Structural & Functional characterization of fatty acids and retinoid binding proteins from nematodes

• Structural & Functional characterization of fatty acids and retinoid binding proteins from nematodes

Thesis · November 2015

DOI: 10.13140/RG.2.1.2459.8487

1 author:

Umer Naveed Chaudhry
The University of Edinburgh
55 PUBLICATIONS **153** CITATIONS

SEE PROFILE

Structural and Functional Characterization of Fatty acid & Retinoid binding proteins (FAR) from Nematodes

Honours Thesis in Biomedicine, 30 ECTS

(15/05/2009 - 20/11/2009)

Report Version II

Umer Naveed Chaudhary
Msc Biomedicine
University of Skövde Sweden
a08chaum@student.his.se

Supervisor

Dr. Paul A. Tucker
Group Leader
European Molecular Biology
Laboratory Hamburg Germany
tucker@embl-hamburg.de

Dr. Rositsa Jordanova
European Molecular Biology
Laboratory Hamburg Germany
rosie@embl-hamburg.de

Examiner
Dr. Karin Klinga Levan
School of Life Sciences
Biomedicine/Systems Biology
University of Skövde
karin.klinga.levan@his.se

Abstract

Parasitic nematodes cause serious diseases in animals, humans and plants. Chemoprophylaxis offers short-term benefits, but due to rapid development of drug resistance in parasites there is a pressing need for novel treatments of nematode infection. Parasitic nematodes secrete a structurally novel class of fatty acid and retinol binding (FAR) protein into the surrounding tissues of the host. These proteins are of interest because they may play an important role in scavenging fatty acids and retinoid from the host that are essential for the survival of the parasite and also because the localised depletion of such lipids may have immunomodulatory affects that compromise the host immune response. The genome of the free-living nematode *Caenorhabditis elegans* encodes eight FAR proteins (Ce-FAR-1 to 8) these fall into three discrete groups as indicated by phylogenetic sequence comparisons and intron positions, the proteins from parasitic nematodes falling into group A.

The first goal of this project was to characterized Ce-FAR proteins by using bioinformatics tools. Second goal was to examine Ce-FAR-1 that includes the expression in competent *Escherichia coli* cells, purification by metal ion affinity chromatography, size exclusion chromatography, crystallization and crystals were cryo-protected with and without glycerol (v/v) for X-ray diffraction followed by bioinformatics based structure predication. Third goal was creation of point mutation in Ce-FAR-6 and Ce-FAR-7, expression analysis in competent *E.coli* cells and purification by metal ion affinity chromatography followed by lipid removal and dialysis for comparative ligand binding studies of Ce-FAR-6 mutant using fluorescence spectroscopy and Ce-FAR-7 mutant was phosphorylated *in vitro* with casein kinase II to measure with mass spectroscopy.

Key words: FAR, *Caenorhabditis elegans*, metal ion affinity chromatography, size exclusion chromatography, crystallization, X-diffraction.

Table of contents

Abbreviations

FAR	Fatty acid & retinoid binding proteins
Ce	*Caenorhabditis elegans*
Ce-FAR	*Caenorhabditis elegans*- Fatty acid and retinoid binding protein
LB	Laria Bertani
E. Coli	*Escherichia coli*
PCR	Polymearse Chain Reaction
cDNA	Copy Deoxyribonucleic Acid
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
IPTG	Isopropyl-Beta-D-Thiogalactopyranoside
BME	ß-mercaptoethanol
CKII	Casein kinase II
EST	Expressed sequence tag
Ni	Nickel
PBS	Phosphate buffered saline

1. Introduction

Nematodes are one of the most abundant groups of multicellular organisms on earth and parasitic forms directly cause more human disease and economic damage to domestic animals and crop plants than any other group of metazoan organisms, with the possible exception of insects (Garafalo *et al.,* 2003). Approximately one sixth of the earth's population, mostly in developing countries, suffers from nematode infections. Four of the fifteen tropical diseases, recorded by the world health organization (WHO) are caused by parasitic nematodes (www.who.int/neglected_diseases/diseases/en/).

It was not until the mid 1980s, however, that nematode-specific retinoid-binding proteins with characteristics different from those of human tissues and plasma were finally identified (Sani *et al.,* 1998). Evidence was built up that nematodes parasites require retinoid for three reasons. First, retinol plays an important role in gene activation, cell signalling, tissue differentiation and repair. Second, there is evidence that parasitic nematodes may require retinol variety of their metabolic and developmental processes, such as growth, differentiation, embryogenesis, glycoprotein biosynthesis and as antioxidants (Gudas *et al.,* 1994). Third, retinol deficiency can alter the character of the host immune response and it is possible that such changes could be beneficial to the parasite (Nikawa *et al.,* 1999). Parasitic worms are unable to synthesise the lipid *de-novo* and depend on import of essential lipids from their host, which makes the lipid binding proteins a good targets for new treatments (Kennedy *et al.,* 1997).

Two structurally novel families of lipid-binding proteins have been identified in parasitic nematodes. These are nematode polyprotein allergens/antigens (NPAs) with molecular weights of approximately 15 kDa and the fatty acid and retinol binding (FAR) proteins with molecular weights of approximately 20 kDa (Kennedy *et al.,* 2000). The first FAR protein to be identified was the Ov-FAR-1 from the filarial agent *Onchocerca volvulus*, which cause river blindness in humans. A recombinant fragment of that protein was originally used as an immunodiagnostic tool to detect infection with *O. volvulus* (Bradley *et al.,* 1991). Ten more FAR proteins from filarial species, all causing serious sickness in humans and animals have been studied. They share high sequence homology (79-100%) and belong to two major clusters i.e. nodule species such as Ov-FAR-1 and Bm-FAR-1 from *Brugia malayi* (causing elephantiasis) and some other lymphatic filarias (Garofalo *et al.,* 2002). Since the characterisation of Ov-FAR-1, genes encoding similar proteins have been identified in a wide range of nematodes, including parasites of humans, animals, and plants (Prior *et al.,* 2001).

The secreted FAR proteins of parasitic nematodes may play a role in the pathogenesis of infection. The ocular and dermal disease manifestations of onchocerciasis, for example, are very similar to those associated with vitamin A deficiency and may reflect the localised depletion of this vitamin by Ov-FAR-1, produced and secreted by populations of *O. volvulus* concentrated in high numbers within the onchocercal nodules (Bradley *et al.,* 2001). Retinol deficiency is immunosuppressive, and it is also possible it could alter the character of the host immune response to the advantage of the parasite. In addition to retinol, nematode FAR proteins are known to bind other lipid modulators of the host immune system such as arachidonic acid, and it may be relevant in this context that some species of tissue dwelling parasitic nematodes have been shown to adsorb and modify host lipids for secretion as biologically active eicosanoids (Liu

et al., 1992). Onchocercal nodules are rich in collagen and preliminary studies have indicated that Ov-FAR-1 may convey a retinoid messenger to tissues of the host which acts as a trigger to induce the formation of these collagenous structures (Nirmalan 1999).

Genes encoding FAR proteins have now been described from many species of parasitic nematode (Tree et al., 1995, Blaxter, 1998, Prior et al., 2001), and inspection of expressed sequence tag and other surveys (see nema.cap. ed.ac.uk/Nest.html and Ref. 28) indicates that some or all parasitic nematodes may produce one or two types of FAR proteins (http://www.nematodes.net) (Garofalo et al., 2002). Recently it was identified that there are one *NPA* and six *FAR* genes in *Haemonchus contortus* and life stage gene expression of Hc-FAR-1 to 6 revealed unique transcription patterns for each of these genes (Kuang et al., 2009).

To understand the function of FAR proteins in parasitism, it is important to establish how many genes encoding FAR proteins exist in given specie, the function of the encoded proteins and one important in host-parasite interactions. At present functional characterisation of the FAR proteins has been impeded by the difficulties encountered in attempting to maintain and manipulate parasitic nematodes in a laboratory setting. None of the parasites can complete their life cycle outside the host. The free living nematode *Caenorhabditis elegans* is one of the most powerful model organisms in developmental biology (Gilleard, 2004). *C. elegans* has six chromosomes and 18,400 annotated genes currently available in wormbase data base (http://www.wormbase.org/, 2009). In the genome of the *C. elegans*, eight FAR proteins have been identified. They belong to three groups, (group A include Ce-FAR-1, 2, 6, group B include Ce-FAR-3, 4, 5 and group C include Ce-FAR-7, 8) (Garafalo *et al.*, 2003). The genes encoding the Ce-FAR proteins were found to be transcribed differentially through the life cycle of *C. elegans*, such that Ce-FAR-4 was weakly transcribed only at fourth larval stage, Ce-FAR-5 weakly transcribed in all stages and Ce-FAR-1, Ce-FAR-2, Ce-FAR-3, Ce-FAR-6 and Ce-FAR-7 predominated in males.

The FAR proteins of parasitic nematodes seem to be secreted from the cell into the extra cellular environment, where they presumably play a role in the sequestration and transport of fatty acids and retinoid for the parasite (Garofalo et al., 2002). It is not known at this stage whether any of the *C. elegans* FAR proteins are secreted by the nematodes, and it is possible that the release of these proteins into the surrounding host tissue by parasites may reflect an adaptation of their function for parasitism. In order to explore further possibilities for targeting FAR proteins for novel drug development, obtaining detailed structural information is essential but there are only two 3D structures of Lipid binding proteins from parasitic worms in protein data bank and both have FABP folds. There is a report of NMR structure of NPA protein (Meenan et al., 2005). Recently the first high resolution crystal structure of Ce-FAR-7 from *C. elegans* has been reported (Jordanova et al., 2009).

In this study, we first characterized Ce-FAR proteins by using bioinformatics tools. Second goal was to express recombinant Ce-FAR-1 in *E. coli* cells, purified it by metal ion chromatography, size exclusion chromatography followed by crystallization, X-ray diffraction and bioinformatics based structure predication. Third goal was the creation of a point mutation in Ce-FAR-6 T50D and Ce-FAR-7 T26A, their expression and purification, followed by lipid removal and dialysis.

2 Materials and Methods

2.1. Characterization of Ce-FAR Protein

Ce-FAR genes were identified by tblastn and blastp searches of the *C. elegans* sequence data base available online at wormbase data base (http://www.wormbase.org/). High scoring matches were inspected to determine the relevance of the match, and a total of eight *FAR* genes (*Ce-FAR-1* to *Ce-FAR8*) were found (Appendix I) previously described by Garafalo et al., 2003. The SignalP program (SignalP 3.0 Server, 2009) was used to predict the presence and location of any signal peptide cleavage sites. The molecular weight and isoelectric point of the proteins were estimated using the ProtParam program (Expasy protparam server, 2009). Consensus sites for various post-translational modifications can be predicated in each of the *Ce-FAR* protein by using Expasy data base server (http://www.expasy.ch, 2009). The conserved amino acid sequences, hydrophilic & hydrophobic position, secondary structures, disulphide bridges and coil-coiled prediction of Ce-FAR-1 to Ce-FAR-8 were predicted using several algorithms programs run by geneDoc (http://motif.genome.jp/), ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html), JPRED server (http://www.compbio.dundee.ac.uk/Software/JPred/jpred.html, DSD Base (http://caps.ncbs.res.in/dsdbase//dsdbase.html) and Ch.EMBnet.org (http://www.ch.embnet.org/software/COILS_form.html).

2.2. Ce-FAR-1 protein

2.2.1. Large scale expression and Purification

Ce-FAR-1 was cloned in pETM-11 LIC previously done by Fatima (Appendix II) and further transformed in *E. coli* BL21(DE3)pLysS competent cells. The few colonies from the LB agar plate were inoculated into $LB^{Kan\&Chm}$ and incubated over night at 37^0C. The pre-culture was poured into 1lit $LB^{Kan\&Chm}$ and the optical density (OD_{600}) were measured after every one hr until it reached 0.6-0.8. Isopropyl-beta-$_D$-thio-galactopyranoside IPTG (1mM) was used to induce the cells for Ce-FAR-1 protein expression and the culture was grown overnight at 20^0C. The cells were spun down at 5500 rpm for 25 min; the pellet was washed with 1X PBS buffer and frozen at -20^0C for further use.

The pellet of Ce-FAR-1was dissolved in binding buffer (20mM Tris base pH 8.5, 50mM NaCl, 5mM β-mercaptoethanol and 20mM Imidazole) with one tablet per 50 ml of Complete Protease Inhibitor Cocktail, Roche, Germany. The cells were lysed by sonication. The high frequency (>20 kHz) of the sound waves cause the cells to burst open. Since heat is generated, samples must be kept on ice and the sonication was limited to short pulses (0.3 to 0.8 sec). After cell disruption the obtained lysate was centrifuged to remove the unwanted cell debris. The collected supernatant contained the solube protein fraction. Supernatant was filtered with 0.22µ filter paper for the removal of other cell debris.

