H  E  L  D  Y  D  N  A  F  H  Y  T  C  Q  R  N  Y  C  C  K  R  *  V  A  Q  *  N  D
Q  G  D  M  N  W  I  M  I  T  L  F  I  I  L  A  N  A  I  T  V  V  N  G  K  L  L  N  K  M

     1090      1100      1110      1120      1130      1140      1150      1160      1170
CACGTCAGTTGCGTTAACACGAAATATCATCGCTGAAGATATAATAAGCATAACAACAGATCATACATCCTAATCATTACATGCCATTGC
 H  V  S  C  V  N  T  K  Y  H  R  *  R  Y  N  K  H  N  N  R  S  Y  I  L  I  I  T  C  H  C
  T  S  V  A  L  T  R  N  I  I  A  E  D  I  I  S  I  T  T  D  H  T  S  *  S  L  H  A  I  A
T  R  Q  L  R  *  H  E  I  S  S  L  K  I  *  *  A  *  Q  Q  I  I  H  P  N  H  Y  M  P  L

     1180      1190      1200      1210      1220      1230      1240      1250      1260
AATCCTAAATTGTCGAGATAATGATATACATTGAATTTCACAGAAGCTTTCACAAAATGAAGAGGCTAAGTTGAAGATGAAACAAATTTT
 N  P  K  L  S  R  *  *  Y  T  L  N  F  T  E  A  F  T  K  *  R  G  *  V  E  D  E  T  N  F
  I  L  N  C  R  D  N  D  I  H  *  I  S  Q  K  L  S  Q  N  E  E  A  K  L  K  M  K  Q  I  L
Q  S  *  I  V  E  I  M  I  Y  I  E  F  H  R  S  F  H  K  M  K  R  L  S  *  R  *  N  K  F
```

```
      1270       1280       1290       1300       1310       1320       1330       1340       1350
ATTTTAAAAAAATTGGTAATACTGTTGACTAAATTTTAAAGAAATAACGACGAAATCGATTTTGAAATATCTTCAAGCATTAGCCTACTTC
 I  L  K  N  W  *  Y  C  *  L  N  F  K  E  I  T  T  K  S  I  L  K  Y  L  Q  A  L  A  Y  F
  F  *  K  I  G  N  T  V  D  *  I  L  K  K  *  R  R  N  R  F  *  N  I  F  K  H  *  P  T  S
Y  F  K  K  L  V  I  L  L  T  K  F  *  R  N  N  D  E  I  D  F  E  I  S  S  S  I  S  L  L

      1360       1370       1380       1390       1400       1410       1420       1430       1440
AGGTTACGTTCGCGGATGTTACTACACAAATTGGGCTCAATATCGACAAGGTTCACTGATGTTTGAAATAATTTTGTTATTAAATAGTCA
 R  L  R  S  R  M  L  L  H  K  L  G  S  I  S  T  R  F  T  D  V  *  N  N  F  V  I  K  *  S
  G  Y  V  R  G  C  Y  Y  T  N  W  A  Q  Y  R  Q  G  S  L  M  F  E  I  I  L  L  L  N  S  Q
Q  V  T  F  A  D  V  T  T  Q  I  G  L  N  I  D  K  V  H  *  C  L  K  *  F  C  Y  *  I  V

      1450       1460       1470       1480       1490       1500       1510       1520       1530
ATAATTTAAAAGGATTGTACAGTTTTTCAAATAAAAATTAGAAAACTGATGGAATCATTTCAAGATTTCGATCTTAAAAGTGAATTTTAA
 I  I  *  K  D  C  T  V  F  Q  I  K  I  R  K  L  M  E  S  F  Q  D  F  D  L  K  S  E  F  *
  *  F  K  R  I  V  Q  F  F  K  *  K  L  E  N  *  W  N  H  F  K  I  S  I  L  K  V  N  F  K
N  N  L  K  G  L  Y  S  F  S  N  K  N  *  K  T  D  G  I  I  S  R  F  R  S  *  K  *  I  L

      1540       1550       1560       1570       1580       1590       1600       1610       1620
ACTATACGAAATATTAGTGGAAAATAAACATCGATTAATGGTTCAAGTTTCAACATAAACTCAAATTTTATTTATATTGATATTTCCTGA
 T  I  R  N  I  S  G  K  *  T  S  I  N  G  S  S  F  N  I  N  S  N  F  I  Y  I  D  I  S  *
  L  Y  E  I  L  V  E  N  K  H  R  L  M  V  Q  V  S  T  *  T  Q  I  L  F  I  L  I  F  P  E
N  Y  T  K  Y  *  W  K  I  N  I  D  *  W  F  K  F  Q  H  K  L  K  F  Y  L  Y  *  Y  F  L

      1630       1640       1650       1660       1670       1680       1690       1700       1710
GTGTACATCGTTTCAACGTGTGCTTTTTTTTTCTATCATAAACTAATGTGCTGAAAAAAATAAATGAGCTAATTAAAGTAAAACAAATAT
 V  Y  I  V  S  T  C  A  F  F  F  Y  H  K  L  M  C  *  K  K  *  M  S  *  L  K  *  N  K  Y
  C  T  S  F  Q  R  V  L  F  F  S  I  I  N  *  C  A  E  K  N  K  *  A  N  *  S  K  T  N  I
S  V  H  R  F  N  V  C  F  F  F  L  S  *  T  N  V  L  K  K  I  N  E  L  I  K  V  K  Q  I

      1720       1730       1740       1750       1760       1770       1780       1790       1800
CGTATTACTGTCGGAAATGTAATATACATTTTCATCTTAAAGGAGAAGGCAAATTCTTGCCGGAAGATATCCCAATAGGCCTGTGCACTC
 R  I  T  V  G  N  V  I  Y  I  F  I  L  K  E  K  A  N  S  C  R  K  I  S  Q  *  A  C  A  L
  V  L  L  S  E  M  *  Y  T  F  S  S  *  R  R  R  Q  I  L  A  G  R  Y  P  N  R  P  V  H  S
S  Y  Y  C  R  K  C  N  I  H  F  H  L  K  G  E  G  K  F  L  P  E  D  I  P  I  G  L  C  T

