NUCLEOTIDE SEQUENCE OF A cDNA FOR *crf-1*,

A STEROID HORMONE RECEPTOR HOMOLOGUE GENE IN

CAENORHABDITIS ELEGANS

by

Rebecca M. Eagen

Hon.B.Sc., University of Western Ontario, 1988

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Biological Sciences

Rebecca M. Eagen 1991

SIMON FRASER UNIVERSITY

October 1991

APPROVAL

Name: **REBECCA EAGEN**

Degree: **Master of Science**

Title of Thesis:

NUCLEOTIDE SEQUENCE OF AcDNA FOR crf-1, A STEROID HORMONE
RECEPTOR HOMOLOGUE IN CAENORHABDITIS ELEGANS

Examining Committee:

Chairman: Dr. A.T. Beckenbach

Dr. B. Honda, Associate Professor,
Senior Supervisor, Dept. Biological Sciences
SFU

Dr. D. Baillie, Professor, Dept.
Biological Sciences, SFU

Dr. M. Smith, Professor, Dept.
Biological Sciences, SFU

Dr. B. Brandhorst, Professor,
Director, IMBB, SFU,
Public Examiner

Date Approved _____Dec 6 /91_____

Title of Thesis/Project/Extended Essay

NUCLEOTIDE SEQUENCE OF A cDNA FOR

crf-1, A STEROID HORMONE RECEPTOR

HOMOLOGUE GENE IN CAENORHABDITIS

ELEGANS

Author:

(signature)

REBECCA EAGEN

(name)

Dec 6/91

(date)

ABSTRACT

The purpose of this work was to obtain sequence of a cDNA for a putative steroid hormone receptor homologue in *Caenorhabditis elegans* and to elucidate its pattern of expression during development.

In previous work, two oligonucleotides that share identity with the DNA binding domain of a variety of human steroid hormone receptors had been used to isolate genomic and cDNA clones of a *C. elegans* steroid receptor homologue, *crf-1* (*C. elegans* receptor finger). Nested deletions of a 2.2 kb cDNA had also been created to allow DNA sequencing.

The nucleotide sequence of the cDNA was determined and a putative protein of 440 amino acids and 50 147 Daltons was derived. A portion of this protein sequence shares identity with the DNA binding domain of other steroid hormone receptors and transcription factors. There are sites for phosphorylation, recognition of a hormone response element and intron splice junctions within the cDNA and protein product of *crf-1* that are characteristic of steroid hormone receptors. The putative ligand binding region shows no apparent homology to any previously reported genes.

Northern blots of stage specific RNA from *C. elegans* were probed with the *crf-1* cDNA to reveal a possible 2.2 kb message expressed during all stages.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Materials and Methods

I.   DNA Sequencing

II. Organization and Expression of *crf-1*

Results

I. Sequence Analysis of *crf-1*

LIST OF FIGURES

INTRODUCTION

Steroid hormone receptors are a class of regulatory proteins that are activated by the binding of a hormone. The binding of the ligand activates the DNA binding domain (Gehring, 1987). This area then interacts with *cis* regulatory elements of specific genes (Yamamoto, 1985). This type of interaction is one mechanism of regulating gene expression. In eukaryotes, transcriptional regulation is thought to be the primary mechanism of control (Chambon *et al*, 1984). Therefore, steroid hormone receptors are a unique group of proteins, the study of which allows one to elucidate details concerning very fundamental elements of the control of growth and differentiation. Steroids are derived from cholesterol and make ideal intercellular messages because they are small, stable and lipophilic (Yamamoto, 1985). These molecules are important for mediating a wide variety of physiological responses from stress to differentiation.

The purpose of this study was to characterize a previously isolated steroid receptor homologue in *Caenorhabditis elegans*. It has been shown that in Drosophila, a steroid hormone receptor *svp* is involved in photoreceptor differentiation during neuronal development (Ritchie *et al*, 1990). Consequently if the paradigm of hormone messenger-activated receptor-gene control is a universal system then even if a different class of message is involved the mechanism of interaction should be more or less standard. All eukaryotes including plants and invertebrates as well as some prokaryotes have the enzymatic components capable of synthesizing and metabolizing steroids (Yamamoto, 1985).

Why choose the organism *C. elegans*? It is an invertebrate but hormonal function is not as well characterized as in the insects. *C. elegans* is a well suited

organism since it has been extensively characterized in terms of genetics, anatomy, development and behavior (Emmons, 1988 and Kenyon, 1988). In addition there exists an ordered cosmid library which will eventually allow researchers to map the genome and the ability to create transgenic animals (Wood, 1988). However, much is still unknown about gene regulation in this organism.

Steroid hormone receptors, although they can range in size from 400-1000 amino acids in length, can be strictly divided into four functional domains. There is a hypervariable region with respect to size and composition, located at the amino terminus. The ultimate function is unknown; however, deletions within this region can diminish but not abolish receptor activity (Evans, 1988). The DNA binding domain, also called region 'C', lies immediately to the carboxy side of the hypervariable region (see Figure 1). This is a highly conserved region of approximately seventy amino acid residues. There are eight conserved cysteine residues which complex with zinc ions to form two fingers. This region may also interact with associated proteins and play a role in nuclear translocation (Beato, 1989). Immediately following this region, less well conserved and of unknown function is the 'D' region (Green and Chambon, 1988). The carboxy terminus contains the 'E' region which is responsible for ligand binding and also possibly involved in transactivation, nuclear translocation and dimerization (Walhi and Martinez, 1991). Binding of the steroid hormone by the receptor in the cytoplasm of the cell causes an allosteric change in the receptor structure. The actual events involved in this process are unknown at present but may involve the dissociation of receptor associated components (hsp90) (Green and Chambon, 1988). The DNA binding site becomes active and the receptor moves into the nucleus to regulate transcription by associating with hormone response elements flanking steroid

Figure 1 A schematic diagram of the general peptide structure of a steroid hormone receptor. The peptide domains A-E commonly share the characteristics listed below each domain.

H2N

A/B	C	D	E	COOH

HYPERVARIABLE

DNA BINDING
RECEPTOR ASSOCIATED
PROTEIN BINDING
DIMERIZATION?

HORMONE BINDING
TRANSACTIVATION
DIMERIZATION
NUCLEAR TRANSLOCATION
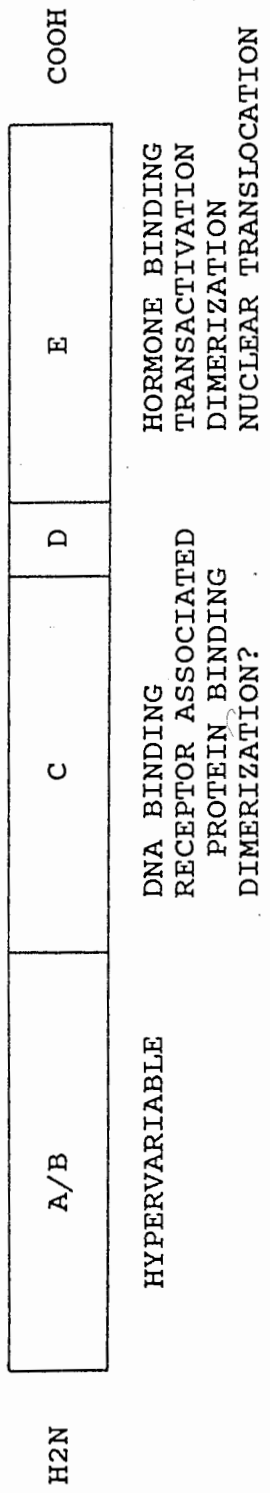
sensitive genes (Gehring, 1987).

Phosphorylation of steroid hormone receptors may play a role in ligand binding (Beato, 1987) but also may be important in mediating the biological actions of steroid hormones (Moudgil, 1990). For progesterone receptors the kinases involved phosphorylate mostly serine residues and the sites of phosphorylation lie within the amino terminus of the receptor. These multiple sites may suggest that multiple kinases are involved. For the estrogen receptor, the sites are found in the carboxy terminus and tyrosine appears to act as the major substrate of the kinase. The role of phosphorylation in steroid hormone receptor function could be diverse. This post translational modification could modulate ligand binding, it could be critical for DNA binding to the hormone response elements and it could also be involved in receptor regulation (Moudgil, 1990). The receptor itself could autophosphorylate or act as a kinase, these two possibilities have not been eliminated. It is however known that under physiological conditions, hormone treatment leads to amplification of phosphorylation and a coincident change in the expression of hormone sensitive genes.