The supernatant was applied to a gravity-flow column (Bio-Rad) packed with Ni-NTA affinity resin (Qiagen, USA). The unbound bacterial proteins were removed from the column using 5-7 column volume of binding or washing buffer (20mM Tris pH 8.5, 50mM NaCl, 5mM β-

mercaptoethanol, 20mM Imidazole). The N-terminal His-tagged Ce-FAR-1 protein was eluted from the column using an elution buffer (20mM Tris pH 8.5, 50mM NaCl, 5mM β-mercaptoethanol, 400mM Imidazole). The protein was further purified by gel filtration on a Superdex 75 HR 16/60 column (Pharmacia) pre-equilibrated with 20mM Tris pH 8.5, 50mM NaCl, 5mM β-mercaptoethanol. Ce-FAR-1 was concentrated using a Millipore concentration kit (Millipore, USA) for crystallization trials. The concentration of the protein was measured spectrophotometrically using the nanodrop method. The molar extinction coefficient (at 280 nm) for the proteins was estimated from their tyrosine and tryptophan content. The purified protein from affinity and gel filtration chromatography was analysed by SDS-PAGE according to the protocols and buffer recipe of Invitrogen, USA, 2009. After running, the gel was then stained with coomassie blue overnight on a rocking platform followed by de-staining and washing.

2.2.2. Crystallization and X-ray diffraction

Purified Ce-FAR-1 protein was sent to the high throughput crystallization facility at EMBL Hamburg for screening. The sparse matrix screens were set up using sitting drop vapour diffusion method in 96 well plates. It is a common method for initially determining potential crystallisation conditions. A wide selection of buffers, pH values, salts, additives and precipitants were screened. The plates were stored at 20°C and were automatically imaged for remote inspection. The best conditions were optimized manually with Grid screening also set up for hanging drop vapour diffusion, followed by streak and micro seeding. Crystals were grown on Easy Xtal tool 24 well plates (Qiagen, 2009) by equilibrating a mixture containing 1μL of protein and 1μL of a reservoir solution against 1ml of reservoir solution.

X-ray diffraction analysis of single crystals with suitable dimensions was carried out on the X13 beamline, EMBL at the synchrotron radiation facility DESY in Hamburg, Germany and ID29 beamline at the European synchrotron radiation facility Grenoble, France. For X-ray diffraction experiments, crystals were transiently soaked in glycerol (v/v) corresponding to the reservoir solution but test supplemented with 10%, 15%, 20% and 25% (v/v) glycerol or the reservoir solution without glycerol as a cryo-protectant with the help a fine loop and mounted into a goniometer, which allow it to be positioned accurately within the X-ray beam and rotated.

2.2.3. Bioinformatics based structure prediction

The homology search against Ce-FAR-1 was performed by using NCBI blastp searches (www.ncbi.nlm.nih.gov, 2009) and domain organization was predicted using program run by motif search server (http://motif.genome.jp/). The secondary structure of Ce-FAR-1 and Ce-FAR-7 was predicted using the consensus of several algorithms run by the JPRED server (http://www.compbio.dundee.ac.uk/Software/JPred/jpred.html) and a multiple sequence alignment was generated using the program ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

The X-ray structural model of Ce-FAR-7 were available at 1.79 Å resolution [PDB: 2w9y] and were used as a template to generate the 3D model of Ce-FAR-1. For this purpose, we choose input de novo protein modelling program ESyPred3D Web Server 1.0 (http://www.fundp.ac.be/

sciences/Biologie/urbm/bioinfo/esypred/). The sequence of Ce-FAR-1 contains 166 amino acids, while Ce-FAR-7 contains 138 amino acids. High-resolution model for Ce-FAR-1 was constructed on the basis of their functional and secondary structural homology with Ce-FAR-7. The 3D structure of Ce-FAR-1 was validated with the PROCHECK (http://deposit.pdb.org/cgi-bin/validate/adit-session-driver). The program generates ramachandran plot and accuracy. The 3D structure of Ce-FAR-1 was used to predict the amino acid residues in the active site domain by using Q-site Finder server (ligand binding site predication) (http://www. modelling. leeds.ac.uk/qsitefinder/)

2.3. Ce-FAR-6 and Ce-FAR-7 proteins

2.3.1. Mutagenesis, Sequencing and Test expression

Ce-FAR-6 and Ce-FAR-7 were amplified from *C.elegans* cDNA and cloned into pETM-11 and pETM-11 LIC vectr previously done Rositsa (Appendix II). The specific primers were designed for the creation of point mutation in Ce-FAR-6 and Ce-FAR-7 by using the QuikChange® Primer Design Program (Stratagene, 2009). The primers were designed to generate the desired mutations. The desired mutation was present in both of the mutagenic primers and it annealed to the opposite plasmid strand on the same sequence. Master Mix was prepared for PCR having all the reagents and its composition is given in QuikChange® II Primer kit (Stratagene 2009). The thermocycler was run according to the program given in QuikChange® Primer kit. The amount of template used was 125ng, annealing temperature (T_A), length of the primers along with melting temperature (Tm) and guanine and cytosine content are given in Table 1.

Table1: Description of Ce-FAR-6 and Ce-FAR-7 primers for mutagenesis (Forward and Reverse).

Assigned Name	Forward Primer	Reverse Primer	T_A	T_M	Length	GC %
Ce-FAR-6 T50E	AAAGTTACTGAATACCTTA AATCCATCGAGACTGAGG AGAAAGCTGCTATTAAGG AA	TTCCTTAATAGCAGCTTTC TCCTCAGTCTCGATGGAT TTAAGGTATTCAGTAACTT T	60 ° C	73 ° C	*FP 57 *RP 57	*FP 36.8 *RP 36.8
Ce-FAR-6 T50D	AGTTACTGAATACCTTAAA TCCATCGATACTGAGGAG AAAGCTGCTATTAAG	CTTAATAGCAGCTTTCTC CTCAGTATCGATGGATTT AAGGTATTCAGTAACT	60 ° C	72° C	*FP 52 *RP 52	*FP 36.5 *RP 36.5
Ce-FAR-7 T26A	CTCGAGTTCTCCTCATCA ATTGCCGCTGACGAG	CTCGTCAGCGGCAATTGA TGAGGAGAACTCGAG	60 ° C	73° C	*FP 33 *RP 33	*FP 54.5 *RP 54.5

* FP: Forward Primer, RP: Reverse Primer

Forward and reverse primers for Ce-FAR-6 T50E, T50D and Ce-FAR-7 T26A were used for site directed mutagenesis by replacing the amino acid threonine (Thr) with asparatic acid (Asp), glutamic acid (Glu) and alanine (Ala) at position number 50 of Ce-FAR-6 and position number 26 of Ce-FAR-7 respectively. The mutant non-methylated DNA was obtained by digestion of the methylated wild type DNA with endonuclease *Dpn-1* by using the QuickChange®II Site-Directed Mutagenesis Kit (Stratagene, 2009). The *DpnI*-treated plasmid DNA was then

transformed into *E. coli* (XL1-Blue) super competent cells, plated on LB^{Kan} plates over night at $37^{o}C$. Four colonies from Ce-FAR-6 T50D mutant plate, one from Ce-FAR-6 T50E mutant plate and six colonies from Ce-FAR-7 T26A mutant plate were transferred to a falcon tube with LB^{Kan} and incubated overnight at $37^{o}C$. All the procedures were performed according to the QuickChange®II Site-Directed Mutagenesis Kit (Stratagene, 2009).

Plasmid purification was done from the culture of bacteria harbouring pETM 11 and pETM-11 LIC plasmid along with Ce-FAR-6 T50D and Ce-FAR-7 T26A mutant insert by using a Miniprep kit (Qiagen, Germany). The DNA concentration was measured using the nanodrop method. Ce-FAR-6 T50D and Ce-FAR-7 T26A clones were sequenced by Eurofins GmbH using T7 Fwd Primer. The obtained nucleotide sequence was then aligned with the wild type Ce-FAR-6 and Ce-FAR-7 using the software Lalign (Expasy, 2009).

The positive Ce-FAR-6 T50D and Ce-FAR-7 T26A clones were transformed in *E. coli* BL21(DE3)pLysS, BL21(DE3)RIL and BL21(DE3)RP competent cells on $LB^{Kan\&Chm}$ for test expression according to the protocols of Qiagen Bench Guide, USA, (2009). Colonies were inoculated into $LB^{Kan\&Chm}$ and incubated over night at $37^{0}C$ followed by pre-culturing. Optical density (OD_{600}) of the cultures was measured every half an hour until reaching 0.6-0.8. 1ml sample was taken before IPTG induction and 1ml sample after each time point (every half an hour) and the pellets were dissolved in SDS sample buffer (Novagen, 2009) and boiled. 1ml sample was taken at the end of expression and and the cells were lysed with Bug Buster lysis buffer (Novagen, 2009) and centrifuged. The supernatant was mixed with SDS sample buffer and boiled. The expression was analysed by SDS PAGE.

2.3.2. Large scale expression and Purification

Ce-FAR-6 T50D and Ce-FAR-7 T26A were expressed in BL21(DE3)RIL and BL21(DE3)pLysS *E. coli* competent cells respectively for large scale expression. Colonies were inoculated into $LB^{Kan\&Chm}$ and incubated over night at $37^{0}C$. The expression was performed as described in *Expression and purification of* Ce-FAR-1.

Ce-FAR-6 T50D and Ce-FAR-7 T26A were purified by a nickel-column affinity and gel filtration chromatography according to the protocols described in *Expression and purification of* Ce-FAR-1. The purity of protein was analysis by SDS-PAGE according to the protocols and buffer recipe of Invitrogen, 2009. After running, the gel was then stained with Coomassie blue overnight followed by de-staining and washing.

2.3.3. Lipidex incubation, Dialysis and Phosphorylation

The purified Ce-FAR-6 T50D mutant protein was further treated with Lipidex-1000 (Perkin-Elmer Life & Analytical Sciences, 2009) followed by dialysis for buffer exchange. The Lipidex was first washed with water and then equilibrated with 1X PBS. The matrix was mixed with the protein and incubated at $37^{0}C$ for 1hr on a shaker for two serial incubations. The supernatant was filtered on a 22µ filter and put into the dialysis cassette (Thermo Scientific, 2009) for overnight dialysis against gel filtration buffer (Tris pH 8.5, 50mM NaCl, 5mM β-mercaptoethanol) at $4^{0}C$ for comparative ligand binding studies using fluorescence spectroscopy.

The purified Ce-FAR-7 T26A mutant protein was further dialyed for buffer exchange. The supernatant was filtered on a 22µ filter and put into the dialysis cassette (Thermo Scientific, 2009) for overnight dialysis against gel filtration buffer (Tris pH 8.5, 50mM NaCl, 5mM β-mercaptoethanol) at 4^0C. The Ce-FAR-7 T26A mutant dialysed product was *in-vitro* phosphorylated with casein kinase II (NEB, 2009) using 500 U/ul CKII per 4 pg protein in the presence of 1mM ATP (NEB) and 1X CKII buffer (NEB) along with a non phosphorylated control mixed in water. The mixtures were incubated for 3 hr at 37^0C.

3 Results and Discussion

Nematodes are probably the most abundant and ecologically diverse group of multicellular organism on Earth, and parasitic forms directly cause more human diseases and economic damage to domestic animals and crop plants than any other group of metazoan organisms, with the possible exception of insects. Lipid binding proteins released by nematode parasites have attracted increasing interest because of their potential role in nutrient acquisition, manipulation of the tissues they occupy, and countering host defence reactions (Kennedy, 2001).These proteins may interfere with intercellular lipid signalling to manipulate the defence reactions of the host or acquire essential lipids for the parasites. Nematodes produce two different types of small retinol and fatty acid binding proteins (14–20 kDa), neither of which have recognizable counterparts in other animal groups, including mammals. Both are helix-rich, in contrast to the beta rich small lipid transporters proteins of vertebrates such as the lipocalins, which are extracellular and secreted (Flower, 1996) and members of the FABP/P2/CRBP/ CRABP1 family (which are cytoplasmic lipid transporters, with exceptions only from nematodes) (Mei, 1997). One of these two families of helix rich lipid binding proteins of nematodes is synthesized as a large nematode polyproteins antigen (NPA) that is post-translationally cleaved into multiple copies of the functional protein of 14 kDa. The other family comprises the fatty acid and retinol binding proteins (FAR), which are the subjects of this report.

Fatty acids and retinoids are poorly water soluble, oxidation-sensitive and therefore require specific carrier proteins to facilitate their intracellular and intercellular transport through aqueous environments (Flower, 1996). An early indication that parasitic nematodes may possess retinoid-binding proteins that differ significantly from those of their host came from work with *O. volvulus* (Sturchler et al,. 1981). The retinol concentration within the onchocercal nodule was found to be eight times greater than that of the surrounding host tissue, suggesting that the parasite has the ability to sequester retinol from the carrier proteins of the host. It was not until the mid-80s, however, that nematode-specific retinoid-binding proteins with characteristics different from those of human tissues and plasma were finally identified (Sani *et al,*. 1985).