      1810       1820       1830       1840       1850       1860       1870       1880       1890
ACATTCTTTATGCATTCGCCAAAGTTGACGAGAAAGGCACATCAATGTAAGTTACAAATAATACTTGTAATGAGAAACATAAACTAAATT
 T  F  F  M  H  S  P  K  L  T  R  K  A  H  Q  C  K  L  Q  I  I  L  V  M  R  N  I  N  *  I
  H  S  L  C  I  R  Q  S  *  R  E  R  H  I  N  V  S  Y  K  *  Y  L  *  *  E  T  *  T  K  F
H  I  L  Y  A  F  A  K  V  D  E  K  G  T  S  M  *  V  T  N  N  T  C  N  E  K  H  K  L  N

      1900       1910       1920       1930       1940       1950       1960       1970       1980
TTCGCATTTTGCTCGTACAGTGGACATTCAGACTATTTAAAATTTTTGTCAGCTATTCAAGTTGTCACATACTGGCAGCATTCATGTGGA
 F  A  F  C  S  Y  S  G  H  S  D  Y  L  K  F  L  S  A  I  Q  V  V  T  Y  W  Q  H  S  C  G
  S  H  F  A  R  T  V  D  I  Q  T  I  *  N  F  C  Q  L  F  K  L  S  H  T  G  S  I  H  V  E
F  R  I  L  L  V  Q  W  T  F  R  L  F  K  I  F  V  S  Y  S  S  C  H  I  L  A  A  F  M  W

      1990       2000       2010       2020       2030       2040       2050       2060       2070
ACATTCAGTAAAGATAGGCGGTTTTGCATAATATGCTCTTCTATGATGACCATTATCAGCAAATATAATCTATCCTGTAAATTTCCTATT
 T  F  S  K  D  R  R  F  C  I  I  C  S  S  M  M  T  I  I  S  K  Y  N  L  S  C  K  F  P  I
  H  S  V  K  I  G  G  F  A  *  Y  A  L  L  *  *  P  L  S  A  N  I  I  Y  P  V  N  F  L  F
N  I  Q  *  R  *  A  V  L  H  N  M  L  F  Y  D  D  H  Y  Q  Q  I  *  S  I  L  *  I  S  Y

      2080       2090       2100       2110       2120       2130       2140       2150       2160
TCAATAACTTTTAATCTTAGGGCATTTGAATGGAATGACGAAGATACCGAATGGTCAAAAGGGATGTACTCACGTGTAATAAAACTGAGA
 S  I  T  F  N  L  R  A  F  E  W  N  D  E  D  T  E  W  S  K  G  M  Y  S  R  V  I  K  L  R
  Q  *  L  L  I  L  G  H  L  N  G  M  T  K  I  P  N  G  Q  K  G  C  T  H  V  *  *  N  *  E
F  N  N  F  *  S  *  G  I  *  M  E  *  R  R  Y  R  M  V  K  R  D  V  L  T  C  N  K  T  E
```

```
      2170      2180      2190      2200      2210      2220      2230      2240      2250
GAGAACGATCCAACACTGAAAATCCTTCTTTCTTATGGTGGCTATAACTTCGGATCGTCAACTTTTACTGTTTGTTTCATTTTTTTAAAG
  E  N  D  P  T  L  K  I  L  L  S  Y  G  G  Y  N  F  G  S  S  T  F  T  V  C  F  I  F  L  K
   R  T  I  Q  H  *  K  S  F  F  L  M  V  A  I  T  S  D  R  Q  L  L  F  V  S  F  F  *  R
R  E  R  S  N  T  E  N  P  S  F  L  W  W  L  *  L  R  I  V  N  F  Y  C  L  F  H  F  F  K

      2260      2270      2280      2290      2300      2310      2320      2330      2340
ATTTACGCTCGCTCAACTCATGAATTTTCTTTTGATTCGGGAAAGCAGATTGGTATTCCTCGATAATTATATTGAACAGGTTATTCGGAA
   I  Y  A  R  S  T  H  E  F  S  F  D  S  G  K  Q  I  G  I  P  R  *  L  Y  *  T  G  Y  S  E
    F  T  L  A  Q  L  M  N  F  L  L  I  R  E  S  R  L  V  F  L  D  N  Y  I  E  Q  V  I  R  N
D  L  R  S  L  N  S  *  I  F  F  *  F  G  K  A  D  W  Y  S  S  I  I  I  L  N  R  L  F  G

      2350      2360      2370      2380      2390      2400      2410      2420      2430
TAGAGTGGAAATATTACGGGAGAACAGTAATGAATAGTACTATTTCTGACGATTGATTTGTGAAACTTTGACGAGTCACTTTTCCCTTTT
   *  S  G  N  I  T  G  E  Q  *  *  I  V  L  F  L  T  I  D  L  *  N  F  D  E  S  L  F  P  F
    R  V  E  I  L  R  E  N  S  N  E  *  Y  Y  F  *  R  L  I  C  E  T  L  T  S  H  F  S  L  L
I  E  W  K  Y  Y  G  R  T  V  M  N  S  T  I  S  D  D  *  F  V  K  L  *  R  V  T  F  P  F

      2440      2450      2460      2470      2480      2490      2500      2510      2520
GATCTGTATATATATATATAGAGTGTCCAACGTAACTATACATATACCAAATATTGGGCGATAACAGGATAGTTATTTAGATACATAAAA
   D  L  Y  I  Y  I  *  S  V  Q  R  N  Y  T  Y  T  K  Y  W  A  I  T  G  *  L  F  R  Y  I  K
    I  C  I  Y  I  Y  R  V  S  N  V  T  I  H  I  P  N  I  G  R  *  Q  D  S  Y  L  D  T  *  N
*  S  V  Y  I  Y  I  E  C  P  T  *  L  Y  I  Y  Q  I  L  G  D  N  R  I  V  I  *  I  H  K

      2530      2540      2550      2560      2570      2580      2590      2600      2610
TCAACATGAAACACGTATATCAGTCATATAATAAATCAATAGAAATCACGTAGAGATACGCGTGATAAGTTACAAGAATCACACAGATTT
   S  T  *  N  T  Y  I  S  H  I  I  N  Q  *  K  S  R  R  D  T  R  D  K  L  Q  E  S  H  R  F
    Q  H  E  T  R  I  S  V  I  *  *  I  N  R  N  H  V  E  I  R  V  I  S  Y  K  N  H  T  D  L
I  N  M  K  H  V  Y  Q  S  Y  N  K  S  I  E  I  T  *  R  Y  A  *  *  V  T  R  I  T  Q  I