The DNA binding fingers (see Figure 2) formed by the tetrahedral association of the sulphur atoms of two pairs of highly conserved cysteine residues with zinc ions are thought to be responsible for interacting with the DNA and consequently altering gene expression (Beato, 1991). The DNA sequence that these fingers interact with have been called the hormone response element. This *cis* acting element has 'enhancer-like' properties in that it functions in a position and orientation independent fashion (Beato, 1987). Commonly, these hormone response elements are comprised of 13-15 base pairs which include two conserved blocks with their centers 10 base pairs apart separated by three nonconserved

nucleotides. They are palindromic with a central axis of dyad symmetry (Beato, 1991). There is a general consensus sequence for the element and depending upon the nucleotide substituted at two locations, three subgroups have been identified. The overall consensus sequence is $5'AGGN_xCAN_{0-5}TGN_yCCT\ 3'$ where the $N_x = T$ and the $N_y = A$ for the estrogens and $N_x = A$ and $N_y = T$ for the glucocorticoids and $N_x = T$ or A for the thyroid response elements (Danielsen *et al*, 1989; Glass and Holloway, 1990; Wahli and Martinez, 1991). The first zinc finger has been implicated as being responsible for the sequence specific interaction (Green and Chambon, 1988; Luisi *et al*, 1991). Precisely three amino acids (marked with asterisks in Figure 2), whose mutation has been shown to abolish this interaction, occur between the third and fourth cysteine and directly after it (Mader *et al*, 1989). These residues lie within what is thought to be an alpha-helical region (Walhi and Martinez, 1991). For the estrogen group the sequence of residues is CEGCKA/G and for the glucocorticoid group it is CGSCKV in which the second, third and fifth residues are the ones previously referred to. A receptor dimer binding to one response element is the current model of the mechanism of interaction (Green *et al*, 1988). The two zinc fingers of the glucocorticoid receptor form a globular domain. Then the amino terminus of the first finger with its tertiary structure of antiparallel beta-pleated sheet is responsible for orientating the residues that make contact with the phosphate backbone of DNA (Luisi *et al*, 1991). The carboxy terminus of the finger is a distorted alpha-helix of which the residue side chains form contacts with the bases in the major groove. The entire carboxy terminus of the first and second finger are involved in dimer contacts and positive transcriptional regulation. The actual DNA binding domain lacks the ability to dimerize and it is only after binding to the hormone response element that this event occurs (Luisi *et al*, 1991).

In previous work to isolate receptor homologues in *C. elegans*, synthetic oligonucleotides that share identity with the DNA binding domain of other steroid hormone receptors were used to screen both Charon 4 genomic (Snutch, 1984) and lambda ZAP cDNA (Barstead et al, 1989) *C. elegans* libraries. The oligonucleotides called SR1 and SR2 lie within the CI region of the 'C' domain (see Figure 14). This region has a higher degree of sequence conservation. SR1 is 23 nucleotides in length and has 32-fold degeneracy based on conservative codon usage and *C. elegans* codon bias (Emmons, 1988). SR2 is 20 nucleotides in length and 16-fold degenerate. Several positive clones were isolated from both libraries by A. Purac.

Previously A. Purac conducted a genomic library screen which produced a positive clone called 9aSR1. EcoRI digestion of this clone produced fragments of 9.0, 3.3, 0.95 and 0.8 kb. These were blotted by the method of Southern (Maniatis *et al*, 1982) and probed with the radiolabeled cDNA fragment. The 3.3 and 0.8 kb fragments hybridized to the probe and were subsequently subcloned into the EcoRI site of pUC19 (Yanisch-Perron *et al*, 1985) and called 9aSR1 3.3Eco and 9aSR1 0.8Eco.

From the cDNA library a 4.5 kb clone hybridized to the original SR1 oligonucleotide in a Southern blot (not shown). Subsequent restriction with EcoRI reduced this clone to fragments of 2.2, 1.6, 0.8 and 0.1 kb. Of these only the 2.2 kb fragment hybridized to the phage 9aSR1 recovered from the genomic library screen. This suggests that the larger cDNA was an artifact produced during library construction. A probable explanation could be that two messages became ligated after the restriction of added EcoRI linkers just prior to vector ligation. The Henikoff nested deletion protocol was used on the 2.2 kb subclone to produce overlapping clones suited to rapid sequencing. The sequence derived from these

clones is represented in Figure 3. This figure shows roughly how the deletions overlap.

In this thesis I undertook to sequence these deletion clones. Synthetic oligonucleotides were also constructed and used as sequencing primers to verify regions of ambiguity (see Figure 3). The interpretation of sequence data revealed a 1340 bp open reading frame which translates into a 440 residue peptide of 50 147 Daltons. This peptide sequence shares homology with the DNA binding domain of other steroid hormone receptors and the hydropathy of this region closely follows the pattern of *err1* and 2 (human estrogen related receptors). Putative intron splice junctions within the cDNA of *crf-1*, located on the carboxy side of each finger, align with known intron positions of other steroid hormone receptors.

Because the 3' UTR of the *crf-1* cDNA was unusually large, we decided to verify that the 3' end was part of the 2.2 kb clone rather than an artifact generated during the construction and cloning of the deletion. This was accomplished by Southern blotting. All deletion clones hybridized to only the appropriate 2.2 kb fragment.

Northern analysis would furnish information regarding the expression of *crf-1*. A blot using RNA isolated from the various developmental stages of *C. elegans* was probed with the 2.2 kb cDNA revealing a 2.2 kb message of low abundance expressed during all stages.

MATERIALS AND METHODS

I. DNA SEQUENCING

MATERIALS

Caenorhabditis elegans var. Bristol, strain N2 was the organism used in this study. The nematodes were maintained on NGM plates supplemented with *Escherichia coli* strain Op50 as a food source (Brenner, 1974). For large scale DNA and RNA preparations, worms were cultured on high peptone plates seeded with a lawn of *E. coli* strain B (Rose *et al*, 1982).

Restriction endonucleases used throughout this research were obtained from Pharmacia and BRL. All radioisotopes were obtained from NEN.

OLIGONUCLEOTIDES

Oligonucleotides were used as sequencing primers to verify regions of ambiguity (see figure 3). They were on average 20 nucleotides in length and were obtained from either Tom Atkinson at the Department of Medical Genetics, UBC or Bruce Brandhorst of the Institute of Molecular Biology and Biochemistry, SFU. The oligonucleotides were manufactured using an ABI 391 DNA synthesizer according to the phosphoamidite method of oligonucleotide synthesis. The oligonucleotides were purified by filtration through Millipore Sep-pack C10 cartridges prepared in the following manner. The cartridge was first equilibrated with HPLC grade acetonitrile and washed with distilled water. The oligonucleotide, dissolved in 1.5 mls of ammonium acetate, was loaded into the cartridge and washed with distilled water. It was eluted from the column with 20% acetonitrile. This fraction was evaporated to dryness using a Savant Speed Vac Concentrator. The

oligonucleotidess were rehydrated with sterile distilled water and the concentration was measured by spectrophotometry. Distilled water was used to make the appropriate dilutions and the stocks were frozen in small aliquots at -20°C for future use.

ISOLATION OF PLASMID DNA

Small scale plasmid DNA was purified from *E. coli* JM83 by Miniprep Spun Column kit purchased from Pharmacia. This procedure is based on the alkaline lysis method (Maniatis *et al*, 1982) with the following modification. If the plasmid DNA was to be used for sequencing it was resuspended in 50 *u*l of 1X TE (10 mM Tris, 1 mM EDTA at pH 8.0), denatured for 10 minutes at room temperature in 2 M sodium hydroxide and then purified by loading and spinning the columns for 2 minutes at 1000 RPM.

RESTRICTION DIGESTS

Restriction enzymes were used with the appropriate buffers under manufacturers recommended reaction conditions. These conditions include limiting the enzyme concentration to less than or equal to 5% of the total reaction volume and maintaining equivalent enzyme and buffer volumes. To ensure that genomic DNA was fully digested, 5 *u*g was incubated in excess enzyme for extended periods of up to 4 hours at 37°C.