Currently more than 40 FAR genes have been documented, although some of these are only represented by short read ESTs (Blaxter, 1998) (http://nema. cap.ed.ac.uk/Nest.html). The genome of the nematode *C. elegans* encodes eight FAR like proteins (Ce-FAR-1 to 8), that fall into three discrete groups as indicated by phylogenetic sequence comparisons as shown in Fig. 1 (Garofalo *et al.,* 2003).

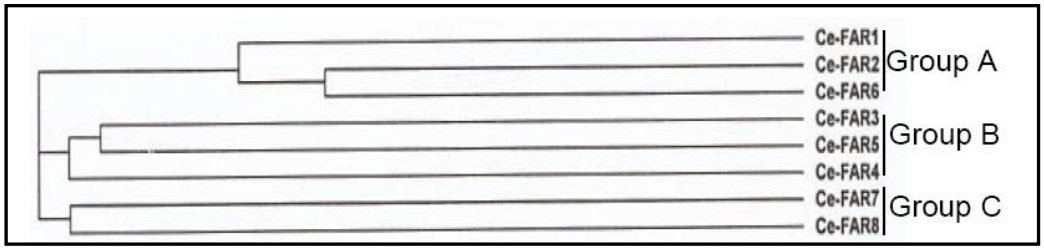


Figure 1: Phylogenetic relationships between the FAR protein sequences of *C. elegans*.

The FAR proteins of parasitic nematodes that have been used for functional studies cluster most closely together with the group comprising Ce-FAR-1, Ce-FAR-2 and Ce-FAR-6, and this group of Ce-FAR proteins may therefore most closely resemble the ancestral homologue from which the parasite proteins evolved (Garofalo *et al.,* 2003). The basic aims of this study are already mentioned, the focus was the bioinformatics characterization of Ce-FAR proteins, mutagenesis and sequence analysis of Ce-FAR-6 T50D and Ce-FAR-7 T26A, expression analysis of Ce-FAR-1, Ce-FAR-6 T50D and Ce-FAR-7 T26A in *E.coli*, protein purification by affinity and size exclusion chromatography and crystallization.

3.1. Characterization of Ce-FAR protein

3.1.1. Classification

Eight *far* genes of *C. elegans* including assigned name, gene ID, groups, chromosomal localization and size (wormbase data base, 2008) are presented in Table 2 as initally described by Garofalo *et al.,* 2003.

Table 2: *far* genes are classified on the basis of assigned name, gene ID, group, chromosomal location and size.

Assigned Name	Gene ID	Groups	Chromosomal Localization	Size (Nucleotide)	Size (Amino acid)
Ce-FAR-1	WB00001385	A	III	549	182
Ce-FAR-2	WB00001386	A	III	549	182
Ce-FAR-3	WB00001387	B	V	570	189
Ce-FAR-4	WB00001388	B	V	618	205
Ce-FAR-5	WB00001389	B	V	531	176
Ce-FAR-6	WB00001390	A	IV	555	184
Ce-FAR-7	WB00001391	C	II	417	138
Ce-FAR-8	WB00001392	C	III	699	232

Eight Ce-FAR proteins according to their molecular weight, iso-electric point and extinction coefficient (Expasy protparam tool, 2009) has been shown in Table 3 as initally described by Garofalo *et al.,* 2003.

Table 3: Ce-FAR proteins are characterized on the basis of molecular weight, iso-electric point and extinction coefficient.

Assigned Name	Molecular weight	Iso-electric point	0.1%Extinction coefficient
Ce-FAR-1	20KDa	6.98	0.445
Ce-FAR-2	20KDa	5.73	0.521
Ce-FAR-3	20KDa	8.76	0.356
Ce-FAR-4	23KDa	10.09	0.322
Ce-FAR-5	19KDa	5.17	0.377
Ce-FAR-6	20KDa	8.63	0.434
Ce-FAR-7	15KDa	6.20	0.113
Ce-FAR-8	26KDa	5.26	0.952

All Ce-FAR proteins, with the exception of Ce-FAR-7, possess a hydrophobic secretory signal peptide, and the proteins are thus predicted to be secreted from the cell into the extracellular environment, where they presumably play a role in the sequestration and transport of fatty acids and retinoids. The signal sequence of Ce-FAR proteins was obtained in the sequenced fragment using SignalP 3.0 Server (2009) shown in Fig. 2 in italic and underscored red (Garofalo *et al.,* 2003). Immunolocalisation studies have shown that the FAR proteins of filarial nematodes are released by the parasites into the extracellular environment (Tree *et al.,* 1995) and similar results have also been reported for the plant parasite *G. pallida* (Prior *et al.,* 2001). It is not known at this stage whether any of the Ce-FAR proteins are infact secreted by the nematode. Recently it has been shown that Ce-FAR-7, which does not have a signal peptide, is not secreted in to the extracellular environment (Jordanova, *et al.,* 2009).

3.1.2. Posttranslational Modifications of Ce-FAR protein

Consensus sites for various post-translational modifications can be identified in each of the FAR proteins, but most of these are short, commonly occurring sequences. With the exception of N-linked glycosylation (Prior *et al.,* 2001), no evidence has yet been provided that any of them are utilised *in vivo*. All the FAR proteins of parasitic nematodes have putative N-linked glycosylation sites, but the number and location of the sites is not always conserved (Garofalo *et al.,* 2002) and the FAR proteins of some parasite species do not appear to be glycosylated (Tree *et al.,* 1995, Garofalo *et al.,* 2002). It is interesting that only those filarial parasites with unsheathed microfilarae produce glycosylated proteins. It is possible that the FAR proteins play

an important role in parasitism and the absence of glycosylation is essential to enable their secretion through the microfilarial sheath into the surrounding host tissue, but the true biological significance of the difference in glycosylation between filarial parasites with sheathed and unsheathed microfilarae remains to be established (Garofalo *et al.*, 2002). Three of the Ce-FAR proteins (Ce-FAR-4, Ce-FAR-6, Ce-FAR-8) that have putative N-linked glycosylation sites are underscored in Fig. 2 (NetNGlyc 1.0 server, 2009) but in Ce-FAR4, the glycosylation site is coincident with the putative casein kinase II phosphorylation site and it may not therefore be utilised. N-linked glycosylation sites were not observed in other Ce-FAR proteins (Garofalo *et al.*, 2003).

The putative casein kinase II phosphorylation site of Ce-FAR-1 to Ce-FAR-8 that is highlighted boldface type with blue mark in Fig. 2 (NetPhos 2.0 Server, 2009), however, is conserved in all the nematode FAR proteins that have been characterised to date, and it may therefore be of importance (Garofalo *et al.*, 2003). Phosphorylation is known to have a significant affect on the biological activity of many proteins, including those involved in gene regulation (Ventura *et al.*, 2001) and it can control homodimerisation (Surette *et al*,. 1996) and even the stability of alpha-helices (Szilak *et al*,. 1997).

```
Ce-FAR1  --miratiilaavaalafsAPVPEVPENYDDIP----------------AEYKSLIPAEV 42
Ce-FAR2  --miraflvvalasvavfsAPIPEVPQNFDDIP----------------AEYKGLIPAEV 42
Ce-FAR3  msrlfafnvfclvllrfsaaAPADDSSPFSQIL----------------KQHKDLLPSEV 44
Ce-FAR4  ----msklllllvlsllffitsaFPFGEPQAGGF----------------QKFKNLLPREL 40
Ce-FAR5  -----mllrffaillifhfsfaADG--IFEAVI----------------ESYKDALPPKV 37
Ce-FAR6  --miriflvialasvavfsAPISRLPQSLDDIP----------------AEFKELIPAKV 42
Ce-FAR7  -------------------------MSVASLP----------------ECVKNFFPTEQ 18
Ce-FAR8  mftlracslilvsvavfvlgrPAEEQMTEKDFTNLVFTVEKFDQILKAYSEYKQFMPSYV 60

Ce-FAR1  SEHLKSITPEEKAILKEVAKGYK----DFKSEDDFLNALKEKSPTLHEKASKLHQIVKDK 98
Ce-FAR2  AEHLKAITAEEKAALKELAQNHK----EYKTEEEFKAALKEKSPSLYEKAGKLEALLTAK 98
Ce-FAR3  VQAYQDLSPEEKAALKDVFKNYK----SYKNEGELIAALKEKSSSLGEKAEKLQAKLQKK 100
Ce-FAR4  VEAYSNLSQKDQPDLKDVFRNHQ----NYRNEQEMVNALKMKNPALGARMERRLMALKQK 96
Ce-FAR5  VAAFDNLSPGESAIMKEVFMNYD----KFTSIADLIVAIKKKSESLGSFFEKLYIEIDAE 93
Ce-FAR6  TEYLKSITTEEKAAIKEFIKSVMG---GNKSVEELSADIKERSPSLYAKVEKLDVLLRTK 99
Ce-FAR7  LEFSSSITADEKPVLHEVFQKHS----CFSQCGEMIDEVSKKHPELGKRLATVLEGNKKR 74
Ce-FAR8  LEPLDNITDEQKTQAVQMVNDYHAGKFEPKNYDEYIAIMKKSYPALAGPYETMYNKYKEQ 120

Ce-FAR1  VNALNDE-AKAFVKKAIAEGRKIHAQYLAGEKPSLDTLKTTAKTHIEAYKGLSQDAK--D 155
Ce-FAR2  FEKLDAT-AQALVKKIIAKGRELHQQYLAGDKPTLDSLKELAKGYIAEYKALSDDAK--A 155
Ce-FAR3  VDALSPK-PKDFVNELIAGGRGLYARSVNGEKISVSEIKLLIETQVAAYKALPAEAQ--D 157
Ce-FAR4  IDGLSSEEAKGFIQNLISTGRQIYAQRLNGQQMDQSQLRQVGMGIAMHYRSLPPYAQ--Q 154
Ce-FAR5  IQALTPE-TKKFVTEMLGIGRAIYTAQIVGIPLDSKEVLPVFAKQFTSFKSLSDATK--E 150
Ce-FAR6  LAKLDPA-ALALFGKVIAQGLSFRQQFHAGYQPTPEMVKKLFKGYIAEYKTLSENAK--A 156
Ce-FAR7  LDGLSPA-AVEYAKKLIHMVTTTLCSLTVGKPIDDADAKRLHQ----EFQSLSSEDQ--A 127
Ce-FAR8  VAKLGPK-GQEYCNGLEAQ---MYTDASPDRVVWACHIFNNAKSAVSGAKALLQDDSEAA 176

Ce-FAR1  SIAKEFPILTGFFK---NEKVQAMVGQYLN-------------------------- 182
Ce-FAR2  TITAEFPILTGFFQ---NEKIQAIVGQYVN-------------------------- 182
Ce-FAR3  ELKKNFGGVAKFLE---DDKTQTLIAKLLEKNNNQ--------------------- 189
Ce-FAR4  ELQSTFPQIFQFMRQMREQRLRSMMGGFFGGGGGIGMGQGMGQGMGMGMGK----- 205
Ce-FAR5  ELEKTFLGLYKFAS---NDKIKTEIDKLL-------------------------- 176
Ce-FAR6  TITDEFPIVVEFFQ---HEKIQAIIQQIVNY------------------------- 184
Ce-FAR7  ALRKNNPDIKF------------------------------------------- 138
Ce-FAR8  KIEEAFPEAVKLLNSKKFEAYSIIVNNLKPLDCVKDREQVFNTIKLMDKQSVLTNN 232
```

Figure 2: Multiple sequence alignment of Ce-FAR protein. The putative hydrophobic signal peptides that were predicated by the signalP program are shown in italic/lowercase letters and underscored. Consensus N-linked glycosylation site are underscored, and conserved casein kinase II phosphorylation site is in boldface type.

3.1.3. Secondary structure of Ce-FAR Protein

FAR proteins are identical and secondary structure prediction programmes indicate that Bm-FAR-1 has the same helix-rich structure as Ov-FAR-1 (Garofalo *et al.*, 2002). Multiple sequence alignment of Ce-FAR proteins, hydrophilic and hydrophobic positions of amino acids and secondary structure predictions indicate that the proteins are rich in alpha helix, with no significant content of beta extended structure as shown in Fig. 3 (Garofalo *et al.*, 2003). Circular dichroism data agree with the predictions, as shown for Ov-FAR-1 (Kennedy et al., 1997), Ce-FAR-1 to 6 (Garofalo *et al.*, 2003) and Ce-FAR-7 (Kostova *et al.*, 2009, unpublished). The high helical content of the FAR family was confirmed by the crystal structure of Ce-FAR-7 (PDB ID 2w9y), showing nine helices and no beta structures in the protein (Jordanova *et al.*, 2009). The predominance of helical structure thus appears to be a universal characteristic of the parasitic proteins and *C. elegans* FAR do not seem to differ significantly in this respect from those of the parasites. It is a factor that clearly sets them apart from the similar sized lipid-binding proteins of vertebrates, such as the lipocalins and members of the cytoplasmic lipid transporters family of proteins, which are all beta rich (Noy, 2000, Cowan *et al.*, 1993).