      2620      2630      2640      2650      2660      2670      2680      2690      2700
GCGCAGGATAAAATGGGTAGCCGACATGCAGACATATTACAGATAATAGAATAGTTAGGCAATGTAAATGAAAAGAAAGAAAGAGCGATG
   A  Q  D  K  M  G  S  R  H  A  D  I  L  Q  I  I  E  *  L  G  N  V  N  E  K  K  E  R  A  M
    R  R  I  K  W  V  A  D  M  Q  T  Y  Y  R  *  *  N  S  *  A  M  *  M  K  R  K  K  E  R  W
C  A  G  *  N  G  *  P  T  C  R  H  I  T  D  N  R  I  V  R  Q  C  K  *  K  E  R  K  S  D

      2710      2720      2730      2740      2750      2760      2770      2780      2790
GGGAAGAAGGAGACTGTTTGATTACTCATTGGCTAGATTTGTTGTTATAGATAAAAAGAAAGATAGAGAACATGTAATAGCGGGAAATGT
   G  K  K  E  T  V  *  L  L  I  G  *  I  C  C  Y  R  *  K  E  R  *  R  T  C  N  S  G  K  C
    G  R  R  R  L  F  D  Y  S  L  A  R  F  V  V  I  D  K  K  K  D  R  E  H  V  I  A  G  N  V
G  E  E  G  D  C  L  I  T  H  W  L  D  L  L  L  *  I  K  R  K  I  E  N  M  *  *  R  E  M

      2800      2810      2820      2830      2840      2850      2860      2870      2880
TGGCGGGTAATTTACTGTGAATGACATTCAAATAGCACAAATATAAAATAACGGATGGCAGGAAATTAAATTTCACATTGAAGAAAGATT
   W  R  V  I  Y  C  E  *  H  S  N  S  T  N  I  K  *  R  M  A  G  N  *  I  S  H  *  R  K  I
    G  G  *  F  T  V  N  D  I  Q  I  A  Q  I  *  N  N  G  W  Q  E  I  K  F  H  I  E  E  R  L
L  A  G  N  L  L  *  M  T  F  K  *  H  K  Y  K  I  T  D  G  R  K  L  N  F  T  L  K  K  D

      2890      2900      2910      2920      2930      2940      2950      2960      2970
GTGAATAGCAACTTCTTTAGCTTGGCTGTCAATTATACAAGTCATAAAACTTCGGGTTTCCCGACACTAACACATTGACAAATGATTTCA
   V  N  S  N  F  F  S  L  A  V  N  Y  T  S  H  K  T  S  G  F  P  T  L  T  H  *  Q  M  I  S
    *  I  A  T  S  L  A  W  L  S  I  I  Q  V  I  K  L  R  V  S  R  H  *  H  I  D  K  *  F  Q
C  E  *  Q  L  L  *  L  G  C  Q  L  Y  K  S  *  N  F  G  F  P  D  T  N  T  L  T  N  D  F

      2980      2990      3000      3010      3020      3030      3040      3050      3060
AAAAGAGCTGCAAACACGTGATGCATTGTATCTCAAATGTTCTCCAACTTTACGTCTCACTTTTTTCTTGCTGAAATATTTGAAGGCAAT
   K  R  A  A  N  T  *  C  I  V  S  Q  M  F  S  N  F  T  S  H  F  F  L  A  E  I  F  E  G  N
    K  E  L  Q  T  R  D  A  L  Y  L  K  C  S  P  T  L  R  L  T  F  F  L  L  K  Y  L  K  A  I
K  K  S  C  K  H  V  M  H  C  I  S  N  V  L  Q  L  Y  V  S  L  F  S  C  *  N  I  *  R  Q
```


3070 3080 3090 3100 3110 3120 3130 3140 3150
AGCAAAGTCGGCGGAAAAACGGAAGCATTTCATTCAGTCAGCTACAACATTTCTTCGCAAGCATAAATTTGATGGATTCGATCTCGACTG
S K V G G K T E A F H S V S Y N I S S Q A * I * W I R S R L
A K S A E K R K H F I Q S A T T F L R K H K F D G F D L D W
* Q S R R K N G S I S F S Q L Q H F F A S I N L M D S I S T
3160 3170 3180 3190 3200 3210 3220 3230 3240
GGAATATCCAACCGGTGTAGCAGAGGAACATGCCAAACTTGTTGAGGTGAGTGTAATTCCATCACCGACTAACGTTTAATGTTTTGTTTA
G I S N R C S R G T C Q T C * G E C N S I T D * R L M F C L
E Y P T G V A E E H A K L V E V S V I P S P T N V * C F V Y
G N I Q P V * Q R N M P N L L R * V * F H H R L T F N V L F
3250 3260 3270 3280 3290 3300 3310 3320 3330
CTTGAAATCTCTGGCACTATTGTAGCAATCAGATATATGCCATCAAATTTTAATTGACATTCCAAAAGAGACAATTCTCGCTTCAAAAAT
L E I S G T I V A I R Y M P S N F N * H S K R D N S R F K N
L K S L A L L * Q S D I C H Q I L I D I P K E T I L A S K M
T * N L W H Y C S N Q I Y A I K F * L T F Q K R Q F S L Q K
3340 3350 3360 3370 3380 3390 3400 3410 3420
GAGTTGGTTTCAAAAATAAAACTGAAATATATTACATCCATCTGAAAGAGGGAGAAAAATTCAAAACGAATTTTGTAGGAAATGAAAGCG
E L V S K I K L K Y I T S I * K R E K N S K R I L * E M K A
S W F Q K * N * N I L H P S E R G R K I Q N E F C R K * K R
* V G F K N K T E I Y Y I H L K E G E K F K T N F V G N E S
3430 3440 3450 3460 3470 3480 3490 3500 3510
GCGTTCGTGGAAGAAGCCAAGGAATCCGGCAAGCAACAATTACTTCTTACCGCTGCTGTATCAGCCGGGAAAATGACTATTGATGAAAGT
A F V E E A K E S G K Q Q L L L T A A V S A G K M T I D E S
R S W K K P R N P A S N N Y F L P L L Y Q P G K * L L M K V
G V R G R S Q G I R Q A T I T S Y R C C I S R E N D Y * * K
3520 3530 3540 3550 3560 3570 3580 3590 3600
TACAATGTTCAATCCCTTGGAAAGTAAAAAAAAATTGCTTTTTCATATACTGACATTAAATTTGACACGCAATAAAATTTACAATACTTCA
Y N V Q S L G K * K K L L F H I L T L N L T R N K I Y N T S
T M F N P L E S K K N C F F I Y * H * I * H A I K F T I L Q
L Q C S I P W K V K K I A F S Y T D I K F D T Q * N L Q Y F
3610 3620 3630 3640 3650 3660 3670 3680 3690
AATATAATCTTGTAGCATTCGAAATTTTCAGAAGTTTGGATCTTCTATTTCTGATGTCGTATGATTTACACGGTAGTTGGGAGAAGAACG
N I I L * H S K F S E V W I F Y F * C R M I Y T V V G R R T
I * S C S I R N F Q K F G S S I S D V V * F T R * L G E E R
K Y N L V A F E I F R S L D L L F L M S Y D L H G S W E K N
3700 3710 3720 3730 3740 3750 3760 3770 3780
TCGACTTACACGGAGAGTTGCGTCCGACGGAGAGAGAAACATCCGGAACTGGCGTCTGTAATACCGTTAGTTCTTGAAATATATAAAAAT
S T Y T E S C V R R R E K H P E L A S V I P L V L E I Y K N
R L T R R V A S D G E R N I R N W R L * Y R * F L K Y I K I
V D L H G E L R P T E R E T S G T G V C N T V S S * N I * K
3790 3800 3810 3820 3830 3840 3850 3860 3870
TTCTAATTTTTTCTTGTTTGCATGAAGACGAATTCTTGATGATTTCAGGAATTTGCAGCAAATTATTGGGCAGAGAAAGGTATGCCGAAA
F * F F L V C M K T N S * * F Q E F A A N Y W A E K G M P K
S N F F L F A * R R I L D D F R N L Q Q I I G Q R K V C R N
F L I F S C L H E D E F L M I S G I C S K L L G R E R Y A E
3880 3890 3900 3910 3920 3930 3940 3950 3960
CAGAAAATTATCATTGGGATTCCGGCTTACAGTCGGGGATGGACATTAAGTAATCCTTCGGAAACGGCAATTGGAGCGGAGGGTGATCGT
Q K I I I G I P A Y S R G W T L S N P S E T A I G A E G D R
R K L S L G F R L T V G D G H * V I L R K R Q L E R R V I V
T E N Y H W D S G L Q S G M D I K * S F G N G N W S G G * S