AGAROSE GEL ELECTROPHORESIS

Agarose was dissolved most often to the concentration of 0.7% in 1X TBE (89 mM Tris, 89 mM borate, 2.5 mM EDTA). This is the concentration most effective for resolving DNA fragments of 0.1 to 10 kb by size (Maniatis *et al*, 1982). Ethidium bromide was added to a concentration of 0.2 *u*g/*u*ls in order to visualize

the fragments under UV light. Loading buffer was added to the DNA to a 1X final concentration (10X LB = 25% Ficoll 400, 0.25% bromophenol blue, 0.25% xylene cyanol in 1X TBE buffer). Gels were run in 1X TBE at 20-100 volts for 2-12 hours. Under short wave ultraviolet light of 302 to 365 nm they were photographed using Kodak Plus X Pan film.

ELECTROELUTION

Recovery of restriction digested DNA fragments from agarose gels was conducted by dissecting the fragment of interest from the gel and placing it in dialysis tubing covered with 600 ul of 1X TBE and sealed using removable clips (Maniatis *et al*, 1982). A current of 20-100 volts was then applied for 2-12 hours such that the DNA in the gel positioned closest to the cathode would be electroeluted into the buffer toward the anode. Electroelution was complete when the ethidium bromide intercalated into the DNA fragment could be visualized under UV light in the buffer fraction. The polarity of the electrodes was then briefly reversed to free up the DNA from contact with the dialysis tubing. The buffer was collected and spun for 15 minutes at $4^{O}C$ in a microcentrifuge in order to precipitate the residual agarose. The supernatant containing the DNA was ethanol precipitated in two volumes of 95% ethanol and 1/10 volume of 5 M ammonium acetate at -$20^{O}C$, washed with 70% ethanol and vacuum desiccated. The pellet was then resuspended in 1X TE.

SEQUENCING OF DNA

Denatured plasmid template DNA was obtained as previously described. The amount of DNA required per sequencing reaction was 3-5 ug. The sequencing

protocol followed was according to Sequenase Version 2.0 Kit (USB) which contained all reagents except radionucleotides which were supplied by NEN. This protocol is based on the dideoxy method of Sanger *et al* (1977) but slightly altered according to Tabor and Richardson (1984). The denatured plasmid was added to 0.5 pmol of appropriate primer (T3 forward, T7 reverse, M13 reverse or -40) and reaction buffer and denatured for 2 minutes at 65°C. The reaction was cooled for 15 minutes at room temperature to allow the primer to anneal to the template. This mixture is added to 0.1 M DTT, 7.5 uM of dGTP, dCTP and dTTP, 5 uCi ^{35}S-dATP and 1 unit of recombinant T7 DNA polymerase and incubated 2-5 minutes at room temperature. This reaction mixture is equally aliquoted into four prewarmed termination mixtures of G, A, T, or C which consist of deoxy/dideoxynucleotides. These are incubated for 5 minutes at 37°C. A stop solution of 95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF is added and the tubes are heated to 85°C for 2 minutes and put on ice just prior to loading.

Samples were electrophoresed on 6% acrylamide, 8 M urea gels at 50 amps for 6 hours in 1X TBE in the top buffer chamber and 1 M ammonium acetate in 1X TBE in the lower chamber. The gels were then transferred on to 3MM Whatman chromatography paper and covered with Saran wrap to be dried at 80°C for 20 minutes under vacuum. Afterwards, the gel was placed directly on Kodak XAR-5 film for 12 hours.

An amended protocol for sequencing was used to resolve DNA regions which displayed high secondary structure (Terry Snutch, personal communication). This method was used in conjunction with the Sequenase 2.0 Kit (USB). The primer at 2.5 pmol concentration is first annealed to 2-4 ug of denatured double stranded template DNA in the presence of 1 ul of DMSO at a higher temperature of 95°C for 3 minutes. The reaction mixture is frozen on dry ice and ethanol for 5 minutes, quickly thawed just prior to the addition of sequencing buffer and incubated at room

temperature for 5 minutes. The labelling mix which consists of DTT, dilute labelling mix, ^{35}S-dATP, dilute Sequenase 2.0 enzyme and 0.5 ul DMSO is added and the mixture incubated for another 5 minutes at room temperature. Aliquots are added to each dideoxy termination mix, incubated for 5 minutes at 37°C and 4 ul of stop mix is added and the tubes are denatured for 5 minutes at 95-100°C just prior to loading on the gel.

PROTEIN SEQUENCE ALIGNMENT AND PHYLOGENY OF *crf-1* AND OTHER STEROID RECEPTORS AND TRANSCRIPTION FACTORS

The DNA binding domain of *crf-1* was aligned with the corresponding region of other steroid hormone receptors and transcription factors (Figure 5 and 6). The other sequences were obtained from the sources cited in the Figure 5 legend. The computer program Clustal 4 compares amino acid identity in order to construct a phylogenetic tree. Clustal uses a distance matrix based on operational taxonomic units for sequence comparison. This program was used on the CI alignment and is shown in Figure 7. See page 35 for further details.

II. ORGANIZATION AND EXPRESSION OF *crf-1*

LABELING OF PROBES

DNA probes used for library screening, Southern or Northerns were labeled by nick translation (Maniatis *et al*, 1985), random priming with oligonucleotides (T7 Quick Prime Kit, Pharmacia) or end labeled by kinase (Maniatis *et al*, 1982).

The nick translation reaction requires approximately 500 ng of DNA to be incubated with alpha ^{32}P-dATP (80 Ci/mmole) as well as unlabeled nucleotides and alpha ^{32}P-dTTP (2.6 uM) in the presence of DNase and DNA polymerase I for

2 hours at 15°C. Unincorporated nucleotides were removed by chromatography through Sephadex G-25.

The T7 Quick Prime Kit protocol based on the method of Feinberg and Vogelstein (1983) requires greater than or equal to 50 ng of denatured DNA and alpha ^{32}P-dCTP (3000 Ci/mmole). This is incubated with a reagent mix of unlabeled nucleotides and T7 DNA polymerase for 15 minutes at 37°C. The reaction product of the T7 Quick Prime Kit was acid precipitated in TCA prior to counting.

End labeling of the 5' end with kinase requires DNA which lacks a 5' phosphate as in the case of synthetic oligonucleotides. This protocol involves incubation with T4 polynucleotide kinase, 100 mM magnesium chloride, 100 mM DTT and alpha ^{32}P-dATP (50 uCi) for 15 minutes at 37°C. This method also requires the removal of unincorporated nucleotides by chromatography through Sephadex G-25.

In all methods the specific activity of the labeled DNA was measured by Cerenkov counting in a scintillation counter. A specific activity of 1×10^7 cpm/ug could be achieved for probes labeled by these methods.

NORTHERN BLOTTING

Total RNA of mixed stages or stage specific total RNA was isolated from *C. elegans* by Terry Snutch according to the method of Chirgwin *et al* (1979). RNA at a concentration of 30 ug was denatured in 50% formamide, 1 mM EDTA, 5 mM sodium acetate, 10X MOPS at pH 7.0 for 15 minutes at 60°C. Loading buffer was added to a concentration of 1X and 2 ul of ethidium bromide (10 mg/ml) just prior to electrophoresis through 1.1% agarose gel containing 1X MOPS and 2.2 mM formaldehyde for 4-5 hours at 80 volts in 1X MOPS running buffer. An 8 ul aliquot

of BRL 0.24 - 9.5 kb RNA ladder was treated like a sample and used as a marker. The gel was photographed under UV light prior to transfer.

In preparation for transfer, the gel was soaked twice for 15 minutes in 0.025 M potassium phosphate at pH 7.5 and once in distilled water for 15 minutes. Genescreen (Dupont), the membrane for transfer, was soaked for 15 minutes in 0.025 M monobasic and dibasic sodium phosphate at pH 6.5. The blotting procedure was according to Maniatis *et al* (1982). The transfer was conducted for at least 12 hours. Upon disassembly, the membrane was rinsed in buffer to remove residual agarose, cross-linked at 254 nm for 40 seconds in a Stratagene cross-linker and baked for 2 hours at $80^{\circ}C$ under vacuum.