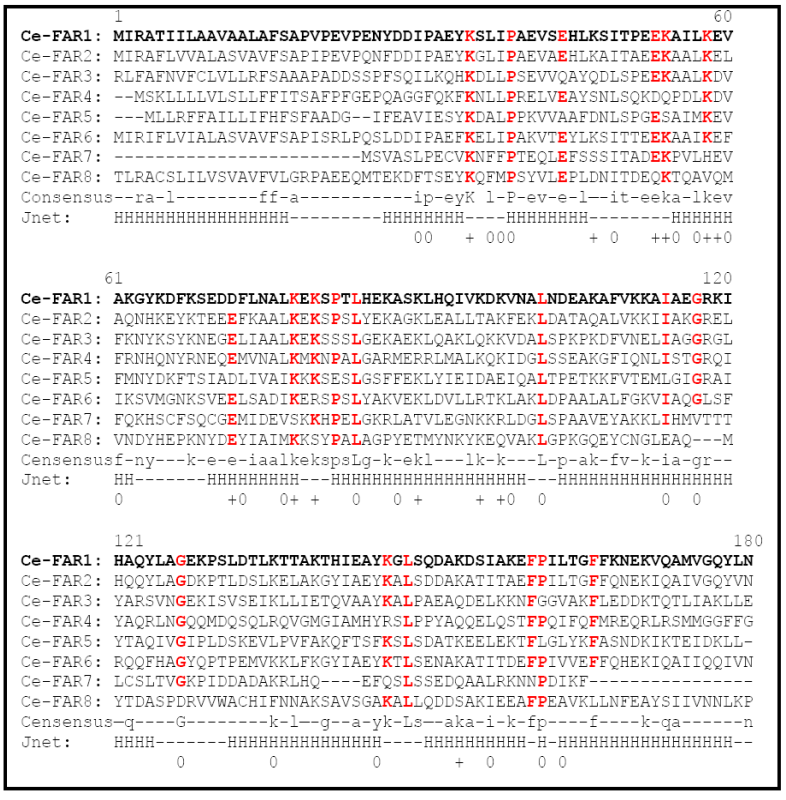


Figure 3: Ce-FAR protein relationships. A multiple sequence alignment and secondary structural prediction for the *C. elegans FAR* proteins. In the consensus line, uppercase letters refer to those amino acids that are conserved at that position in all of the sequences, and lowercase letters refer to amino acids that are conserved at that position in more than half of the sequences. The Jnet line shows the secondary structure prediction from submission of the multiple alignments to the Jpred secondary structure prediction program. H represents a prediction for alpha helix, and the gaps indicate regions for which no structural prediction emerged. No beta structure was predicted by Jpred or any other secondary structure prediction programs. The bottom line indicates conserved hydrophobic (O) and hydrophilic (+) positions.

A characteristic feature of Ce-FAR-7 and Ce-FAR-8 is the possession of four (including signal peptides) cysteines, respectively, in the predicted mature proteins. The other Ce-FAR proteins possess no cysteines, except Ce-FAR-3, which has one in the putative leader/signal peptide (position 11). The cysteine positions in Ce-FAR-7 (9, 41, 46, 97) and Ce-FAR-8 (6, 132, 152, 208) are dissimilar, and the fourth cysteine in Ce-FAR-8 may be lost if the C-terminal extension of this protein is trimmed post-translationally (Fig. 4). Given the weighting typically attributed to the occurrence and position of cysteines in protein families, it is arguable that Ce-FAR-7 and Ce-FAR-8 are not typical of the FAR family. They may however represent specialized or ancestral forms of FAR proteins with distinctive functions (Garofalo *et al.*, 2003).

```
Ce-FAR3  MSRLFAFNVFCLVLLRFSAAAPADDSSPFSQIL----------------KQHKDLLPSEV 44
Ce-FAR7  --------------------------MSVASLP---------------ECVKNFFPTEQ 18
Ce-FAR8  MFTLRACSLILVSVAVFVLGRPAEEQMTEKDFTNLVFTVEKFDQILKAYSEYKQFMPSYV 60


Ce-FAR3  VQAYQDLSPEEKAALKDVFKNYK----SYKNEGELIAALKEKSSSLGEKAEKLQAKLQKK 100
Ce-FAR7  LEFSSSITADEKPVLHEVFQKHS----CFSQCGEMIDEVSKKHPELGKRLATVLEGNKKR 74
Ce-FAR8  LEPLDNITDEQKTQAVQMVNDYHAGKFEPKNYDEYIAIMKKSYPALAGPYETMYNKYKEQ 120


Ce-FAR3  VDALSPK-PKDFVNELIAGGRGLYARSVNGEKISVSEIKLLIETQVAAYKALPAEAQ--D 157
Ce-FAR7  LDGLSPA-AVEYAKKLIHMVTTTLCSLTVGKPIDDADAKRLHQ----EFQSLSSEDQ--A 127
Ce-FAR8  VAKLGPK-GQEYCNGLEAQ---MYTDASPDRVVWACHIFNNAKSAVSGAKALLQDDSEAA 176


Ce-FAR3  ELKKNFGGVAKFLE---DDKTQTLIAKLLEKNNNQ---------------------- 189
Ce-FAR7  ALRKNNPDIKF------------------------------------------------ 138
Ce-FAR8  KIEEAFPEAVKLLNSKKFEAYSIIVNNLKPLDCVKDREQVFNTIKLMDKQSVLTNN 232
```

Figure 4: Multiple sequence alignment of Ce-FAR protein (Ce-FAR3, Ce-FAR-7 and Ce-FAR8). Figure shows Cysteines residue in Ce-FAR3, Ce-FAR-7 and Ce-FAR8 highlighted in Bold/red mark.

The most notable structural feature associated with parasite FAR proteins is the strong prediction for coiled-coil structures (Prior *et al.*, 2001, Garofalo *et al.*, 2002, Kennedy *et al.*, 1995). There was a strong prediction for the occurrence of coiled-coils in Ce-FAR-1, Ce-FAR-2 and Ce-FAR-3, but not for the others. The significance of these observations is unclear. Coiled-coil interactions can impart significant stability to proteins and such interactions could underlie the finding that FAR proteins are extremely thermostable (Kennedy *et al.*, 1997). Where two stretches of predicted coiled-coil occur within a single protein, coiled-coil interactions could either act between the two regions/domains of the same protein molecule or be involved in intermolecular associations and perhaps even be responsible for specific recognition between molecules (Burkhard *et al.*, 2001).

3.2. Ce-FAR-1 protein

3.2.1. Expression and Purification

The Ce-FAR-1 (His-tag) construct was transformed into BL21(DE3)pLysS cells and expression was induced by IPTG. Growth was monitored every one hour until OD_{600} was between 0.6-0.8 (Fig. 5). The constructs were successfully expressed and the pellets were stored for futher purification of the desired protein.

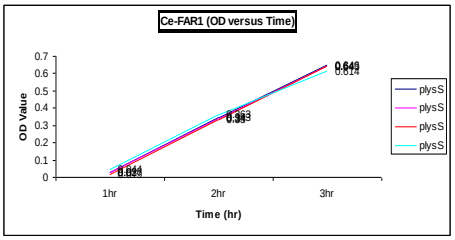


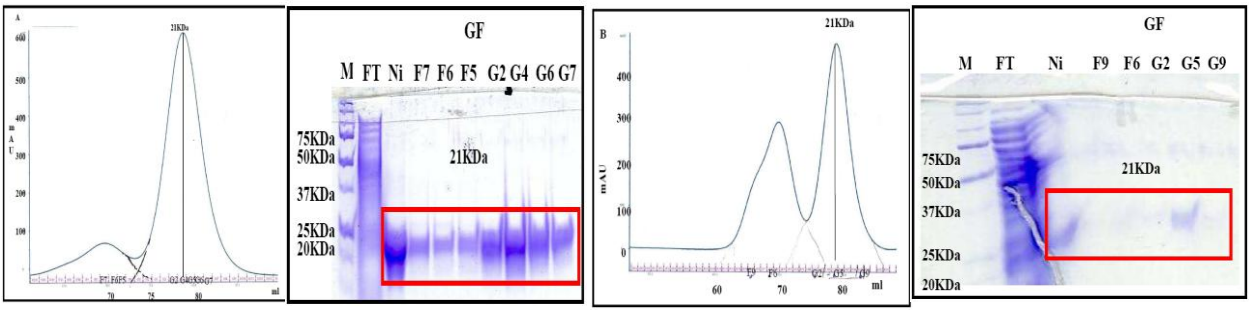
Figure 5: Growth curve of BL21(DE3)pLysS. The plot shows the OD value against time of Ce-FAR-1 (h).

A two steps purification of the Ce-FAR-1 protein was carried out by affinity chromatography by using a nickel column followed by gel filtration. Cell lysates from *E. coli* BL21(DE3)pLysS containing recombinant Ce-FAR-1 were applied to a Ni column. Ce-FAR-1 protein was eluted and concentrated. It was applied to a Superdex 75 column for gel filtration. The protein peak was detected by A_{280} absorbance with an Äkta Purifier system (HPLC) as described by Jinsong *et al.*, 2009. The fraction eluted from Superdex 75 columns contained pure protein and it was concentrated for further crystallization experiments. The final yield of pure Ce-FAR-1 from 2 L culture is given in Table 4.

Table 4: Protein yield during two step purification of Ce-FAR-1.

Purification Steps	Protein (mg/ml)	Protein (mg/ml)	Protein (mg/ml)	Protein (mg/ml)	Protein (mg/ml)	Protein (mg/ml)
Ni-affinity column	4.5 mg/ml	4 mg/ml	3 mg/ml	4.6 mg/ml	5.5 mg/ml	2.5 mg/ml
Superdex 75 column	8.5 mg/ml	8.9mg/ml	6 mg/ml	5.5 mg/ml	5 mg/ml	6 mg/ml

The purification was repeated six times and the purity of protein collected for crystallization was analysed by SDSPAGE as shown in Fig 6.



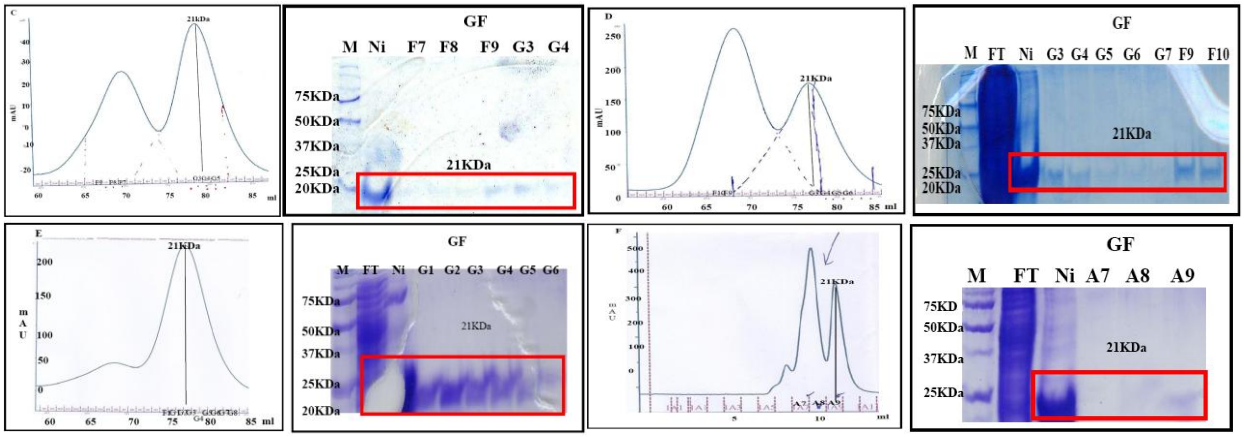


Figure 6: (A, B, C, C, D, E and F) the peaks show the profile of gel filtration chromatography of Ce-FAR-1 protein. The volume in ml is plotted on X axis and on Y axis is the absorbance at 280nm. An SDS-PAGE image shows the profile of affinity and gel filtration chromatography of *Ce-FAR-1* protein. The size of the product was 21KDa compared with marker (M) and FT is the Flow through from affinity chromatography. In SDS-PAGE image, affinity chromatography (Ni) shows 21KDa band along with the bands of gel filtration chromatography (GF) fraction taken from micro well plates.

3.2.2. Crystallization and X-ray diffraction

Purified protein Ce-FAR-1 after gel filtration was further used in the crystallization experiment with concentrations between 6-9 mg/ml. Five conditions {(i) 10% PEG4000, Hepes pH 7.5, 20% Isopropenol (ii) 10% PEG8000, Hepes pH 8 (iii) 8% PEG8000, Tris pH 8.5 (iv) 10% PEG6000, MES pH 6.0 (v) 10% PEG6000, Hepes pH 7.0} gave promising crystals from the sparse matrix screens set up at the high throughput crystallization facility at EMBL Hamburg (Fig. 7A). Further the condition 13% PEG6000 as precipitant with 0.1M MES buffer pH 6.4 was selected for screening with an additive screen (Hampton Research). Three additives were chosen after getting the results for crystal optimization as shown in Fig.7B - 0.1M Adenosine-5-triphosphate disodium salt hydrate, 40% v/v Pentaerythritol ethoxylate (3/4 EO/OH), 50% v/v Jeffamine M-600 pH 7.0 (Hampton research additive screening, 2009).