```
      3970       3980       3990       4000       4010       4020       4030       4040       4050
CCATCATCGCCATCGACAACAAATCCCGCNGGCGGTACTGCGGCTTATTGGGAGGTTTAATTTAAAGTTCTCATTTACTTTTAAATACAT
 P  S  S  P  S  T  T  N  P  A  G  G  T  A  A  Y  W  E  V  *  F  K  V  L  I  Y  F  *  I  H
  H  H  R  H  R  Q  Q  I  P  X  A  V  L  R  L  I  G  R  F  N  L  K  F  S  F  T  F  K  Y  I
S  I  I  A  I  D  N  K  S  R  X  R  Y  C  G  L  L  G  G  L  I  *  S  S  H  L  L  L  N  T

      4060       4070       4080       4090       4100       4110       4120       4130       4140
TTCTTGCTGCTGATGAATTTCTCAAATTCCATTATTAATGAATGTAACCATTTTAAGATATGCAAATATGTGAAAGAAGGTGGCAAGGAA
 F  L  L  L  M  N  F  S  N  S  I  I  N  E  C  N  H  F  K  I  C  K  Y  V  K  E  G  G  K  E
  S  C  C  *  *  I  S  Q  I  P  L  L  M  N  V  T  I  L  R  Y  A  N  M  *  K  K  V  A  R  K
F  L  A  A  D  E  F  L  K  F  H  Y  *  *  M  *  P  F  *  D  M  Q  I  C  E  R  R  W  Q  G

      4150       4160       4170       4180       4190       4200       4210       4220       4230
ACGGTGGATAAAAAAAAGTGTTGGAGCGTATATGGTAAAAGGAGATCAGTGGCACGGTTACGATAATGAGGAAACTATTAAAATTAAAGTA
 T  V  D  K  K  S  V  G  A  Y  M  V  K  G  D  Q  W  H  G  Y  D  N  E  E  T  I  K  I  K  V
  R  W  I  K  K  V  L  E  R  I  W  *  K  E  I  S  G  T  V  T  I  M  R  K  L  L  K  L  K  *
N  G  G  *  K  K  C  W  S  V  Y  G  K  R  R  S  V  A  R  L  R  *  *  G  N  Y  *  N  *  S

      4240       4250       4260       4270       4280       4290       4300       4310       4320
ATTTTTATTTCGCTTATTGTTGTTGATAGTCAAAAATAAAAAAGTGATTTGATAGTCAAAAAATAAGAAAAGATAATAAGATAGTCGAAA
 I  F  I  S  L  I  V  V  D  S  Q  K  *  K  S  D  L  I  V  K  K  *  E  K  I  I  R  *  S  K
  F  L  F  R  L  L  L  L  I  V  K  N  K  K  V  I  *  *  S  K  N  K  K  R  *  *  D  S  R  K
N  F  Y  F  A  Y  C  C  *  *  S  K  I  K  K  *  F  D  S  Q  K  I  R  K  D  N  K  I  V  E

      4330       4340       4350       4360       4370       4380       4390       4400       4410
AATAAAAAAATTAAAAAATAACGTTGCAAATATATCTTGCCTTAACTTTAATTTAAAGAGCACCATAAATTTTATAATGTTTAATAAGAAG
 N  K  K  L  K  N  N  V  A  N  I  S  C  L  N  F  N  L  K  S  T  I  N  F  I  M  F  N  K  K
  I  K  N  *  K  I  T  L  Q  I  Y  L  A  L  T  L  I  *  R  A  P  *  I  L  *  C  L  I  R  R
K  *  K  I  K  K  *  R  C  K  Y  I  L  P  *  L  *  F  K  E  H  H  K  F  Y  N  V  *  *  E