The membrane was prehybridized for at least 16 hours at $42^{\circ}C$ in 50% deionized formamide, 5X Denhardts, 5X SSC, 1% SDS and 100 ug/ml herring sperm DNA (Denhardt, 1966). Hybridization conditions were of moderate stringency using fresh prehybridization solution containing the denatured probe ($1x10^{7}$ cpm/ug) for 16-48 hours at $42^{\circ}C$. Two washes under each of the following conditions were performed: 2X SSC at room temperature for 5 minutes, 2X SSC and 1% SDS at $65^{\circ}C$ for 30 minutes and 0.1X SSC for 30 minutes at room temperature. The membrane was sealed in plastic and exposed to preflashed Kodak XAR-5 film at $-70^{\circ}C$ with an intensifying screen for over 48 hours.

SOUTHERN BLOTTING

DNA was prepared as previously described and 2-30 ug were restriction digested and size fractionated by electrophoresis on 0.7% agarose gels in 1X TBE for 12 hours at 20 volts. A photograph of the gel was taken. The excess gel was trimmed off and the remainder soaked in 0.25 M hydrochloric acid for 10 minutes at room temperature then briefly rinsed in distilled water. The gel was soaked three times each for 20 minutes at room temperature in 25 mM monobasic and dibasic

potassium phosphate (transfer buffer) with gentle agitation. Meanwhile, the membrane (Genescreen, Dupont) and 3MM Whatman chromatography paper were cut to the size of the gel and soaked for 5 minutes at room temperature in 25 mM transfer buffer. Transfer was set up according to Maniatis *et al* (1982) for 12 hours in transfer buffer.

After transfer the membrane was rinsed briefly in transfer buffer to remove residual agarose and crosslinked at 254 nm for 40 seconds in a Stratagene cross-linker, then dried for 30 minutes at 80°C under vacuum.

Prehybridization was conducted at 60°C in 5X SSPE, 5X Denhardts and 0.3% SDS for 4-12 hours. Hybridization conditions were identical to prehybridization conditions with the exception of the addition of 50 ul of denatured probe at $1.5x10^6$ cpm in a total volume of 10 ml for a filter of 70 cm^2. The minimum time of incubation was 12 hours. Four post hybridization washes were done; 1) 5 minutes at room temperature in 2X SSPE and 0.5% SDS, 2) 15 minutes at room temperature in 2X SSPE and 0.1% SDS 3) 30 minutes at 37°C in 0.1X SSPE and 0.5% SDS, 4) 30 minutes at 60°C in the previous solution. The damp membrane was then sealed in plastic and exposed to preflashed Kodak XAR-5 film at -70°C with a Dupont Cronex intensifying screen for not less than 12 hours.

RESULTS

I. SEQUENCE ANALYSIS OF THE *crf-1* cDNA

Sequence data for the *crf-1* cDNA was obtained by sequencing a series of overlapping cDNA deletions delta 1-12 (see Figure 2). Regions difficult to read through were subjected to the amended sequencing protocol. For verification, synthetic oligonucleotides were made, with identity to known flanking regions, to serve as sequencing primers. These allowed alternate access to these problem areas. Figure 2 shows the sequencing strategy of the deletion clones including oligonucleotide position and orientation in difficult to sequence regions. The sequence in its entirety is shown in Figure 3.

The sequence is 2292 bp in length with a predicted protein encoded by nucleotides 32 to 1348. There are two putative intron splice junctions, which conform to *C. elegans* consensus, located at nucleotide positions 341-347 and 451-457. The position of the second intron splice junction has been confirmed by alignment with partial genomic sequence. The genomic sequence (not shown) diverges from homology with the cDNA sequence at the 5' intron boundary. The genomic intron sequence shares nine out of ten identities with *C. elegans* 5' intron border consensus sequence (Emmons, 1988). These are shown in parentheses in Figure 3.

The putative protein sequence of the largest open reading frame is 440 amino acid residues in length and 50 147 Daltons, assuming an initiating methionine at nucleotide position 32. This sequence also has homology to the original oligonucleotide SR1 - GCCACTGGATATCACTACGGA (underlined in Figure 3) whose synthesis was based on identity to the DNA binding domain (C

Figure 2. Representation of the amount of sequence obtained from the deletion clones of crf-1. The sequence of the oligonucleotides used as sequencing primers is shown as well as their position and orientation with respect to the deletion clones. Numbers at the right end of the boxes indicate the number of nucleotides sequenced from each particular deletion clone. The hatched boxes indicate sequence obtained from oligonucleotide primers.

2.2 KB

CRF1 192

1 246

2 320

3 200

5 342

6 357

7 381

8 317

9 279

10 272

10.5 364

11 392

12 335

1 ⇒ 2 ⇒ 3 ⇒ 4 ⇐ 5 ⇐

OLIGONUCLEOTIDES

#1 AATCAGCCCAAAATCGTC

#2 CCGAGGCTTAGGCGTAGGC

#3 AGAAGATTGTAAGACGGGG

#4 GTCTCCACAGGCGAAACAG

#5 CAGCCACAACACAGGTCC
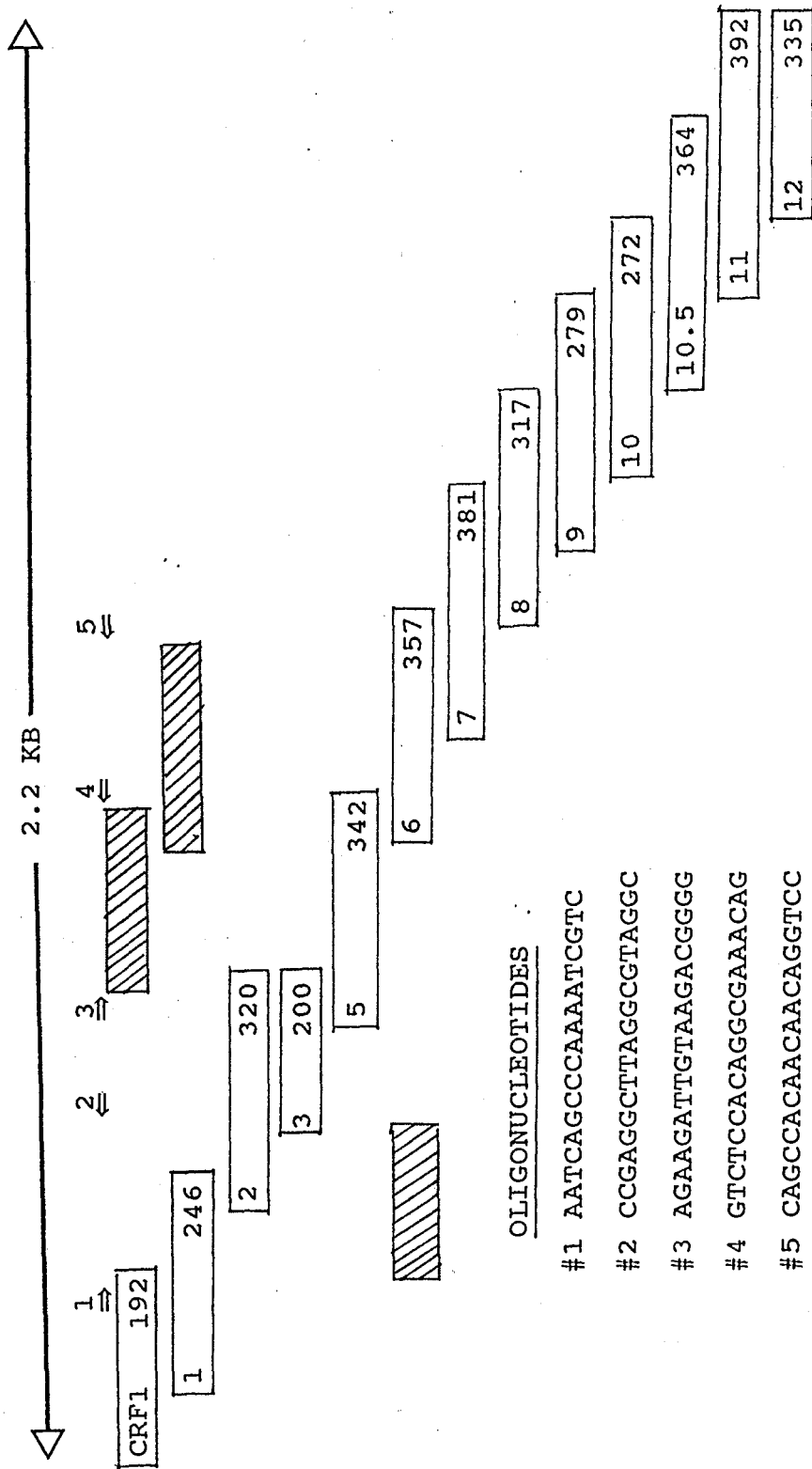
Figure 3 The complete nucleotide sequence of crf-1. The putative protein product is indicated below the nucleotide sequence. Circled amino acid residues indicate putative phosphorylation sites. The domains corresponding to known steroid hormone receptors are indicated in the right margin. The boxed residues are the conserved cysteines in the DNA binding region. The underlined region indicates the sequence and position of oligonucleotides originally used to isolate this clone. The boxed methioine represents the putative start codon. The two set of brackets denote putative intron splice junctions (Fields, 1989). The asterisk within the protein sequence indicates a stop codon.