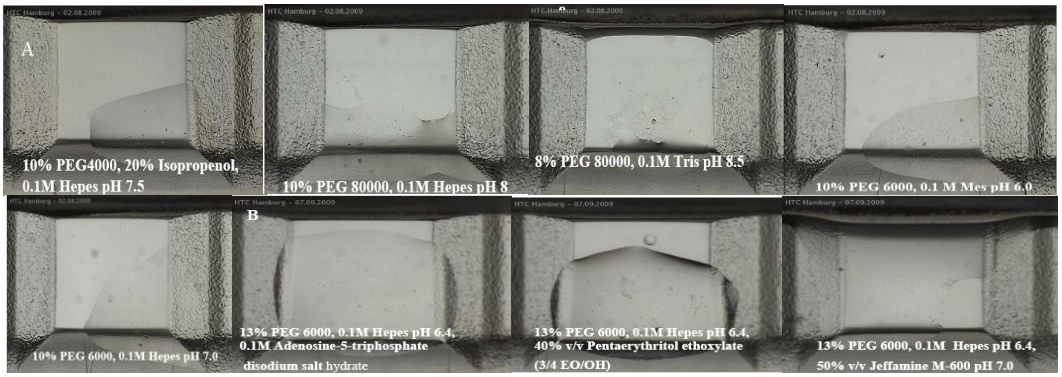


Figure 7 (A) HTC Hamburg screening for crystallization of Ce-FAR-1 protein. All five conditions show small crystal formation of Ce-FAR-1 having 10% PEG4000, Hepes pH 7.5, 20% Isopropenol, 10% PEG8000, Hepes pH 8, 8% PEG8000, Tris pH 8.5, 10% PEG6000, MES pH 6.0, 10% PEG6000, Hepes pH 7.0. (B) Images from HTC Hamburg additive screening show three best conditions having 13% PEG6000 in 0.1M MES pH 6.4, along with 0.1M Adenosine-5-triphosphate disodium salt hydrate and 12%, 14% and 15% PEG6000 in 0.1M MES pH 6.6, 40% v/v Pentaerythritol ethoxylate (3/4 EO/OH).

Several grid optimization screens were set up with Polyethylene glycol 6000 (PEG6000) as a precipitant in the concentration of 10 % to 15% and 0.1M MES buffer at pH 6.0 to 6.8. After few days, small rectangular shape crystals appeared (Fig. 8B). In order to obtain bigger crystals and check diffraction, initial conditions were further optimized by seeding techniques or crystallization with additives. Streak seeding was performed as well, the idea behind is to differentiate the process of crystal growth from nucleation and obtain bigger crystals. In general, the degree of super saturation required for nucleation is higher than that required for crystal growth. When crystal seeds are added (take a small drop from best crystal from other screening), the equilibrium can shift towards crystal formation, and avoids the random nature of spontaneous nucleation. The growth was satisfactory having single crystals as shown in Fig. 8D. Finally, successful micro seeding and co-crystallization with additives were achieved. The additives, that gave the best results were 0.1M Adenosine-5-triphosphate disodium salt hydrate, 40% v/v Pentaerythritol ethoxylate (3/4 EO/OH) and 50% v/v Jeffamine M-600 pH 7.0 from Hampton Additive screen (Fig. 7B). Larger rectangular shape crystals were obtained (Fig. 8E, F & G).

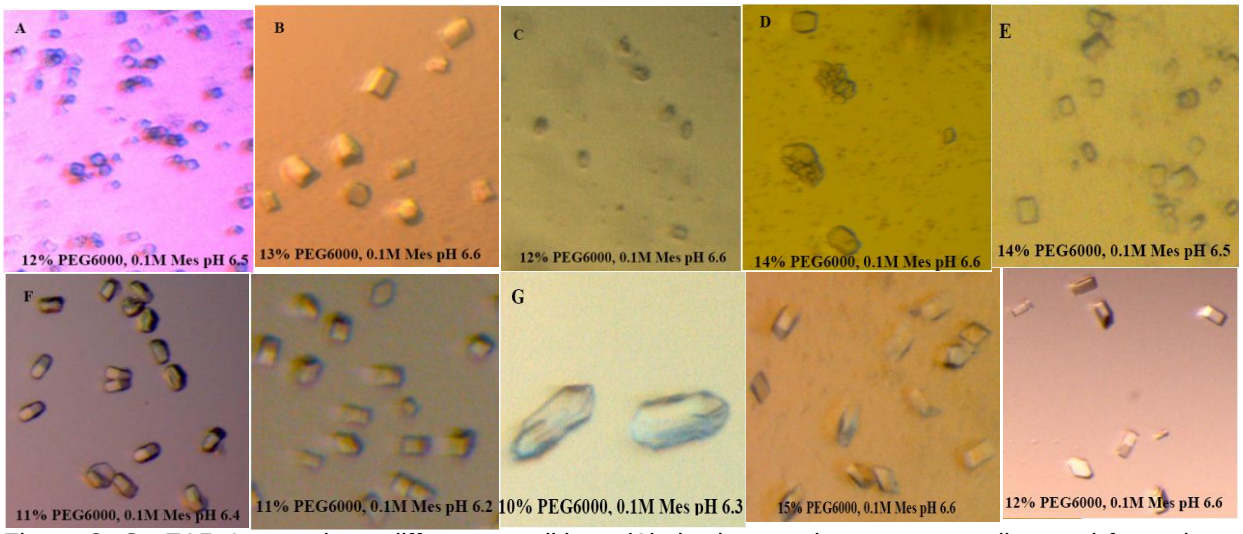


Figure 8: Ce-FAR-1 crystals at different conditions (A) the image shows very small crystal formation of Ce-FAR-1 with 12% PEG6000 and 0.1M MES 6.5. (B) The Image shows small rectangular shape crystals of Ce-FAR-1 with 13% PEG6000 and 0.1M MES pH 6.6. (C) The image shows small streak seeded crystals of Ce-FAR-1 into 12% PEG6000 and 0.1M MES pH 6.6 or (D) 14% PEG6000 and 0.1M MES pH 6.5. (E) The image shows large micro seeded crystals of Ce-FAR-1 into 14% PEG6000 and 0.1M MES pH 6.6 or (F) 11% PEG6000 and 0.1M MES pH 6.4 or 0.1M Mes pH 6.2 (G) The Image shows the crystals of Ce-FAR-1 with 10% PEG6000 and 0.1M MES pH 6.3and 0.1M Adenosine-5-triphosphate disodium salt hydrate as an additive and 12% and 15% PEG6000 in 0.1M MES pH 6.6, 40% v/v Pentaerythritol ethoxylate (3/4 EO/OH) as an additive.

Single crystals of above conditions with suitable dimension were selected for X-ray diffraction tests. Crystals were either mounted directly from the well solution, since it contained high percentage of PEG or where necessary cryo-protected with 10% to 25% glycerol (v/v). Unfortunately, so far Ce-FAR-1 did not diffract better than 20 Å and further crystal optimization is ongoing.

3.2.3. Structural model of Ce-FAR-1

The homology search against Ce-FAR-1 suggested that the amino acid sequence of Ce-FAR-1 show 27% identity (red mark) and 51% similarity (blue mark) with Ce-FAR-7 (Group C) as shown in Fig. 9A. The alignment score was high as 36.6 bits and E value was low as 2e-07 (date not shown). E value indicate that the number of hits see when searching a database of a particular size and alignment score is the sum of the scores specific for each of the aligned pairs of letters and as the value alignment score is high when the similarity will be high. The predications of domain indicate that there was single domain present in Ce-FAR-1 protein. Secondary structure prediction programme shows that Ce-FAR-1 has the same helix-rich structure as Ce-FAR-7. Multiple sequence alignment indicates that Ce-FAR-1 has nine alpha helixes containing 116 residues and 50 residues are coiled as shown in Fig. 9A.

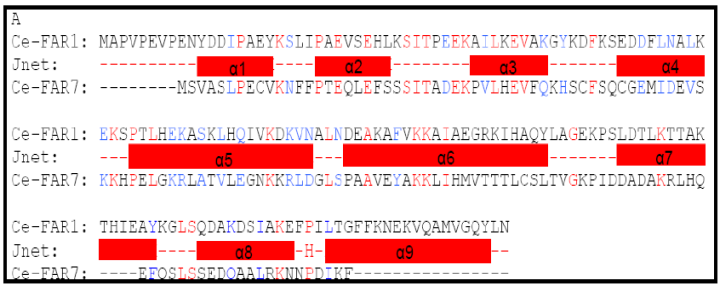
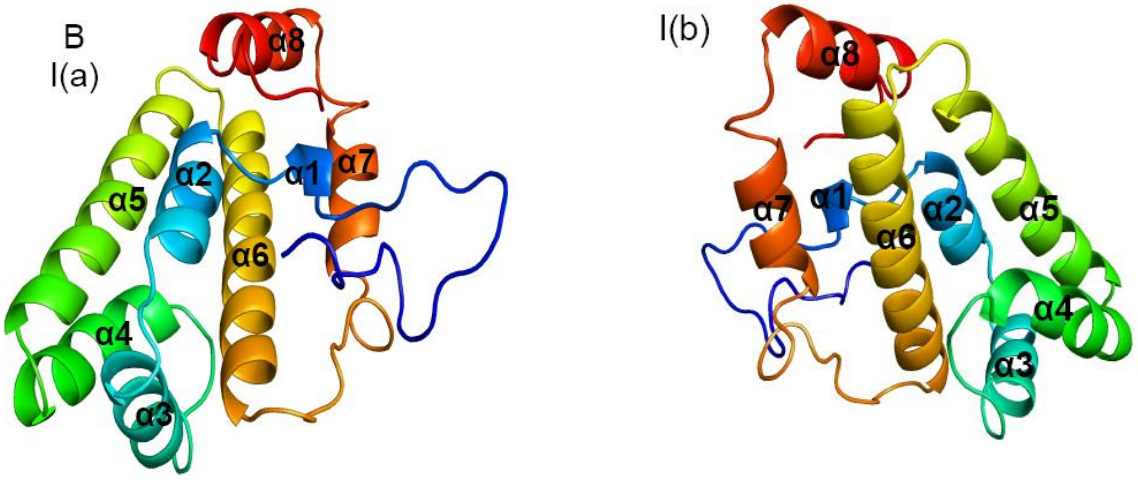


Figure 9: (A) Multiple sequence alignment between representative of the FAR protein from C. elegans group A and C (Ce-FAR-1 and Ce-FAR-7). Red marks refer to those amino acids that show sequence identity and blue mark refer to amino acids that show sequence similarity. Alpha helices represent the secondary structure of Ce-FAR-1 and its comparison with Ce-FAR-7 shown in red colour.

The high helical contents of the FAR family was confirmed by the first crystal structure of Ce-FAR-7 (PDB: 2w9y), showing nine helices and no beta structures in the protein (Jordanova *et al.*, 2009). The single polypeptide chain of Ce-FAR-1 crystal structure contains nine helixes. The structure is cantered on two long amphipathic helices, α5 and α6 which do not direct contact with each other but are inclined to each other by about 20 degrees. Due to low number of identity (27%), α9 helix and some residues of α1 helix was missing in the 3D model of Ce-FAR-1 (Fig. 9B, Ia, Ib). Ramachandran plot of Ce-FAR-1 fulfilled the test from 89.6% in most favoured region, 8.1% additional allowed region, 0.7% generously allowed region. There were two (1.5%) amino acids residues (Ala 16 & Glu9) present in the disallowed region of Ramachandran plot (Fig. 9B, III). So the results deduced that 3D structure of Ce-FAR-1 was not favored for good satisfactory model.



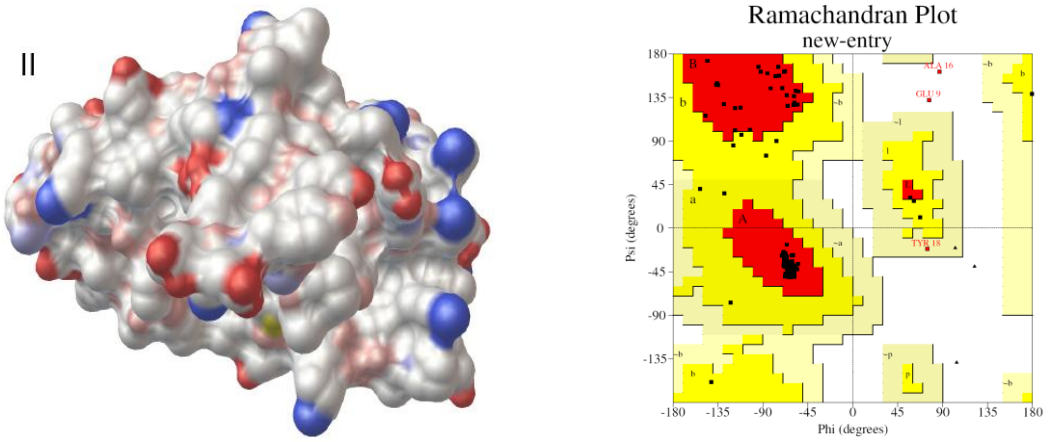


Figure 9: (B) (Ia, Ib) The alpha helices are labelled in two cartoon representation rotated relative to each other by 90° about the vertical axis. Blue colour shows the N-terminus region and red colour shows the C-terminus region of the protein. (II) The van der waal Molecular surface are coloured as follow- Carbon: white, Oxygen: red, Nitrogen: blue, Sulphur: yellow. (III) Favoured amino acids residues in red area in Ramachandran plot for 3D structure of Ce-FAR-1.