      4420       4430       4440       4450       4460       4470       4480       4490       4500
AAATTAGTTTCGGAGTTTACTCGTGTAATTGTTTTACTTTAGAATTAAAATTTTCAGATGAAATGGCTGAAAGAGGAGGATTANGGCGGT
 K  L  V  S  E  F  T  R  V  I  V  L  L  *  N  *  N  F  Q  M  K  W  L  K  E  E  D  X  G  G
  N  *  F  R  S  L  L  V  *  L  F  Y  F  R  I  K  I  F  R  *  N  G  *  K  R  R  I  X  A  V
E  I  S  F  G  V  Y  S  C  N  C  F  T  L  E  L  K  F  S  D  E  M  A  E  R  G  G  L  X  R

      4510       4520       4530       4540       4550       4560       4570       4580       4590
GCCTTTATGTGGACTCTTGATTTTGATGATTTCAAAGGTACAAGTTGTGGCAAAGGTCCATATCCACTGTTAAATGCTATCAATAATGGG
 A  F  M  W  T  L  D  F  D  D  F  K  G  T  S  C  G  K  G  P  Y  P  L  L  N  A  I  N  N  G
  P  L  C  G  L  L  I  L  M  I  S  K  V  Q  V  V  A  K  V  H  I  H  C  *  M  L  S  I  M  G
C  L  Y  V  D  S  *  F  *  *  F  Q  R  Y  K  L  W  Q  R  S  I  S  T  V  K  C  Y  Q  *  W

      4600       4610       4620       4630       4640       4650       4660       4670       4680
CTTGAAAGTGAGGTAAGACATGTTTCAAAGAAATATTTTTAACAATTCTGTTTTTCTTTCAATACGTTTGGATTTAAATTATGCTGATGT
 L  E  S  E  V  R  H  V  S  K  K  Y  F  *  Q  F  C  F  S  F  N  T  F  G  F  K  L  C  *  C
  L  K  V  R  *  D  M  F  Q  R  N  I  F  N  N  S  V  F  L  S  I  R  L  D  L  N  Y  A  D  V
A  *  K  *  G  K  T  C  F  K  E  I  F  L  T  I  L  F  F  F  Q  Y  V  W  I  *  I  M  L  M

      4690       4700       4710       4720       4730       4740       4750       4760       4770
TTGTAAATAAATTAATAAAAAAAAAGGCGAATGGATGAATCAGAATAGACTTTCGGTGGGACCATACGGAAAAATGACTCTATGAATTAT
 L  *  I  N  *  *  K  K  R  R  M  D  E  S  E  *  T  F  G  G  T  I  R  K  N  D  S  M  N  Y
  C  K  *  I  N  K  K  K  G  E  W  M  N  Q  N  R  L  S  V  G  P  Y  G  K  M  T  L  *  I  I
F  V  N  K  L  I  K  K  K  A  N  G  *  I  R  I  D  F  R  W  D  H  T  E  K  *  L  Y  E  L

      4780       4790       4800       4810       4820       4830       4840       4850       4860
AGGATAAAATAGAATTTCTGTCCGAAAGTGATGACTCACTAACTTTTGATTAGCTGGATGGGGCATCAATTTCTATATTGACGTTGTTCAA
 R  I  N  R  I  S  V  R  K  *  *  L  T  N  F  *  L  A  G  W  G  I  N  F  Y  I  D  V  V  Q
  G  *  I  E  F  L  S  E  S  D  D  S  L  T  F  D  *  L  D  G  A  S  I  S  I  L  T  L  F  K
*  D  K  *  N  F  C  P  K  V  M  T  H  *  L  L  I  S  W  M  G  H  Q  F  L  Y  *  R  C  S
```

```
      4870       4880       4890       4900       4910       4920       4930       4940       4950
ACAAAATCTTAAATTTTTAATTATCTCTTGAAAATTAATTTTTAAAATTTAACACATTTTTAGCAAACACCATCAACAAGAGCAGTACCG
 T  K  S  *  I  F  N  Y  L  L  K  I  N  F  *  N  L  T  H  F  *  Q  T  P  S  T  R  A  V  P
  Q  N  L  K  F  L  I  I  S  *  K  L  I  F  K  I  *  H  I  F  S  K  H  H  Q  Q  E  Q  Y  R
N  K  I  L  N  F  *  L  S  L  E  N  *  F  L  K  F  N  T  F  L  A  N  T  I  N  K  S  S  T

      4960       4970       4980       4990       5000       5010       5020       5030       5040
GAAACTACAGAAGACACTGAAGTTGAAACAATGACGGAAGCACCGGGAACTGCAAATGATACTGAAATGGAAACAGCAGAAATGCCAGAA
 E  T  T  E  D  T  E  V  E  T  M  T  E  A  P  G  T  A  N  D  T  E  M  E  T  A  E  M  P  E
  K  L  Q  K  T  L  K  L  K  Q  *  R  K  H  R  E  L  Q  M  I  L  K  W  K  Q  Q  K  C  Q  K
G  N  Y  R  R  H  *  S  *  N  N  D  G  S  T  G  N  C  K  *  Y  *  N  G  N  S  R  N  A  R

      5050       5060       5070       5080       5090       5100       5110       5120       5130
ACTACGGAAGATATTGAAATTGAAACATCGACCGAGGCATCCGACTTTTTCAACAAATTTGGGATTAAATCCTTCATGATATTTATAAAA
 T  T  E  D  I  E  I  E  T  S  T  E  A  S  D  F  F  N  K  F  G  I  K  S  F  M  I  F  I  K
  L  R  K  I  L  K  L  K  H  R  P  R  H  P  T  F  S  T  N  L  G  L  N  P  S  *  Y  L  *  N
N  Y  G  R  Y  *  N  *  N  I  D  R  G  I  R  L  F  Q  Q  I  W  D  *  I  L  H  D  I  Y  K

      5140       5150       5160       5170       5180       5190       5200       5210       5220
TTAGTAAAAAAATGTAAAAATCGGCAGATGAAGAGATGTACACAGGATCATACGGAGAATTGACTCGATGATTTGTGGCATAAATGGAATTT
 L  V  K  N  V  K  S  A  D  E  E  M  Y  T  G  S  Y  G  E  L  T  R  *  F  V  A  *  M  E  F
  *  *  K  M  *  N  R  Q  M  K  R  C  T  Q  D  H  T  E  N  *  L  D  D  L  W  H  K  W  N  F
I  S  K  K  C  K  I  G  R  *  R  D  V  H  R  I  I  R  R  I  D  S  M  I  C  G  I  N  G  I