18a


```
1    1    GAATTAACGAGCCTCCGCGCTCGCCGAAT ATG AACGGTACACTGGCCGATCAACTGCA      60
                                        M   N  G  T  L  A  D  Q  L  Q    10

61   11   AAGCCATCTGCCAACTACGGGCCAGCCAATGGTCAACTGCAGCCAACGCAACGAGGATCCGTC  120
          S  H  L  A  N  Y  G  Q  P  M  V  N (S) Q  R  N  E  D  P  S     30   } A/B

121  31   AATGTATATGAATGATCAGCCGCCTCCGTTCCCACCAACGGCTCATCCTCCTCGATGGG     180
          M (Y) M  N  G  S  A  A  S  V  S  H  T  N  G  S  S  S  M  G     50

181  51   CAATGATCAAAAATTCCGTCTTATAAGCGAATGGCTCAACGAGAAAGTGCCGAAGG       240
          N  D  Q  K  F  P (S) Y  K  R  M  A  Q  R  R  K  V  P  E  G     70

241  71   AGAGCTGTGCCGCAGTTGCTGGACCTTGCTGGGATATCACTACGGATGGCTAGCTG       300
          E  L  A  V [C] S  D  L  A  T  G  Y  H  Y  G  V  A  S [C]       90

301  91   CAATGGATGCAAAAACCTTCTTCCGACGAACAATAGTCAGTGA(CAA)ACCTTCATTTGCCA   360  } C
          N  G [C] K  T  F  F  R  T  I  V  S (E)(Q) T  F  I [C] Q      110

361  111  ATATAATGGAAATTGTGATGTTAATAAAAACATTCGATGTGCATGCGCCGTCACTGCCGCTT  420
          Y  N  G  N [C] D  V  N  K  N  I  R  C  A [C] R  H  C  R  F    130

421  131  CAACAAATGCCTATTGGTGGGATGGACGCAA(AAG)(CAA)TCCAGAATGATCGAGATCGGAT  480
          N  K [C] L  L  V  G  M  D  A (K)(X) I  Q  N  D  R  D  R  I    150

481  151  TGGGCCAACGAAGAGAAAATCAAGATGAGCAGCGGATCTGACGAGCAAGCCACCACTCC     540
          G  P (T) K  K  I  K  M  S  S  G  S  D  D  E  Q  A  T  T  P    170

541  171  ACACCGTCTCTTCAGGACCAGGAAATCATCGACCAGCTCACACAAGTTGAAGGGCTGTGCCA  600
          H  R  L  Q  D  Q  E  I  I  D  Q  L  T  Q  V  E  G  L  C  Q    190

601  191  GGAGTTGCGACGCTGCATAATTCCTGAAGTGACCGGAGTGACACATGCACTGACGTCCCC    660  } D/E
          E  L  R  R  C  I  I  P  E  V  T  G  V  T  H  A  L  T  S  P    210

661  211  TTGCCTACTTTTTGAGACCACTGACCTTAAAGTAGACGTGAGTCTGACCAATACGATTT    720
          C  L  F  E  T  T  D  L  K  V  D  V  S  L  T  N  T  I  F       230
```

```
721   CAAAGAGCTCTTCCCGGGCGTCGATGAACGACATTCGTATGTGGAACATTCGCGAGATGCG   780
231    K  E  L  F  P  A  S  M  N  D  I  R  M  W  N  I  R  E  M  R    250

781   CATCTGCATCGAGTGGGCAAAGACGTTCGATGTTTACCAGAGGCTCAACCTTTTTGATCA    840
251    I  C  I  E  W  A  K  T  F  D  V  Y  Q  R  L  N  L  F  D  Q    270

841   GTTTGCCCTCGTCCGAAACTTTGCCTTCGCATTTAATCTGCTCAACAGAGTGTTCTACTC    900
271    F  A  L  V  R  N  F  A  F  A  F  N  L  L  N  R  V  F  Y  S    290

901   GCCGGACCATGGACCAGATAAGATTGTCTTCCAAAATGGAGCCTTCATCATGAGACAGCC    960
291    P  D  H  G  P  D  K  I  V  F  Q  N  G  A  F  I  M  R  Q  P    310

961   ACAACAACAGGTCCAACTCTCGGGATGCGCCGGCCTATCTACACTCGACAAATGGACGAGAT  1020
311    Q  Q  Q  V  Q  L  S  G  C  R  P  I  Y  T  R  Q  M  D  E  I    330

1021  TATGATTCCATTCCGCAAATTGCAGTTGGCCGAGTTGCCACCTTCAAAGCCGC         1080
331    M  I  P  F  R  K  L  Q  L  S  V  A  E  F  A  T  F  K  A  A    350

1081  TTTGTTTTTTAATCCAGACGCCTTGGATTTGTCTCCCAGGGCAAACAGGAAGTTTTCGA    1140
351    L  F  F  N  P  D  A  L  D  L  S  P  Q  A  K  Q  E  V  F  E    370

1141  GGAGAGAAACAAATATTGGGGTGGACTTTTACCTGTATCACCAGAAAATTGGAATCCC    1200
371    E  R  N  K  Y  L  G  G  L  F  T  C  I  (T) Q  K  I  G  I  P  390

1201  GACTGGAGTTCAGAAATATGGAAGTCTTTAATGATGACCGCTAGTATACAGAATATTCT    1260
391    T  G  V  Q  K  Y  G  S  L  L  M  M  T  A  S  I  Q  N  I  L    410

1262  GGCGCAAAATGAGGAAAATATGCAAGTGATGGAATTGTTCAAAAACTGGGAAGTGGATCC   1320
411    A  Q  N  E  E  N  M  Q  V  M  E  L  F  K  N  W  E  V  D  P    430

1321  ATTTGTCAAGGAACTTTGTATGAAGAGAGCCATAAACGCCGAAATTGTTTATTATAAATAT  1380
431    F  V  K  E  L  C  M  K  R  A  *
```

```
1381   ATTGAAGTTTTTATGTTTAAGACATTAGTTGCCTTCAGATCTCTCACCAAGT          1440
1441   TTTGAGCAAAATGTTTCTCCCTATCATTATTTTATTATTATTGACTTTGACTTCG        1500
1501   CTGGGCCCAATACTTTTTTTCTGTGAAAAAAAACCAACTAGGACACGTCATAGATTC      1560
1561   CCCCGTTCCACACTCACAAACAGTTCCCCGTCTTACAATCTTCTAGTTCTGCTGTATATA   1620
1621   TTTGTGATAAAATAATATAAATGTTCAGTTTGTCCTTATTCTTTTCCCACCAGGTTTTT    1680
1681   TGTGCCATCGCTGAAATTAGTTGAGCACACTTTTTATAAATTAGCTGAGGTTTATGCT     1740
1741   TAGGTTTAGGCCGGAGGCTTAGGCGTAGGCTTGGATCTCTTTAGATCTTCAGTTTT       1800
1801   AGGCTCAGCTTTAGGCTTAACCTTCCCCATGCGCTCCTCAAAAAGCCTGTTAAT         1860
1861   GGCCGGTTGCCCTTACCTGAAAATTCACTAATTCCTCCCGTTTTCACAAGCCAGTTTTTA   1920
1921   AACCCTTTCCCAAACCAACCAACCTCGCCCAGTTCTGATCCAACCTTCCCACCTCTT      1980
1981   CCTCTTCCAAGCATTGTCATCGATTGCAATTTTTCAAATTATTCATGTTTCAACAAAAA    2040
2041   ATGTTTCCTCTATAAAACCCCTATTATTTACTGGTATCAATTCTTTTATTTCTCCCTA     2100
2101   CCACGATTTTGGGCTGATTTATCTTTTTTTTCCCAAATTTTTAACCAGTAAAAAATTGG    2160
2161   TTTGGTATTTGATTTCCGATTGCCCCTTACAAAATTCTTGATGTTCAAGTTTATTTCC     2220
2221   CCCACTTTTTTCTTTGTCATTCATTTTTTTTTAAATTTTTTTTCTCCCTCAATAAATCTT   2280
2281   CATTTCTTAAAA                                                   2292
```