The structural orgination of Ce-FAR-1 (Fig. 9C, I) results in two binding pockets (P1 and P2) joined by cleft, which would allow to accommodate a veriety of ligands with different lengths of aliphatic chain. It was previously proved by determining the structure of Ce-FAR-7 (Jordanova et al., 2009). The cleft is capped by the helix α1 and the loop (L1 and L2) which joins them. The binding pocket P1 is formed 12 residues, that includes Pro3, Val4, Pro5, Pro8, Ile104, Gly111, Pro114, which are hydrophobic and Gln107, Tyr108, Glu112, Lys113, Ser115, which are hydrophilic in nature. The P1 binding pocket resisdues are covered by helices α1 and α6 and the loop (L1 &L7) (fig. 9C, IIa). The binding pocket P2 is formed 15 residues more than the resisdues of binding pocket P1, that includes Leu21, Phe93, Ala123, Ile127, Leu148, Gly150, which are hydrophobic and Glu17, Tyr18, Ser20, Glu100, Thr122, Lys124, His126, Tyr130, Thr149, which are hydrophilic in nature. The P2 binding pocket resisdues are covered by helices α1, α6 and α7 and the loop (L2 &L9) (fig. 9C, IIb).

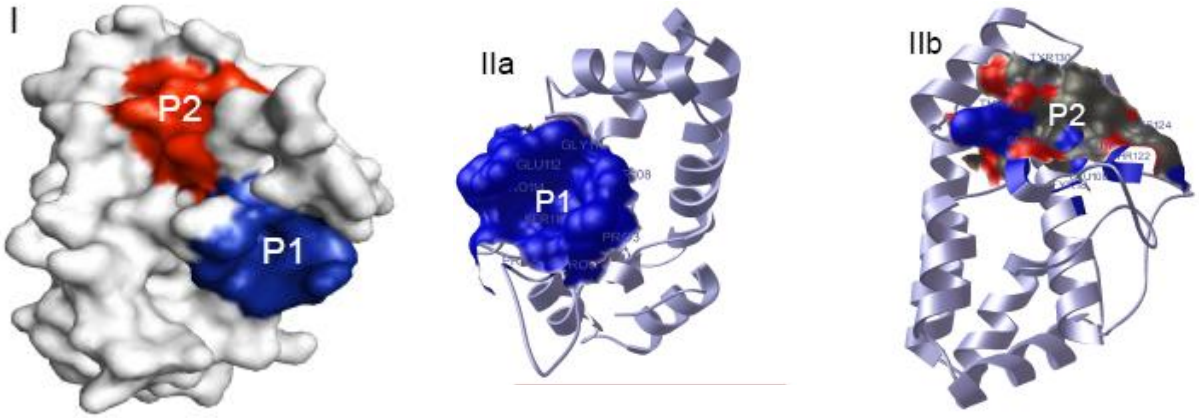


Figure 9: (C) (I) the van der waal Molecular surface illustrate the binding pockets (P1 and P2) of Ce-FAR-1. (IIa) 3D structure of Ce-FAR-1 shows the active site/pocket P1 include Pro3, Val4, Pro5, Pro8, Ile104, Gln107, Tyr108, Gly111, Glu112, Lys113, Pro114, and Ser115. (IIb) 3D structure of Ce-FAR-1 shows the active site/pocket P2 include Glu17, Tyr18, Ser20, Leu21, Phe93, Glu100, Thr122, Ala123, Lys124, His126, Ilu127, Tyr130, Leu148, Thr149 and Gly150.

The sequence alignment and the structural analysis would suggest that the structure of Ce-FAR-1 is respresntative of the complete family of FAR protein. Recently the first high resolution crystal structure of Ce-FAR-7 from *C. elegans* has been reported (Jordanova et al., 2009). Previous structural analysis of this family has been limited to SAXS study (Solovyova et al., 2003) and although the ovall molecular dimensions and the shape are consistent with our work the assumption, of a structural similarity of the FAR proteins with the ligand binding domain of the retinoic acid receptor (RXR α) and the nematode polyprotein allergens (Meenan et al., 2005) is incorrect.

3.3. Ce-FAR-6 T50D and Ce-FAR-7 T26A mutant

3.3.1. Mutation, Sequencing and Test expression

To mimic phosphorylation threonine (T) at position 50 from the conserved casein kinase II site of Ce-FAR-6 was replaced with aspartic acid (D). The mutated primers successfully replaced the nucleotides ACT with GAT in the vector pETM-11 containing the Ce-FAR-6 gene. On other hand the CK2 site was blocked in another Ce-FAR protein, Ce-FAR-7, by replacing the threonine (T) at position 26 with alanine (A). The mutated primers successfully replaced the nucleotide ACC with GCC in the vector pETM-11 LIC containing the Ce-FAR-7 gene.

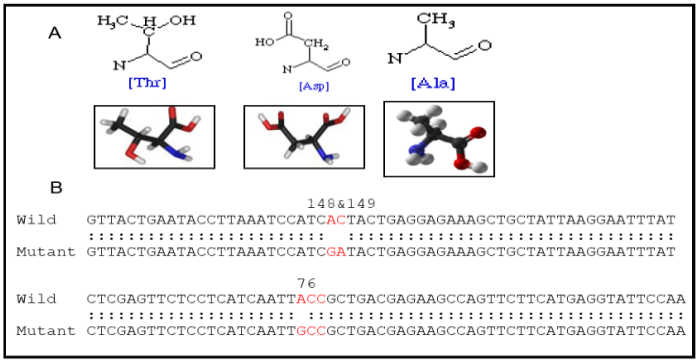


Figure 10: (A) Figure shows the chemical formula of Threanine (T), Aspartic acid (D) and Alanine (A). (B) Alignment of wild and mutant type sequence of Ce-FAR-6 shows mutation on position number 149 and150 that was successfully replaces nucleotide ACT with GAT. Alignment of wild and mutant type sequence of Ce-FAR-7 shows mutation on position number 76 that was successfully replaces nucleotide ACC with GCC.

The mutant (non-methylated) DNA was obtained and uniform growth of the kanamycin resistant plasmid showed that it was successfully transformed into the *E. coli* XL 1 blue competent cells. Sequencing confirmed that one clone of Ce-FAR-6 T50D and three clones of Ce-FAR-7 T26A contained the desired mutation. The web based bioinformatics tool shows the mutation in Ce-FAR-6 T50D and Ce-FAR-7 T26A mutant compared to the wild type (Sequences align, 2009) as shown in Fig. 10B. The molecular weight, iso-electric point and extinction coefficient of Ce-FAR-6 (wild type), Ce-FAR-7 (wild type) and Ce-FAR6 T50D and Ce-FAR-6 T26A mutant have been calculated theoretically and shown in Table 4 (Expasy protparam tool, 2009).

Table 5: Molecular weight, extinction coefficient and iso-eletric point of Ce-FAR-6 T50D and Ce-FAR-7 T26A mutant and their comparison with the wild type Ce-FAR-6 and Ce-FAR-7.

Assigned Name	Molecular weight	Iso-electric point	0.1%Extinction coefficient
Ce-FAR-6 (wild type)	22031Da	6.71	0.541
Ce-FAR-6 T50D mutant	22045Da	6.45	0.541
Ce-FAR-7	18502Da	6.24	0.255
Ce-FAR-7 T26A mutant	18471Da	6.24	0.255

To test the expression, a positive clone of the Ce-FAR-6 T50D mutant was transformed into three expression strains of *E. coli* competent cells (BL21(DE3)pLysS, BL21(DE3)RI, BL21(DE3)RP) and a positive clone of the Ce-FAR-7 T26A mutant was transformed into expression strains of *E. coli* competent cells (BL21(DE3)pLysS). OD_{600} was measured every half hour until it reached 0.6-0.8 and then the cells were induced to initiate expression.

After reaching the desired OD_{600} (cells grown at 37°C) the cells were induced with 1mM IPTG to initiate recombinant protein expression and then grown overnight at 20^0C. To analyse the results and choose the best expression conditions, samples were taken at different time intervals and run on SDS-PAGE as shown in Fig. 11. The results indicate that Ce-FAR-6 T50D mutants were successfully expressed in Bl21(DE3)RIL, weakly expressed in Bl21(DE3)RP with no expression in Bl21(DE3)plysS. On the other hand, Ce-FAR-7 T26A mutants were successfully expressed in Bl21(DE3)plysS (Fig. 11).

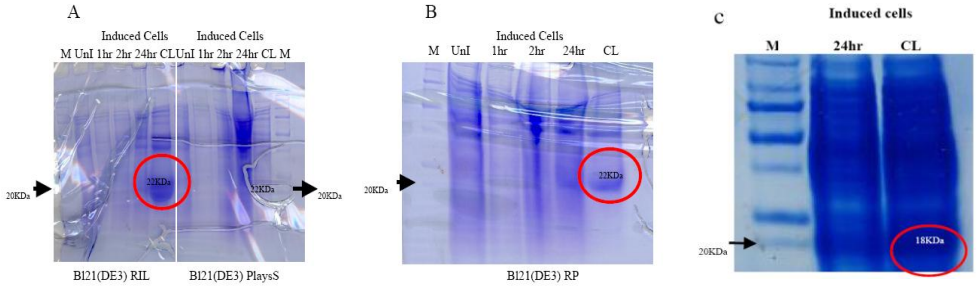


Figure 11: Expression pattern of Ce-FAR-6 T50D and FAR-7 T26A mutant (A) The gels show the expression of His-tagged Ce-FAR-6 T50D mutant protein in Bl21(DE3)RIL, Bl21(DE3)RP and Bl21(DE3)pLysS cells (B) The gel shows the expression of His-tagged Ce-FAR-6 (T50D) mutant protein in Bl21 (DE3) RP Competent cell. (C) The gel shows successful expression of His-tagged Ce-FAR-7 T26A protein in Bl21(DE3)pLysS cells (red mark circle). Uninduced cell: UnI, Induced cell: 1hr, 2hr, 24hr, Cell lyses: CL.

3.3.2. Expression and Purification

Bl21(DE3)RIL competent cells were used for large scale expression of Ce-FAR-6 T50D and Bl21(DE3)pLysS for Ce-FAR-7 T26A. Cell growth before induction is shown on Fig. 12. The recombinat proteins were successfully expressed and the yields were 24mg and 18mg respectively, per litre Culture.

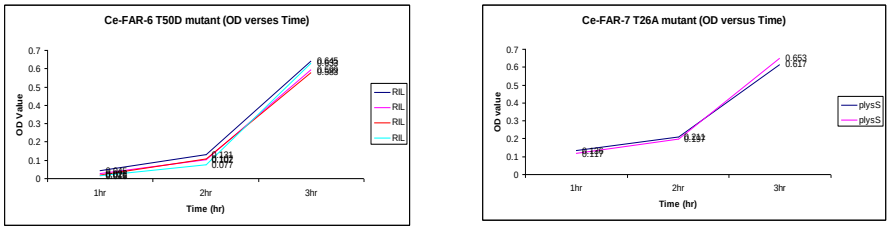


Figure 12: The graphs represent expression cells growth before induction, where OD_{600} is plotted against time (h). A. Ce-FAR-6 T50D; B. Ce-FAR-7 T26A.

Ce-FAR-6 T50D and Ce-FAR-7 T26A mutant were purified by Ni affinity chromatography, using the batch method as described in Materials and Methods section. Mutant proteins were eluted with high imidazole and concentrated by ultra filtration to 10mg/ml, 4mg/ml and 8mg/ml. The purity of the eluted protein fraction was analysis by SDS-PAGE as shown in Fig 13.

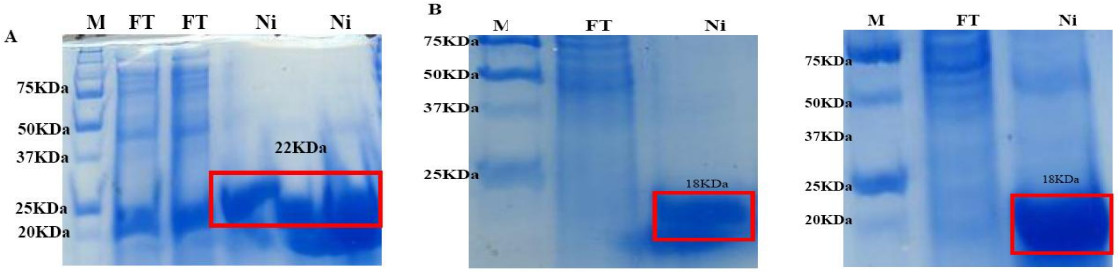


Figure 13: (A) SDS-PAGE showing the profile of affinity chromatography purification of Ce-FAR-6 T50D mutant protein. The size of the product was 22KDa compared with marker (M) and FT is the Flow through from the affinity column. (B) SDS-PAGE gel showing the profile of affinity chromatography of Ce-FAR-7 T26A mutant protein. The size of the product was 18KDa compared with marker (M) and FT is the Flow through from the affinity column.