      5230       5240       5250       5260       5270       5280       5290       5300       5310
CTGCCGGAAAGTAATGACTCATAACTTTTGACTACCTAGATGGAACGTCAATTTCTATATTGATCATTACGTGCATCAAACACAGTCTTT
 L  P  E  S  N  D  S  *  L  L  T  T  *  M  E  R  Q  F  L  Y  *  S  L  R  A  S  N  T  V  F
  C  R  K  V  M  T  H  N  F  *  L  P  R  W  N  V  N  F  Y  I  D  H  Y  V  H  Q  T  Q  S  L
S  A  G  K  *  *  L  I  T  F  D  Y  L  D  G  T  S  I  S  I  L  I  I  T  C  I  K  H  S  L

      5320       5330       5340       5350       5360       5370       5380       5390       5400
AAATCTTGAATTATCTCTTGAAATTTAATTTATAAAATTTAACAAACATTTTTAGCAGACATCATCAACAACAGCAGCACCGGAAACCAC
 K  S  *  I  I  S  *  N  L  I  Y  K  I  *  Q  T  F  L  A  D  I  I  N  N  S  S  T  G  N  H
  N  L  E  L  S  L  E  I  *  F  I  K  F  N  K  H  F  *  Q  T  S  S  T  T  A  A  P  E  T  T
*  I  L  N  Y  L  L  K  F  N  L  *  N  L  T  N  I  F  S  R  H  H  Q  Q  Q  Q  H  R  K  P

      5410       5420       5430       5440       5450       5460       5470       5480       5490
AGAAGGCACTGAAGTTGAAACAACGACGAGCGTTGTTAAACCCGGTTTGTTATTTTCAAGTTGAAAAAGCTATTCCGTTTCTCTTATTAT
 R  R  H  *  S  *  N  N  D  E  R  C  *  T  R  F  V  I  F  K  L  K  K  L  F  R  F  S  Y  Y
  E  G  T  E  V  E  T  T  T  S  V  V  K  P  G  L  L  F  S  S  *  K  S  Y  S  V  S  L  I  I
Q  K  A  L  K  L  K  Q  R  R  A  L  L  N  P  V  C  Y  F  Q  V  E  K  A  I  P  F  L  L  L

      5500       5510       5520       5530       5540       5550       5560       5570       5580
CATTTGGATAATGGTCAGTCATCTGAGTTACTGCTTGTGCATTCTAGGCAGTACAGAAAATACTCATGCAAAGATGAGAAATAACTTAGG
 H  L  D  N  G  Q  S  S  E  L  L  L  V  H  S  R  Q  Y  R  K  Y  S  C  K  D  E  K  *  L  R
  I  W  I  M  V  S  H  L  S  Y  C  L  C  I  L  G  S  T  E  N  T  H  A  K  M  R  N  N  L  G
S  F  G  *  W  S  V  I  *  V  T  A  C  A  F  *  A  V  Q  K  I  L  M  Q  R  *  E  I  T  *

      5590       5600       5610       5620       5630       5640       5650       5660       5670
AAATCTTCAGTATTTTTTTTAAAATGAATATTTATTAAAATTTGATAAGCACACCGACATGGATCACTGGCAAATCATTTAAAATATTCA
 K  S  S  V  F  F  L  K  *  I  F  I  K  I  *  *  A  H  R  H  G  S  L  A  N  H  L  K  Y  S
  N  L  Q  Y  F  F  *  N  E  Y  L  L  K  F  D  K  H  T  D  M  D  H  W  Q  I  I  *  N  I  Q
E  I  F  S  I  F  F  K  M  N  I  Y  *  N  L  I  S  T  P  T  W  I  T  G  K  S  F  K  I  F

      5690       5690       5700       5710       5720       5730       5740       5750       5760
ATATGTTTATTATTTCTTCATTGCCACATATGCCGCATACAATTTCTTCATTCGTTTTTTGTCTCGAAAAACTTCTCGATTCTACGGTAT
 I  C  L  L  F  L  H  C  H  I  C  R  I  Q  F  L  H  S  F  F  V  S  K  N  F  S  I  L  R  Y
  Y  V  Y  Y  F  F  I  A  T  Y  A  A  Y  N  F  F  I  R  F  L  S  R  K  T  S  R  F  Y  G  M
N  M  F  I  I  S  S  L  P  H  M  P  H  T  I  S  S  F  V  F  C  L  E  K  L  L  D  S  T  V
```

```
      5770       5780       5790       5800       5810       5820       5830       5840       5850
GAGAACCCTGAAACAATGAATGCACTACTGAGATCTGAAATCTATTTTAAGGTGAGGAATGTCCGGAGCCCGATGGTTTATTTCCGCATC
 E  N  P  E  T  M  N  A  L  L  R  S  E  I  Y  F  K  V  R  N  V  R  S  P  M  V  Y  F  R  I
  R  T  L  K  Q  *  M  H  Y  *  D  L  K  S  I  L  R  *  G  M  S  G  A  R  W  F  I  S  A  S
*  E  P  *  N  N  E  C  T  T  E  I  *  N  L  F  *  G  E  E  C  P  E  P  D  G  L  F  P  H

      5860       5870       5880       5890       5900       5910       5920       5930       5940
ACAGTGATTGCCATTTGTTCATTAATTGCGCAAATAACTATCCACACATAATGGAATGTCCGGTGGGCACCTTTTTTGATGACACAATCA
 T  V  I  A  I  C  S  L  I  A  Q  I  T  I  H  T  *  W  N  V  R  W  A  P  F  L  M  T  Q  S
  Q  *  L  P  F  V  H  *  L  R  K  *  L  S  T  H  N  G  M  S  G  G  H  L  F  *  *  H  N  Q
H  S  D  C  H  L  F  I  N  C  A  N  N  Y  P  H  I  M  E  C  P  V  G  T  F  F  D  D  T  I