Figure 4 A schematic diagram of the putative protein sequence of the DNA binding domain of crf-1. Zinc ions are shown interacting with eight highly conserved cysteine residues. The positions of oligo SR1 and SR2 used to screen the libraries are indicated by the lines. The regions enclosed by brackets indicate putative alpha-helical tertiary structure. Circled amino acids indicate those that are thought to be involved in dimer contacts. Boxed residues are thought to interact with phosphates of the DNA backbone. The residues marked by asterisk denote amino acids important in recognizing hormone response elements. The arrows indicate intron and putative intron positions of various steroid hormone receptors (Mader *et al*, 1989; Ritchie *et al*, 1990; Luisi *et al*, 1991).

domain) of various known steroid hormone receptors proteins (Beato, 1989). This identity spans nucleotides 241-291. The putative secondary structure of the DNA binding domain of *crf-1* is illustrated in Figure 4. The A/B domain could presumably span from nucleotides 9-241. The hinge region and the ligand binding domain (domains D and E) could occur after nucleotide 446. Homology to other proteins was investigated for all domains outside of the 'C' region. However, nothing significant was revealed. These homologies were calculated by a Fasta (Wilbur and Lipman, 1983) search of the Swiss protein database.

Putative kinase C and tyrosine kinase phosphorylation sites as predicted by Prosite (PC Gene) are indicated as circled residues in Figure 3.

PROTEIN SEQUENCE ALIGNMENT AND PHYLOGENY OF *crf-1* AND OTHER STEROID RECEPTORS AND TRANSCRIPTION FACTORS

An alignment of the DNA binding regions of the peptide product of *crf-1* and other steroid receptors and trans-cription factors is shown in Figure 5. These receptors were chosen because preliminary analysis indicated homology to the progesterone family of receptors. In addition several invertebrate receptors were included. Identical residues are denoted by asterisks and conservative changes are indicated by dots. Figure 6 is an alignment of just the CI region of the DNA binding domain. This alignment was used as data by the computer program Clustal 4 to construct a phylogenetic tree (Figure . This tree indicates that the *crf-1* protein product is most homologous to the CI region of estrogen and the estrogen related receptors. The next closest group includes; MTH, *svp*, COUP, E75A, *crf-2* and *tll*. *Crf-2* shares homology with E75A. On the other branch

Figure 5 An alignment of the DNA binding domain of a variety of steroid receptors. The program Clustal 4 was used with the following parameters: ktup = 1, window size = 10, filtering level = 2-5 and gap penalty = 3. The asterisks denote residues which are identical and the dots denote conservative amino acid substitutions. HMR - human mineralcorticoid, Err1 and 2 - human estrogen related receptors, COUPTF1 - chicken ovalbumin upstream transcription factor (Wang et al, 1989), HGR - human glucocorticoid, HPR - human progesterone, HER - human estrogen, HT3R - human thyroid, HVD3R - human vitamin D$_3$ (Schwabe and Rhodes, 1991), crf-1 and 2 - C. elegans receptor finger (R. Eagen and A. Sluder, personal communication), tll - Drosophila tailless, E75A - Drosophila ecdyson transcription factor, svp - Drosphila seven up, KNIRP, KNI and KNRL - Drosophila Knirp related, egon - Drosophila embryonic gonad (reviewed by Pignoni et al, 1990).

```
CRF1      ELCAVCSDLATGYHYGVASCNGCKTFFRRTIVSE--QTFI-CQYNGNCDV-------NKN IRCACRHCRFNKCLLVGM
CRF2      DRCMVCGDNSTGYHYGVQSCEGCKGFFRRSVHKNIA--YL-CTKGENCTFSYENCAANRG VRTRCQACAFAKCLAVGM
ERR1      RLCLVCGDVASGYHYGVASCEACKAFFKRTIQG--SIEYS-CPASNECEI------TKR RRKACQACRFTKCLRVGM
ERR2      RLCLIVCGDIASGYHYGVASCEACKAFFKRTIQG--NIEYS-CPATNECEI------TKR RRKSCQACRFMKCLKVGM
TLL       --CKVCRDHSSGKHYGIYACDGCAGFFKRSIRRSRQ--YV-CKSQKQ-----GLCVVDKT HRNQCRACRLRKCFEVGM
HT3R      ELCVVCGDKATGYHYRCITCEGCKGFFRRTIQKNLHPSYS-CKYEGKCVI------DKV TRNQCQECRFKKCIYVGM
E75A      --CRVCGDKASGFHYGVHSCKGCKGFFRRSIQQKIQ--YRPCTKNQQCSIL------RI NRNRCQYCRLKKCIAVGM
SVP       --CVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLT--YS-CRGSRN------CPIDQHHRNQCQYCRLKKCLK-GM
COUPTFI   IECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLT--YT-CRANRN------CPIDQH-RNQCYCRLKKCLKVGM
KNIRP     QTCKVCGEPAAGFHFGAFTCAGCKSFFGRSYNNI--STISECKNEGKCII------DKKNRTTCKACRLRKCYNVGM
KNI       --CKVCGEPAAGFHFGAFTCEGCKSFFGRSYNNI--STISECKNEGKCII------DKKNRTTCKACRLRKCYNVGM
EGON      --CKVCGEPAAGFHFGAFTCEGCKSFFGRSYNNI--STISECKNEGKCII------DKKNRTTCKACRLRKCYNVGM
KNRL      --CKVC.____GFHFGAFTCEGCKSFFGRSYNNL--SSISDCKNNGECII------NKKNRTACKACRLLLCLMCGM
HVDR      RICGVCGDK.__.GFHFNAMTCEGCKGFFRRSMKR---KAMFTCPFNGDCKI------TKDNRRHCQACRLKRCVDIGM
MTH       GRCAVCGDNASCQHYGVRTCAGCKGFFKRTVQK--SAKYI-CLANKDCPV-------DKRRRNRCGFCRFQKCLAVGM
HGR       KLCLVCSDEASGCHYGVLTCGSCKVFFKRAVEG--QHNYL-CAGRNDCII------DKIRRKNCPACRYRKCLQAGM
HMR       KICLVCGDEASGCHWGVVTCGSCKVFFKRAVEG--QHNYL-CAGRNDCII------DKIRRKNCPACRLQKCLQAGM
HPR       KICLICGDEASGCHYGVLTCGSCKVFFKRAMEG--QHNYL-CAGRNDCIV------DKIRRKNCPACRLRKCCQAGM
HER       RYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQG--HNDYM-CPATNQCTI------DKNRRKSCQACRLRKCYEVGM

            *  .   .    *     *     *  .  *      .           *         *   *       *   **
```

Figure 6 An alignment of the CI region of the DNA binding domain of a variety of steroid receptors and transcription factors. This alignment was calculated by the same program and under the same parameters as those in Figure 5. Sequence sources are identical to those in Figure 5.

```
MTH        CAVCGDNASCQHYGVRTCAGCKGFFKRT
SVP        CVVCGDKSSGKHYGQFTCEGCKSFFKRS
COUPTFI    CVVCGDKSSGKHYGQFTCEGCKSFFKRS
E75A       CRVCGDKASGFHYGVHSCKGCKGFFRRS
CRF2       CMVCGDNSTGYHYGVQSCEGCKGFFRRS
TLL        CKVCRDHSSGKHYGIYACDGCAGFFKRS
HER        CAVCNDYASGYHYGVWSCEGCKAFFKRS
ERR1       CLVCGDVASGYHYGVASCEACKAFFKRT
ERR2       CLVCGDIASGYHYGVASCEACKAFFKRT
CRF1       CAVCSDLATGYHYGVASCNGCKTFFRRT
HVDR       CGVCGDRATGFHFNAMTCEGCKGFFRRS
HT3R       CVVCGDKATGYHYRCITCEGCKGFFRRT
KNIRP      CKVCGEPAAGFHFGAFTCAGCKSFFGRS
KNI        CKVCGEPAAGFHFGAFTCEGCKSFFGRS
EGON       CKVCGEPAAGFHFGAFTCEGCKSFFGRS
KNRL       CKVCGEPAAGFHFGAFTCEGCKSFFGRS
           *  **  .  ..   *.     .*  .*  .** *.
```

Figure 7 A phylogenetic tree of the DNA binding domain of various steroid receptors and transcription factors. This tree was constructed using the alignments in Figure 6 and the parameters of gap penalty (fixed) = 10 and gap penalty (varying) = 10. The vertical axis represents relative time since divergence.