3.3.3. Lipid removal, Dialysis and Phosphorylation

All FAR proteins have a conserved casein kinase II phosphorylation site as shown in Fig. 2. Jordanova et al (2009) showed that *in vitro* phosphorylation with casein kinase II results in a 81Da difference between phosphorylated and non-phosphorylated sample, confirming that Ce-FAR-7 wild type could be phosphorylated. The phosphorylated product was not stable and a mutant, mimicking phosphorylation was designed – Ce-FAR-7 T26D and used in comparative ligand binding experiments. The result showed that Ce-FAR-7 T26D bound retinol with higher affinity then Ce-FAR-7 wild type (Jordanova et al, 2009). The affinity for fatty acids however was not significantly different to the wild type. So it was proposed that the retinol binding site/pocket is most likely regulated by casein kinase II *in vivo* as the pocket is very close to phosphorylation site (Jordanova *et al.,* 2009). Two CKII genes exist in *C. elegans* corresponding to the two subunits of CKII (www.wormbase.org).

Based on the described above experiments (Jordanova et al, 2009) we initiated the study on Ce-FAR-6 CK II site by designing, cloning and purifying Ce-FAR-6 T50D mutant (as described elsewhere). The final goal is this mutant to be used in comparative lipid binding experiments as already shown for Ce-FAR-7 (Jordanova et al, 2009). To remove bound lipids the purified protein was incubated with hydrophobic matrix Lipidex-1000, followed by dialysis. The protein was stable and it was stored at -70°C for future experiments.

As described above, phosphorylation of Ce-FAR-7 *in vitro* with casein kinase II showed that Ce-FAR-7 has been phosphorylated (Jordanova *et al.*, 2009). In order to prove that the protein is phosphorylated, a mutant (T26A) with blocked phosphorylation site was designed, expressed and purified (as described elsewhere). After dialysis the protein was *in vitro* phosphorylated and prepared for mass spectral analysis.

4 Conclusions

The nematode FAR proteins may have an important effect on the efficacy of some classes of antihelminthic drugs. The 19 kDa retinol-binding protein of O. volvulus (presumably Ov-FAR-1), for example, has been demonstrated to bind ivermectin (Lal, 1996) which is the principle drug used to control onchocerciasis in West Africa. The protein could therefore play a role in the delivery of the drug to, or within the parasite, or alternatively it could diminish its effect or possibly even be involved in the evolution of drug resistance. It is even conceivable that, with their highly novel structures, the FAR proteins themselves may constitute a target for the design of new antihelminthic drugs.

5 Acknowledgements

I pay special thanks to Dr. Paul A. Tucker (Group Leader) European Molecular Biology Laboratory Hamburg Germany for facilitating me to conduct the study and giving pragmatic suggestions for making the manuscript. My especial thanks go to Dr. Rositsa Jordanova, European Molecular Biology Laboratory Hamburg Germany for helping me to solve the various practical problems.

6 References

Bradley JE, Helm R, Lahaise M and Maizels RM (1991) cDNA clones of Onchocerca volvulus low molecular weight antigens provide immunologically specific diagnostic probes. Mol Biochem Parasitol 46:219–228

Bradley JE, Nirmalan NJ, Klager SL, Faulkner H and Kennedy MW (2001) River blindness: a role for parasite retinoid-binding proteins in the generation of pathology? Trends Parasitol 17:471–475.

Burkhard P, Stetfeld J and Strelkov SV (2001) Coiled-coils: a highly versatile protein folding motif. Trends Cell Biol 11:82–88

Blaxter M (1998) Caenorhabditis elegans is a nematode. Science 282:2041–2103.

Cowan SW, Newcomer ME and Jones TA (1993) Crystallographic studies on a family of cellular lipophilic transport proteins. Refinement of P2 myelin protein and the structure determination and refinement of cellular retinol-binding protein in complex with all-trans-retinol. J Mol Biol 230:1225–1246

Couthier, A., Smith, J., Mcgarr, P., Craig, B. and Gilleard, J. S. (2004). Ectopic expression of a Haemonchus contortus GATA transcription factor in Caenorhabditis elegans reveals conserved function in spite of extensive sequence divergence. Molecular and Biochemical Parasitology 133, 241–253.

E, Kostova (2008) Expression, purification, Crystallization and some functional studies of new representative of nematode Fatty acid and Retinol binding proteins FAR. European Molecular Biology Laboratory Hamburg Germany.

Flower DR (1996) the lipocalin protein family: structure and function. Biochem J 318:1–14

Gudas, L. J., Sporn, M. B., and Roberts, A. B. (1994) in The Retinoids (Sporn, M. B., Roberts, A. B., and Goodman, D. S., eds) pp. 443–520, Raven Press, New York

Gudas LJ, Sporn MB and Roberts A (1994) Celluar biology and biochemistry of retinoids. Raven, New York, pp 443–520.

Garofalo A, Kla¨ ger SL, Rowlinson MC, Nirmalan N, Klion A, Allen JE, Kennedy MW and Bradley JE (2002) The FAR proteins of filarial nematodes: secretion, glycosylation and lipid binding characteristics. Mol Biochem Parasitol 122:161–170

Garofalo A, Rowlinson MC, Amambua NA, Hughes JM, Kelly SM, Price NC, Cooper A, Watson DG, Kennedy MW, Bradley JE. The FAR protein family of the nematode Caenorhabditis elegans. Differential lipid binding properties, structural characteristics, and developmental regulation. J Biol Chem. 2003 Mar 7;278(10):8065-74. Epub 2002 Dec 26

Garofalo A, M. W. Kennedy and Janette E. B (2002) The FAR proteins of parasitic nematodes: their possible involvement in the pathogenesis of infection and the use of Caenorhabditis elegans as a model system to evaluate their function Med Microbiol Immunol 192: 47–52

Jinsong Xuan a, Hongwei Yao, Yingang Feng and Jinfeng Wang (2009) Cloning, expression and purification of DNA-binding protein Mvo10b from Methanococcus voltae, Protein Expr Purif. Apr; 64(2):162-6. Epub 2008 Nov 17.

J. S. Gilleard (2004) the use of Caenorhabditis elegans in parasitic nematode research Parasitology 128.

Jordanova R, Matthew R. Groves, Elena Kostova, Christian Woltersdorf, Eva Liebau, Paul A. Tucker. (2009) Fatty acid and retinoid binding proteins have distinct binding pockets for the two types of cargo. JBC.

Kennedy, M. W. (2001) In Parasitic Nematodes: Molecular Biology, Biochemistry and Immunology CABI Publishing, Wallingford and New York pp. 309–330.

Kuang L, Colgrave ML, Bagnall NH, Knox MR, Qian M and Wijffels G (2009) The complexity of the secreted NPA and FAR lipid-binding protein families of Haemonchus contortus revealed by an iterative proteomics-bioinformatics approach. Mol Biochem Parasitol.

Kennedy MW, Garside LH, Goodrick LE, McDermott L, Brass A, Price NC, Kelly SM, Cooper A and Bradley JE (1997) The Ov20 Protein of the Parasitic Nematode Onchocerca volvulus. J Biol Chem 272:29442–29448.

Kennedy MW (2000) the polyprotein lipid binding proteins of nematodes. Biochim Biophys Acta 1476:149–164

Kennedy MW, Allen JE, Wright AS, McCruden AB and Cooper A (1995) the gp15/400 polyprotein antigen of Brugia malayi binds fatty acids and retinoids. Mol Biochem Parasitol 71:41–50

Liu LX, Buhlmann JE and Weller PF (1992) Release of prostaglandin-E2 by microfilarae of Wucheria bancrofti and Brugia malayi. Am J Trop Med Hyg 46:520–523.

Lal PG, James ER (1996) Onchocerca retinol-and-ivermectinbinding protein. Parasitology 112:221–225

Mei, B., Kennedy, M. W., Beauchamp, J., and Komuniecki R. (1997) Secretion of a novel, developmentally regulated fatty acid-binding protein into the perivitelline fluid of the parasitic nematode, Ascaris suum. J. Biol. Chem. 272, 9933–9941.

Meenan NA, Cooper A, Kennedy MW, Smith BO. (2005) Resonance assignment of ABA-1A, from Ascaris suum nematode polyprotein allergen. J Biomol NMR 32(2):176.

Noy N (2000) Retinoid-binding proteins: mediators of retinoid action. Biochem J 348:481–495.

Nirmalan NJ (1999) A comparative analysis of novel filarial retinol binding proteins. Thesis, University of Salford.

Nikawa, T., Odahara, K., Koizumi, H., Kido, Y., Teshima, S., Rokutan, K., and Kishi, K. (1999) J. Nutr. 129, 934–941

Prior A, Jones JT, Blok VC, Beauchamp J, Mcdermott L, Cooper A and Kennedy MW (2001) A surface-associated retinol and fatty acid binding protein (Gp-FAR-1) from the potato cyst nematode Globodera pallida: lipid binding activities, structural analysis and expression pattern. Biochem J 356:387– 394

Prior, A., Jones, J. T., Blok, V. C., Beauchamp, J., McDermott, L., Cooper, A., and Kennedy, M. W. (2001) Biochem. J. 356, 387–394

Surette MG, Levit M, Liu Y, Lukat G, Ninfa EG, Ninfa A and Stock JB (1996) Dimerization is required for the activity of the protein histidine kinase CheA that mediates signal transduction in bacterial chemotaxis. J Biol Chem 271:939–945

Szilak L, Moitra J, Krylov D and Vinson C (1997) Phosphorylation destabilizes alpha-helices. Nat Struct Biol 4:112–114

Sturchler D, Wyss F, Hanck A (1981) Retinol, onchocerciasis and Onchocerca volvulus. Trans R Soc Trop Med Hyg 75:617– 625

Sani BP, Comley JCW (1985) Role of retinoids and their binding proteins in filarial parasites and host tissues. Trop Med Parasitol 36:20–23

Sani BP, Vaid A (1988) Specific interaction of ivermectin with retinol-binding proteins from filarial nematodes. Biochem J 249:929–932.

Solovyova AS, Meenan N, McDermott L, Garofalo A, Bradley JE, Kennedy MW, Byron O. (2003) the polyprotein and FAR lipid binding proteins of nematodes: shape and monomer/dimer states in ligand-free and bound forms. Eur Biophys J. ;32(5):465-76.

Tree TIM, Gillespie AJ, Shepley KJ, Blaxter ML, Tuan RS and Bradley JE (1995) Characterisation of an immunodominant glycoprotein antigen of Onchocerca volvulus with homologues in other filarial nematodes and Caenorhabditis elegans. Mol BiochemParasitol 69:185–195.

Ventura C and Maioli M (2001) Protein kinase C control of gene expression. Crit Rev Eukaryot Gene Expr 11:243–267.

Vogel S, Gamble MV and Blaner WS (1999) Biosynthesis, absorption, metabolism and transport of retinoids. In: Nau H, Blaner WS (eds) The biochemical and molecular basis of vitamin A and retinoid action. Springer, Berlin Heidelberg New York, pp 31–95.