      5950       5960       5970       5980       5990       6000       6010       6020       6030
AAGTTTGCAATTATATGAGGAACGCGCCGGATACATGCAAATGATTTGGCATTTTATGTGATATATATATTTGTTGGAATAAAAATCTAG
 K  F  A  I  I  *  G  T  R  R  I  H  A  N  D  L  A  F  Y  V  I  Y  I  F  V  G  I  K  I  *
  S  L  Q  L  Y  E  E  R  A  G  Y  M  Q  M  I  W  H  F  M  *  Y  I  Y  L  L  E  *  K  S  R
K  V  C  N  Y  M  R  N  A  P  D  T  C  K  *  F  G  I  L  C  D  I  Y  I  C  W  N  K  N  L

      6040       6050       6060       6070       6080       6090       6100       6110       6120
GGTAAAAGTTCTAGAACAAAAAAAAAAAGACGCATAAATCATATTAAATAATCTGCAAAAATTGCATTAAAAATTTATTGAATTTCAAGG
 G  K  S  S  R  T  K  K  K  R  R  I  N  H  I  K  *  S  A  K  I  A  L  K  I  Y  *  I  S  R
  V  K  V  L  E  Q  K  K  K  D  A  *  I  I  L  N  N  L  Q  K  L  H  *  K  F  I  E  F  Q  G
G  *  K  F  *  N  K  K  K  K  T  H  K  S  Y  *  I  I  C  K  N  C  I  K  N  L  L  N  F  K

      6130       6140       6150       6160       6170       6180       6190       6200       6210
AACGAGTTTGCATGTCATTTGCTCTCTCTAAAGTATTATTTCAAACCCACTCCTGAATCTTTCACAATTTAAAAATTCTTTCAGCGAAAT
 N  E  F  A  C  H  L  L  S  L  K  Y  Y  F  K  P  T  P  E  S  F  T  I  *  K  F  F  Q  R  N
  T  S  L  H  V  I  C  S  L  *  S  I  I  S  N  P  L  L  N  L  S  Q  F  K  N  S  F  S  E  I
E  R  V  C  M  S  F  A  L  S  K  V  L  F  Q  T  H  S  *  I  F  H  N  L  K  I  L  S  A  K

      6220       6230       6240       6250       6260       6270       6280       6290       6300
ATAAGCTATAAATGAATAATATTGATTCAATTTTATTAGTTAATCATCCAACTAAATCTTGAAAGCATAAAGATGTTCAATTTATGAAAC
 I  S  Y  K  *  I  I  L  I  Q  F  Y  *  L  I  I  Q  L  N  L  E  S  I  K  M  F  N  L  *  N
  *  A  I  N  E  *  Y  *  F  N  F  I  S  *  S  S  N  *  I  L  K  A  *  R  C  S  I  Y  E  T
Y  K  L  *  M  N  N  I  D  S  I  L  L  V  N  H  P  T  K  S  *  K  H  K  D  V  Q  F  M  K

      6310       6320       6330       6340       6350       6360       6370       6380       6390
TTTTTTTAATAAAAAATTAAAACATAATTTACATAAAAAAGTAATGATACGCATTAAACAATTTAATACATTATTTAACATTTATAAATT
 F  F  *  *  K  I  K  T  *  F  T  *  K  S  N  D  T  H  *  T  I  *  Y  I  I  *  H  L  *  I
  F  F  N  K  K  L  K  H  N  L  H  K  K  V  M  I  R  I  K  Q  F  N  T  L  F  N  I  Y  K  L
L  F  L  I  K  N  *  N  I  I  Y  I  K  K  *  *  Y  A  L  N  N  L  I  H  Y  L  T  F  I  N>

      6400       6410       6420       6430       6440       6450       6460       6470       6480
ATATCAGTACAATGCAAGCATCCTCCAAATGAATAAACAAACAACCGCTAAAGAAGAAAAAGTCCCAACAAATCATTTAGTCATTCAATA
 I  S  V  Q  C  K  H  P  P  N  E  *  T  N  N  R  *  R  R  K  S  P  N  K  S  F  S  H  S  I
  Y  Q  Y  N  A  S  I  L  Q  M  N  K  Q  T  T  A  K  E  E  K  V  P  T  N  H  L  V  I  Q  *
Y  I  S  T  M  Q  A  S  S  K  *  I  N  K  Q  P  L  K  K  K  K  S  Q  Q  I  I  *  S  F  N

      6490       6500
AACTACTCACGCCAAATGTAATTTCA
 N  Y  S  R  Q  M  *  F  X
  T  T  H  A  K  C  N  F  X
K  L  L  T  P  N  V  I  S
```

Appendix 4: Consecutive alignment of exon L (Gene I) to exon Z (Gene III)

```
L   ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ACACCATCAACAACAGCAGCACCGGAAACCACAGAAGACACTGAAGTTGAAACAATGACG 60

    ************ ***** ********* ******************** ********

L   ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ------------ACGGAAGCACCGGGAACTGCAAATGACACTGAAATGGAAACAACAAAT 108
                 *  * ** ***** **** ** * ********* * *** ***  *

L   ------------
Z   ATACCAGAAATT 120

L   ------------ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ACAGAAGATACTGAAACTGAAACATCGGAAATGCCGGAAACTACAGATAATACTGAAATG------------ 180
                  * *   ****   * * * **************  * ****** *

L   ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ----------------------CCAGGAACTGCAGATGATACTGAAACTGAAACATCGGAA 219
                         ** * **** **** ** ******  **** **  *
L   ----------------------
Z   ATGCCGGAAACTACAGATAAT 240

L   ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ------------ACTGAAATGCCGGGAACTGCAGATGATACTGACGAAACATCGGAAATG 288
                *  * * * **** **** **** ** ***** *    * *    * *
L   ------------
Z   CCGGAAACTACA 300

L   -ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG60
Z   GAAATTACAGAGGATACTGAAACTGAAACATGGGAAATGCCGGAAACTACAGATGATATT-420
     * *  *   *  * *     ** * **  *    **  *    ** *  *    **

L   ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ---------------GAAATGCCAGGAACTGCAGATGATACTGAAACTGAAACATCGGAA 465
                  * * * ** * **** **** ** ******  **** **  *

L   ---------------
Z   ATGCCGGAAATTACA 480

L   ACACCATCAACAAGA--GCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGA 58
Z   --------GATAATACTGAAATGCCGGGAACTGCGGATGATACTGAAATGGAAACAACAG 532
          * ** *  * * * **** **** * ** ** ****** * *** ***