MTH
SVP
COUPTFI
E75A
CRF-2
TLL
HER
ERR1
ERR2
CRF-1
HVDR
HT3R
KNIRP
KNI
EGON
KNRL
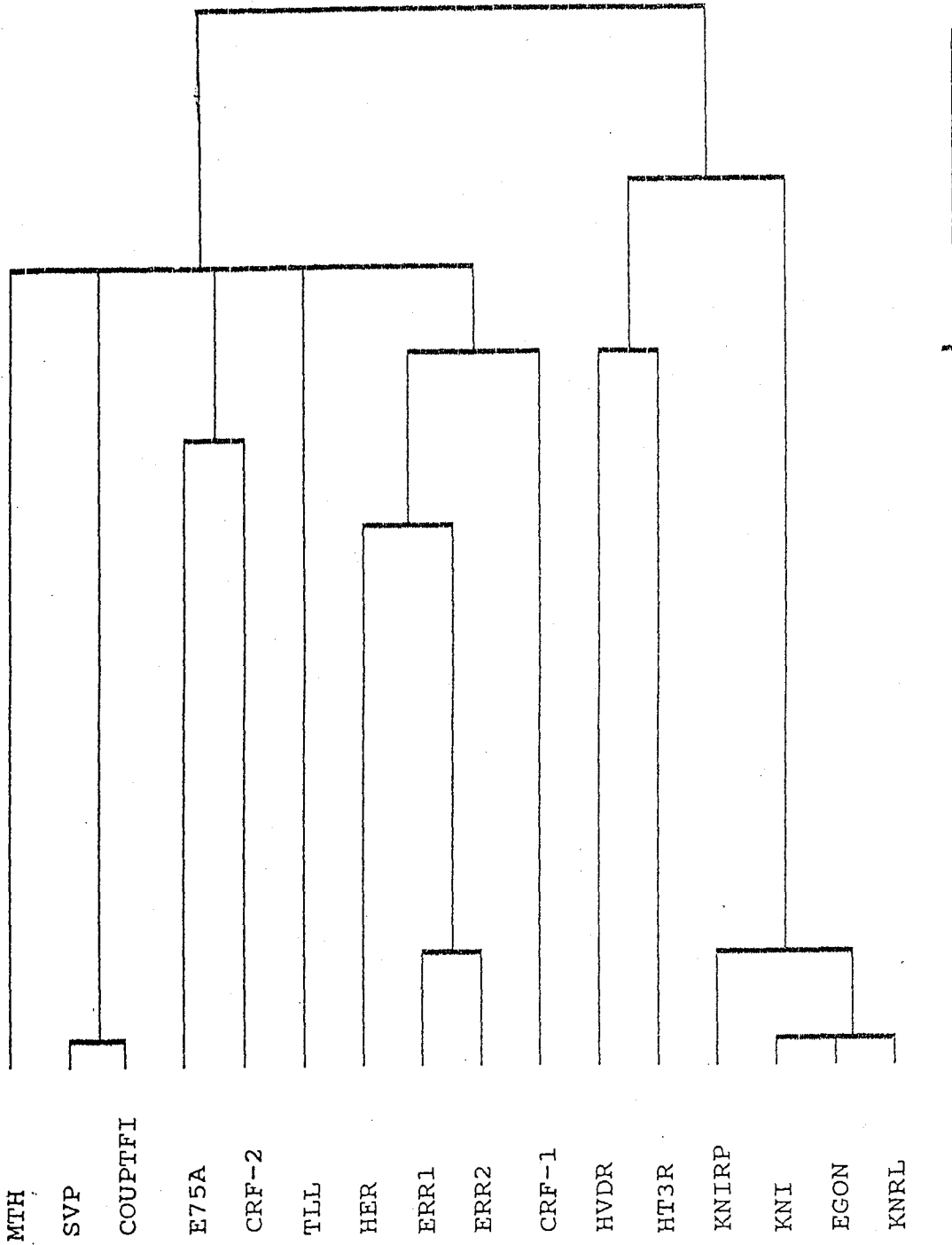
lie HVDR, HT3R, and the Drosophila KNIRP, KNI, *egon* and KNRL.

HYDROPATHY PLOT OF THE PREDICTED PROTEIN SEQUENCE

Figure 8 shows a hydropathy plot of a portion of the largest open reading frame of *crf-1* as predicted by SOAP program (PC Gene). The plots of human estrogen related steroid receptors *err1* and *err2* are superimposed over *crf-1*. The first has 55.2% identity over 67 amino acids as predicted by a Fasta (Wilbur and Lipman, 1983) search of the Swiss protein data base. *Err2* has 53.7% identity over 67 amino acids. The DNA binding region spans amino acids 70-140 in *crf-1*. The plot in this region is similar in shape and amplitude to regions 1-120 of *err1* and 80-200 of *err2*.

II. ORGANIZATION AND EXPRESSION OF *crf-1*

CHROMOSOMAL LOCALIZATION

J. Sulston and A. Coulson have localized *crf-1* to the right terminus of the *C. elegans* X chromosome (Figure 9). This was accomplished by fingerprinting phage 9aSR1 and comparison to the fingerprint database for the overlapping cosmids.

NORTHERN ANALYSIS

A Northern blot of the various developmental stages of *C. elegans* RNA was probed with the 2.2 kb EcoR1 fragment of the cDNA subclone (Figure 10). This was done in order to show that the size of the cDNA clone corresponded to the actual message size. The arrows indicate the position of 28s and 18s rRNA. This blot also reveals information concerning the level of expression during various