7 Appendixes

Appendix I

Sequence of Ce-FAR-1 (549 Nucleotide)

```
ATGATCCGTGCCACTATCATCCTCGCCGCTGTCGCTGCCCTCGCCTTCTCTGCTCCAGTCCCGGAGGTCCCAGAGAA
CTATGACGATATCCCAGCTGAATACAAGTCCCTTATCCCAGCCGAGGTCTCCGAGCATCTTAAGTCCATCACCCCAG
AAGAGAAGGCTATTCTTAAGGAGGTCGCCAAGGGATACAAGGACTTCAAGAGCGAGGATGATTTCTTGAACGCCCTC
AAGGAAAAGTCTCCAACTCTCCACGAGAAGGCCTCCAAGCTCCACCAAATCGTCAAGGACAAGGTCAACGCTCTCAA
TGATGAGGCTAAGGCTTTCGTCAAGAAGGCTATCGCTGAAGGACGTAAGATCCACGCTCAATACTTGGCCGGAGAGA
AGCCATCCCTCGACACCCTCAAGACCACCGCCAAGACCCACATCGAAGCCTACAAGGGACTTTCCCAAGATGCCAAG
GACTCCATCGCCAAGGAATTCCCAATCCTCACCGGATTCTTCAAGAACGAGAAGGTTCAAGCCATGGTCGGACAATA
CCTCAACTAA
```

Amino acid sequence of Ce-FAR-1 (182 AA)

MIRATIILAAVAALAFSAPVPEVPENYDDIPAEYKSLIPAEVSEHLKSITPEEKAILKEVAKGYKDFKSEDDFLNAL
KEKSPTLHEKASKLHQIVKDKVNALNDEAKAFVKKAIAEGRKIHAQYLAGEKPSLDTLKTTAKTHIEAYKGLSQDAK
DSIAKEFPILTGFFKNEKVQAMVGQYLN*

Sequence of Ce-FAR-2 (549 Nucleotide)

ATGATCCGCGCCTTCCTCGTCGTAGCCCTTGCCTCCGTGGCTGTCTTCTCTGCCCCAATCCCAGAGGTTCCACAGAA
CTTCGACGACATCCCAGCCGAGTACAAGGGACTCATCCCAGCTGAAGTTGCCGAGCACCTCAAGGCCATCACCGCTG
AGGAGAAGGCTGCTCTTAAGGAACTCGCCCAAAACCACAAGGAATACAAGACCGAGGAAGAATTCAAGGCTGCTCTC
AAGGAAAAGTCTCCATCCCTTTACGAGAAGGCTGGAAAGCTCGAGGCTCTCCTCACCGCTAAATTCGAGAAACTCGA
TGCCACCGCTCAGGCTCTTGTCAAGAAGATCATCGCCAAGGGACGTGAACTCCACCAACAATACCTCGCCGGAGATA
AGCCAACTCTTGATTCTCTTAAGGAACTCGCCAAGGGATACATCGCTGAATACAAGGCTCTTTCTGATGACGCCAAG
GCTACCATCACCGCTGAGTTCCCAATCCTCACTGGATTCTTCCAAAACGAAAAGATTCAAGCCATCGTCGGACAATA
CGTCAACTAA

Amino acid sequence of Ce-FAR-2 (182 AA)

MIRAFLVVALASVAVFSAPIPEVPQNFDDIPAEYKGLIPAEVAEHLKAITAEEKAALKELAQNHKEYKTEEEFKAAL
KEKSPSLYEKAGKLEALLTAKFEKLDATAQALVKKIIAKGRELHQQYLAGDKPTLDSLKELAKGYIAEYKALSDDAK
ATITAEFPILTGFFQNEKIQAIVGQYVN*

Sequence of Ce-FAR-3 (570 Nucleotide)

ATGTCTCGTCTTTTCGCTTTCAACGTTTTCTGCTTGGTTCTTCTCCGTTTCTCAGCTGCTGCTCCAGCTGATGATTC
TTCTCCATTCTCTCAAATTTTGAAGCAACACAAAGATCTCCTTCCATCCGAAGTTGTTCAAGCCTATCAGGATTTGT
CTCCAGAAGAGAAGGCTGCATTGAAGGATGTATTCAAGAACTACAAGAGCTACAAAAACGAAGGAGAATTGATTGCT
GCTCTCAAAGAAAAGTCTTCAAGCTTGGGAGAAAAGGCTGAGAAGCTTCAAGCTAAGCTCCAAAAGAAGGTTGATGC
ATTGAGCCCAAAACCAAAGGATTTTGTTAACGAGCTCATTGCTGGAGGACGTGGTCTTTATGCTCGTTCTGTTAATG
GAGAGAAGATCTCAGTTTCCGAGATCAAGCTTCTCATTGAAACCCAAGTTGCTGCATACAAGGCATTGCCAGCTGAG
GCTCAAGACGAGTTGAAGAAAAATTTCGGAGGAGTCGCCAAGTTTTTGGAGGATGACAAGACTCAAACACTCATTGC
CAAGCTTCTTGAGAAGAACAATAACCAGTAA

Amino acid sequence of Ce-FAR-3 (189 AA)

MSRLFAFNVFCLVLLRFSAAAPADDSSPFSQILKQHKDLLPSEVVQAYQDLSPEEKAALKDVFKNYKSYKNEGELIA
ALKEKSSSLGEKAEKLQAKLQKKVDALSPKPKDFVNELIAGGRGLYARSVNGEKISVSEIKLLIETQVAAYKALPAE
AQDELKKNFGGVAKFLEDDKTQTLIAKLLEKNNNQ*

Sequence of Ce-FAR-4 (618 Nucleotide)

ATGAGTAAATTACTTCTACTTGTACTTTCTCTTCTCTTTTTTATTACATCAGCTTTTCCATTCGGAGAACCACAAGC
GGGAGGATTTCAAAAATTTAAGAACTTACTTCCAAGGGAGCTTGTGGAGGCCTACAGTAACCTGAGTCAAAAAGATC
AACCTGATCTGAAAGATGTATTCCGTAATCATCAAAACTACCGAAATGAACAAGAAATGGTTAATGCTTTGAAAATG
AAAAATCCAGCACTCGGAGCCCGAATGGAACGGAGATTGATGGCTTTGAAGCAGAAAATCGATGGATTGAGTAGCGA
AGAAGCGAAAGGATTTATTCAGAATTTGATTTCAACTGGAAGACAAATTTATGCTCAACGGCTTAATGGACAACAAA
TGGATCAGTCACAATTGAGACAAGTTGGAATGGGAATTGCAATGCACTATCGATCATTACCTCCATATGCACAACAG
GAACTTCAAAGTACTTTTCCACAAATCTTCCAATTCATGAGACAAATGCGTGAACAACGTCTTCGAAGCATGATGGG
CGGATTCTTTGGCGGCGGAGGTGGAATCGGAATGGGACAAGGCATGGGACAAGGCATGGGAATGGGTATGGGAAAAT
AA

Amino acid sequence of Ce-FAR-4 (205 AA)

MSKLLLLVLSLLFFITSAFPFGEPQAGGFQKFKNLLPRELVEAYSNLSQKDQPDLKDVFRNHQNYRNEQEMVNALKM
KNPALGARMERRLMALKQKIDGLSSEEAKGFIQNLISTGRQIYAQRLNGQQMDQSQLRQVGMGIAMHYRSLPPYAQQ
ELQSTFPQIFQFMRQMREQRLRSMMGGFFGGGGGIGMGQGMGQGMGMGMGK*

Sequence of Ce-FAR-5 (531 Nucleotide)

ATGTTACTCCGTTTCTTTGCTATTTTATTAATTTTTCATTTCTCATTTGCTGCTGATGGAATATTTGAAGCTGTTAT
TGAGAGCTATAAAGATGCTCTTCCACCAAAAGTTGTTGCGGCTTTTGATAATTTGAGTCCAGGGGAAAGTGCTATTA

TGAAAGAAGTATTCATGAATTATGATAAATTTACAAGTATTGCTGATCTGATTGTTGCAATCAAGAAGAAATCGGAA
TCACTTGGATCGTTTTTCGAGAAACTTTACATTGAAATTGACGCTGAAATTCAGGCATTAACACCGGAAACAAAGAA
ATTTGTAACTGAAATGTTAGGAATTGGCAGAGCAATTTATACTGCACAAATTGTTGGAATTCCATTAGATTCTAAAG
AAGTCCTACCTGTATTTGCAAAACAATTTACTTCGTTCAAATCACTATCGGATGCTACAAAGGAAGAACTTGAAAAA
ACATTCCTCGGTCTCTACAAATTCGCTTCAAACGACAAGATCAAGACTGAGATTGATAAATTGTTGTGA

Amino acid sequence of Ce-FAR-5 (176 AA)

MLLRFFAILLIFHFSFAADGIFEAVIESYKDALPPKVVAAFDNLSPGESAIMKEVFMNYDKFTSIADLIVAIKKKSE
SLGSFFEKLYIEIDAEIQALTPETKKFVTEMLGIGRAIYTAQIVGIPLDSKEVLPVFAKQFTSFKSLSDATKEELEK
TFLGLYKFASNDKIKTEIDKLL*

Sequence of Ce-FAR-6 (555 Nucleotide)

ATGATCCGCATCTTCCTTGTCATAGCACTTGCCTCCGTCGCTGTCTTTTCTGCTCCAATTTCACGCTTACCACAAAG
CTTAGACGACATCCCAGCTGAGTTCAAGGAACTCATCCCAGCTAAAGTTACTGAATACCTTAAATCCATCACTACTG
AGGAGAAAGCTGCTATTAAGGAATTTATCAAAAGTGTTATGGGAGGAAACAAATCTGTTGAAGAATTGAGTGCTGAT
ATCAAGGAGCGGTCCCCATCCCTGTACGCTAAGGTTGAAAAACTCGACGTCCTACTCCGCACAAAACTCGCGAAACT
CGACCCCGCCGCTCTGGCTTTATTCGGGAAAGTCATCGCTCAGGGACTTTCTTTCCGGCAACAGTTCCATGCCGGAT
ACCAGCCAACTCCTGAGATGGTAAAGAAACTCTTCAAGGGATACATTGCTGAATATAAAACACTTTCTGAAAACGCC
AAGGCCACCATCACCGATGAGTTCCCCATCGTCGTTGAATTCTTTCAACACGAGAAAATTCAAGCCATCATCCAACA
GATCGTGAACTACTAA

Amino acid sequence of Ce-FAR-6 (184 AA)

MIRIFLVIALASVAVFSAPISRLPQSLDDIPAEFKELIPAKVTEYLKSITTEEKAAIKEFIKSVMGGNKSVEELSAD
IKERSPSLYAKVEKLDVLLRTKLAKLDPAALALFGKVIAQGLSFRQQFHAGYQPTPEMVKKLFKGYIAEYKTLSENA
KATITDEFPIVVEFFQHEKIQAIIQQIVNY*

Sequence of Ce-FAR-7 (417 Nucleotide)

ATGAGCGTTGCTTCACTTCCAGAATGTGTCAAAAACTTTTTCCCAACTGAACAACTCGAGTTCTCCTCATCAATTAC
CGCTGACGAGAAGCCAGTTCTTCATGAGGTATTCCAAAAGCATTCATGTTTCTCACAATGTGGTGAAATGATTGACG
AGGTCTCGAAAAAGCATCCAGAATTGGGAAAACGGTTGGCAACTGTGTTGGAGGGGAACAAGAAACGTTTGGATGGT
TTGAGCCCGGCTGCTGTTGAGTATGCCAAGAAGCTCATACACATGGTAACCACCACCTTGTGCTCCTTAACCGTCGG
AAAACCAATTGATGATGCAGATGCAAAACGTCTTCACCAGGAATTCCAAAGCCTATCTTCAGAAGATCAGGCTGCGC
TGAGAAAGAATAATCCGGATATTAAATTTTGA

Amino acid sequence of Ce-FAR-7 (138 AA)

MSVASLPECVKNFFPTEQLEFSSSITADEKPVLHEVFQKHSCFSQCGEMIDEVSKKHPELGKRLATVLEGNKKRLDG
LSPAAVEYAKKLIHMVTTTLCSLTVGKPIDDADAKRLHQEFQSLSSEDQAALRKNNPDIKF*

Sequence of Ce-FAR-8 (699 Nucleotide)

ATGTTTACCTTACGGGCATGCTCCTTAATTCTTGTGTCAGTGGCTGTGTTCGTTTTGGGTCGCCCCGCTGAAGAACA
AATGACCGAGAAGGATTTTACAAACTTGGTGTTCACTGTGGAGAAGTTTGATCAAATTTTGAAGGCCTATTCTGAAT
ACAAACAATTCATGCCTTCTTACGTTCTGGAACCACTTGATAACATCACCGACGAACAGAAAACTCAGGCAGTTCAG
ATGGTGAATGACTATCACGCAGGCAAGTTTGAGCCGAAGAACTACGATGAGTACATCGCAATCATGAAGAAGAGCTA
TCCAGCATTGGCAGGTCCATATGAGACAATGTACAACAAGTACAAAGAACAGGTTGCCAAATTGGGACCAAAGGGAC
AAGAATATTGCAATGGGCTCGAAGCTCAAATGTACACCGATGCCTCCCCTGACCGTGTTGTCTGGGCTTGCCATATC
TTCAACAATGCCAAGTCCGCAGTTAGTGGAGCAAAAGCTCTTCTTCAGGATGATTCGGAAGCGGCAAAGATTGAAGA
AGCATTCCCAGAGGCTGTGAAACTTCTTAATAGCAAAAAATTTGAAGCCTACTCGATTATCGTTAACAACCTGAAAC
CATTGGATTGCGTCAAGGATCGTGAGCAAGTCTTCAATACTATCAAATTGATGGATAAGCAAAGTGTGCTCACCAAT
AATTGA

Amino acid sequence of Ce-FAR-8 (232 AA)

MFTLRACSLILVSVAVFVLGRPAEEQMTEKDFTNLVFTVEKFDQILKAYSEYKQFMPSYVLEPLDNITDEQKTQAVQ
MVNDYHAGKFEPKNYDEYIAIMKKSYPALAGPYETMYNKYKEQVAKLGPKGQEYCNGLEAQMYTDASPDRVVWACHI
FNNAKSAVSGAKALLQDDSEAAKIEEAFPEAVKLLNSKKFEAYSIIVNNLKPLDCVKDREQVFNTIKLMDKQSVLTN
N*

Appendix II