L   CG------ 60
Z   AAATGCCA 540

L   -ACACCATCAACAAGAGCAGTACCGGAAACT--ACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   GAAATTACAGAAGATACTGACACTAAAACATCGACCGATGCATCTGAAACTACAGAAGAT--- 600
     * *  *   *  * *     **   **  *  ** ** *   *****  *  ** *
```

```
L   --ACACCAT-CAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   GCAGATGATACTGAAACTGAGATATCGGAAATGCCGGAAACTACAGATAGTAGTGAAATG--- 720
      * *  ** *   **  *   ** ******   * ***   ** **   *   * ****

L   ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ----------------------CCAGGAACTGCAGATGATACTGAAACTGAAACATCGGAA 759
                         ** * **** **** ** ******  **** **  *

L   ----------------------
Z   ATGCCGGAAACTACAGATAAT 780

L   ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ------------ACTGAAATGCCAGGAACTTCGGATGATACTGAAATGGAAACAACAGAA 828
               *  * * * ** * **** * ** ** ****** * *** ***

L   ------------
Z   ATGCCAGAAATT 840

L   ------------ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ACAGAAGATACTGAAACTGAAGCATCAGAAATGCCGGAAACTACAGATAATACTGAAATG------------900
                  * *   * **  ** * * **************  * ****** *

L   ----ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   CCAGGAACTGCGGATGAT-GCTGAAATGGAAACAACAGAAATGCCAGAAATTACAG--AGGAT- 960
          **     *  *  ** * *  ****** ******    * *** **  **  * **

L   ---ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ATTGAAACTGAAACATCGGAAATGCCGGAAACTACAGATAATAGTGAAATGCCAGGAACT--- 1020
         * *   ****   * * * **************  * * **** *   ** **

L   --ACACCAT-CAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   GCAGATGATACTGAAACTGAAACATCGGAAATTCCGGAAACTACAGATAATACTGAAATG--- 1080
      * *  ** *   **  * *  * ****** * * ***   ** **   *   * ****

L   ----ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   CCAGGAACTGCGGATGATA-CTGAAATGGAAACAACAGAAATGCCAGAAATTACAG--AGGAT- 1140
          **     *  * * * * *  ****** ******    * *** **  **  * **

L   ---ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ATTGAAACTGAAACATCGGAAATGCCGGAAACTACAGATAATAGTGAAATGCCAGGAACT--- 1180
         * *   ****   * * * **************  * * **** *   ** **

L   ------------ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTT 48
Z   GCAGATGATACTGAAACTGAAACATCGGAAATGCCGGAAACTACAGATAATACTGAAATG 1240
                  * *   ****   * * * **************  * ****** *

L   GAAGCAATGACG 60
Z   ------------

L   ---ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATG 57
Z   CCAGGAACTTCGGATGATACTGAAATGGAAACAACAGAAATGCCAGAAATTACAGAAGAT 1300
         * * **        * * *  ****** ******    * *** **  ** *
```

```
L   ACG 60
Z   ---
L   ---ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATG 57
Z   ACTGAAACTGAAGCATCAGAAATGCCGGAAACTACAGAGGGGGCTGAAACTGAAACAACG 1360
         * *   * **  ** * * ************** *   *****  **** *** *

L   ACG 60
Z   ---
L   ACACCATCAACAAGAGCAGTACCGGAAACTACAGAAGACACTGAAGTTGAAGCAATGACG 60
Z   ------------ACCGAAGCATCAGAAACTACGGAAACTATTAAACCCACCAAAAAACCT 1408
                 *  * ** * * ******** ***   * * **        **   *
L   ----
Z   TCTG 1412
```

Appendix 5: Arguments for differences and similarities in gene sequences of clone 1 and 9

1) Identities in intergenic region between clones= 97 (Differences below)

```
One          591   AGATGAAGAAAAGTATTTTGCAGTCTCAGATGAAGAAAAGTATTTTGCAGTTGCTT 700
Nine         564   ---------------------------AGATGAAGAAAAGTATTTTGCAGTTGCTT 673
                                               ****************************

One             1856  TTTCAGTTAC-TGTGAACTTCA--AACTGGTAATTAAACCATAGCAATGAT----AGTAA
Nine          1830 TTTCAGTTACATAGCAGCGTTGTCGTCTTCTTCTCTGACCGTGATAATATTTACGGGTTT
                    ********** *   * * *      **  *  *   *** *   ***  *     **

One            AAACACAATATTAAAACAAGAAACAAGCGGTACAA 1944
Nine           GATTGCAATCTT----TCAGTTACANGCGGTGCAA 1922
                *   **** **      **  *** ***** ***
```

2) Introns and Exons of identical regions of clones 1 and 9 showing discrepancies

I versus III

Exon	Identities	Intron	Identities	Exon	Identities	Intron	Identities
		f	37	A	100	a	100
G	**95***	g	100	B	100	b	100
H	100	h	100	C	100	c	100
I	100	i	100	D	100	d	100
J	100	j	100	E	**71**	e	**53**
M	100						

9 Publications and conference abstracts

Manuscript from this study

Babila J. Tachu, Smitha Pillai, Richard Lucius, and Thomas Pogonka: *Acanthocheilonema viteae* Chitinase Genes: Structure and Expression. (In preparation).

Conference abstracts from this study

Babila J. Tachu, Richard Lucius, Jörg Hirzmann, Vincent P.K.Titanji, Thomas Pogonka, (2003): Molecular characterisation of *Acanthocheilonema viteae* chitinase. Meeting of the Federation of African Societies for Biochemistry and Molecular Biology held in Yaounde, Cameroon.

Babila J. Tachu, Richard Lucius, Jörg Hirzmann, Vincent P.K. Titanji, Thomas Pogonka (2004): Molecular characterisation of *Acanthocheilonema viteae* chitinase genes. 21st Meeting of the German Society for Parasitology meeting held in Würzburg, Germany.

Michel J. Sereda, Smitha Pillai, Babila J. Tachu, Susanne Hartmann, Thomas Pogonka, and Richard Lucius (2005): Studies on subunit vaccines against infection with the parasitic nematode *Acanthocheilonema viteae* in a rodent model. Vaccine Congress Berlin 2005