Figure 8 Hydropathy plot of crf-1, err1 and err2 by SOAP (PCGene). The solid line represents the plot of crf-1 from amino acids 40-160. The dotted line profiles err1 from 1-120 and the dashed line is err2 from 80-200. The sequences from Evans (1988). The x axis represents amino acid residues and the y axis is representative of relative hydrophobicity and hydrophilicity of stretches of 15 amino acids.
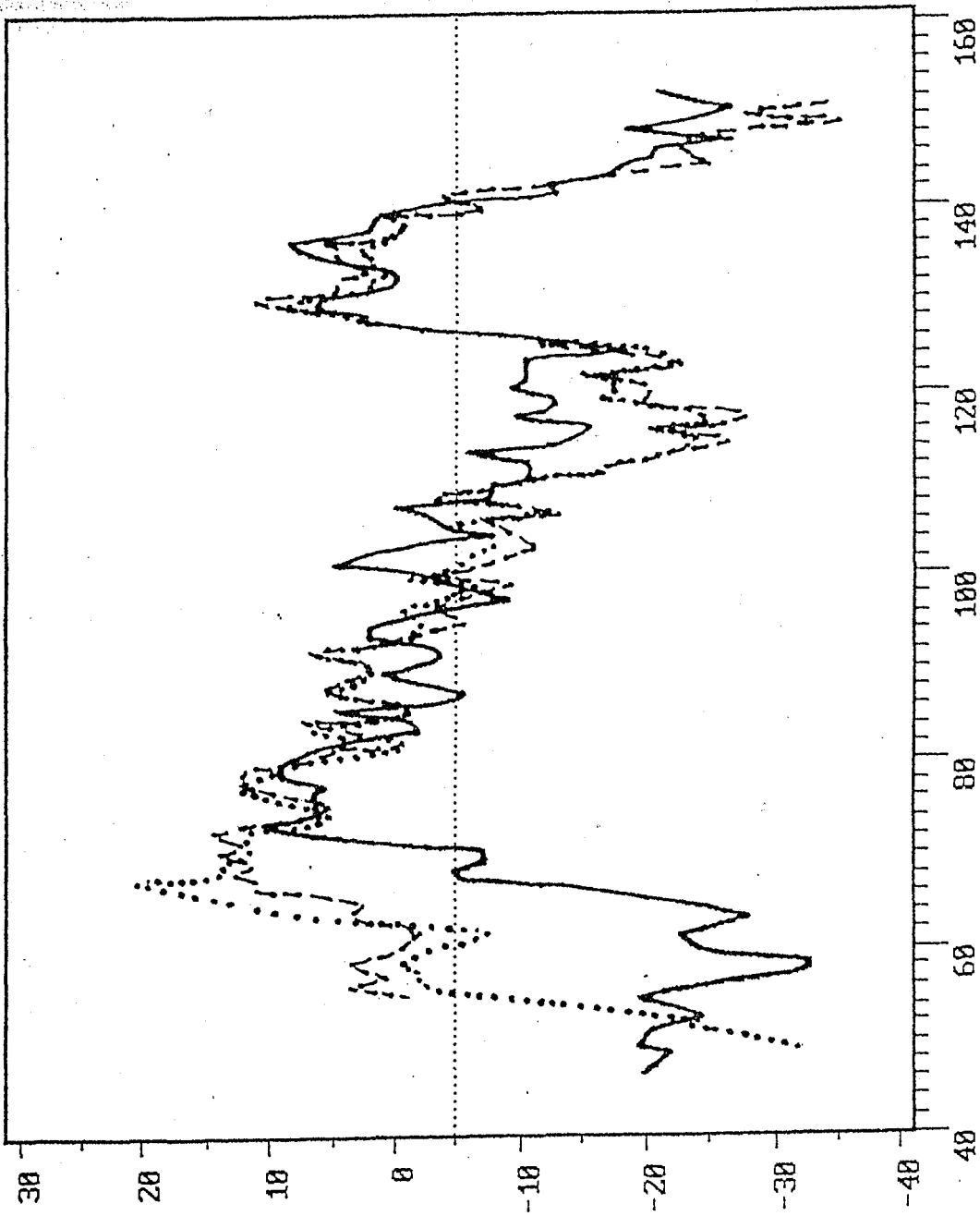
Figure 9 Chromosome maps of the right end of the X chromosome of *C. elegans*. The symbols indicate the positions of three markers.

a. The genomic clone called BH#SR19A which hybridizes to crf-1 occurs at the extreme right end of YAC Y52G2.

b. This diagram indicates by symbols the relative positions of several genetic markers.

a

(X)

(Y12EI1)

K10H3 TU#52
C53C11
ZK1061 *
(Y40C7)
C01B1 *
C33E10 *
K02C2 *

(X)

C01G11 *
C4B63
ZK693 *
F11C7 *
N05C12
T08G *
BB3 *
C52B5 *

(Y5262)

C43H4

ZK54
T20.11A *
ZK52 *
F58A10
T04F1
C36E6 *
C46H5

(K)

BH#SR19A *

R03G11
B0574
C2404
R105
TU#10
(Y25H1)
(Y13A12)

B03B5 *
C15E10 *
F39B3
K09F12 *
T19A3
F31A3 *
F57C11
C2T06

C04B11 *
T19G6

(Y6BA3)

T08B5 *
K03F3 *
F20B4 *
(Y22E4)
K09G11 *
M62
R2F3 *
R03F8 *

b

lin-15

let-2
sdc-1 osm-1

TCMEC4D
BW#108 TCMEC4B
TCMEC4C sup-10
stP2 TCMEC4A

sup-10

mec-4, osm-1 *

let-2

sdc-1

let-3, 10, 11, 35

let-15, 18, 38, 40
lin-15
mec-5

27
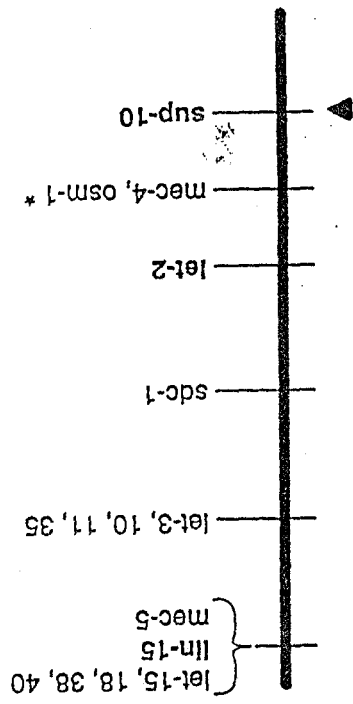
Figure 10 Developmental northern blot analysis of crf-1. *C.elegans* RNA (30 ug/lane) from all developmental stages was probed with the 2.2 kb cDNA of *crf-1*. The position of the ribosomal bands is indicated by the arrows. Marker RNA was run in the far right lane. The original size estimate of 2.2 kb may be an underestimate. See discussion.

ADULT L3/4 L2 L1 TOTAL POLY A+
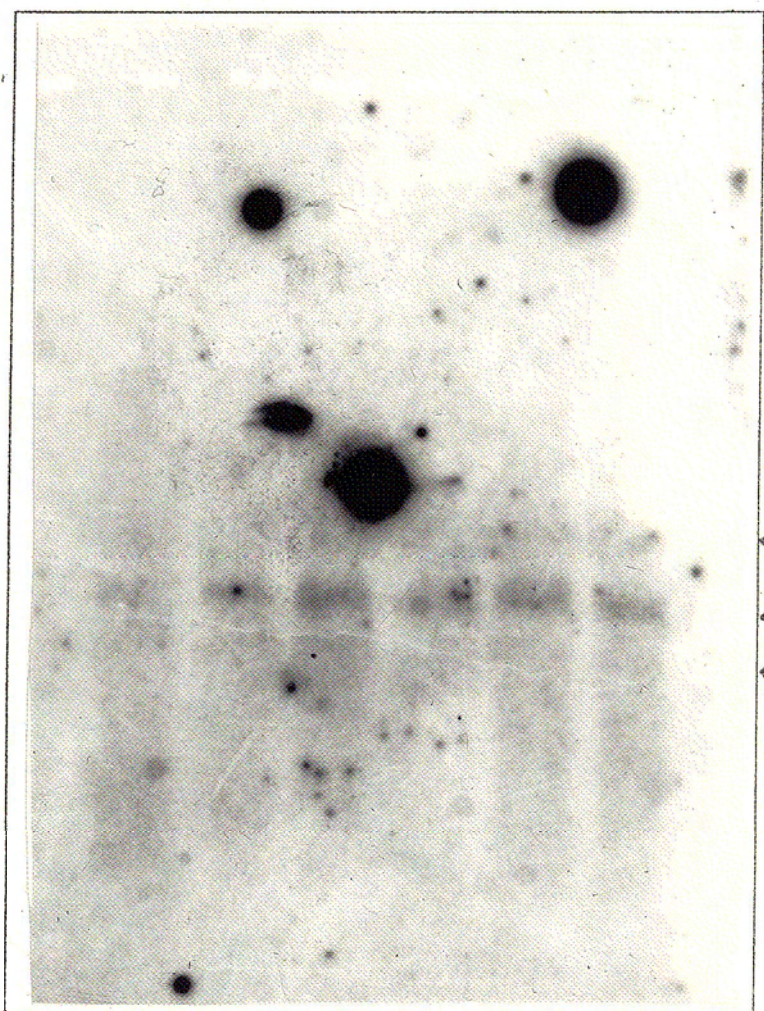
ORIGIN

28s rRNA

2.2 kb

18s rRNA

developmental stages. A potential band is visible in all lanes at the 2.2 kb size and the intensity of hybridization is similar. Figure 11 shows the same Northern blot stripped and reprobed with hsp70A (Heschl, 1988) a control whose message is 2.4 kb in size. The position of 28s and 18s rRNA is indicated by arrows. The expression of this message is low during L1 but increases throughout L2, L3/4 and adult (Prasad and Baillie, 1989).
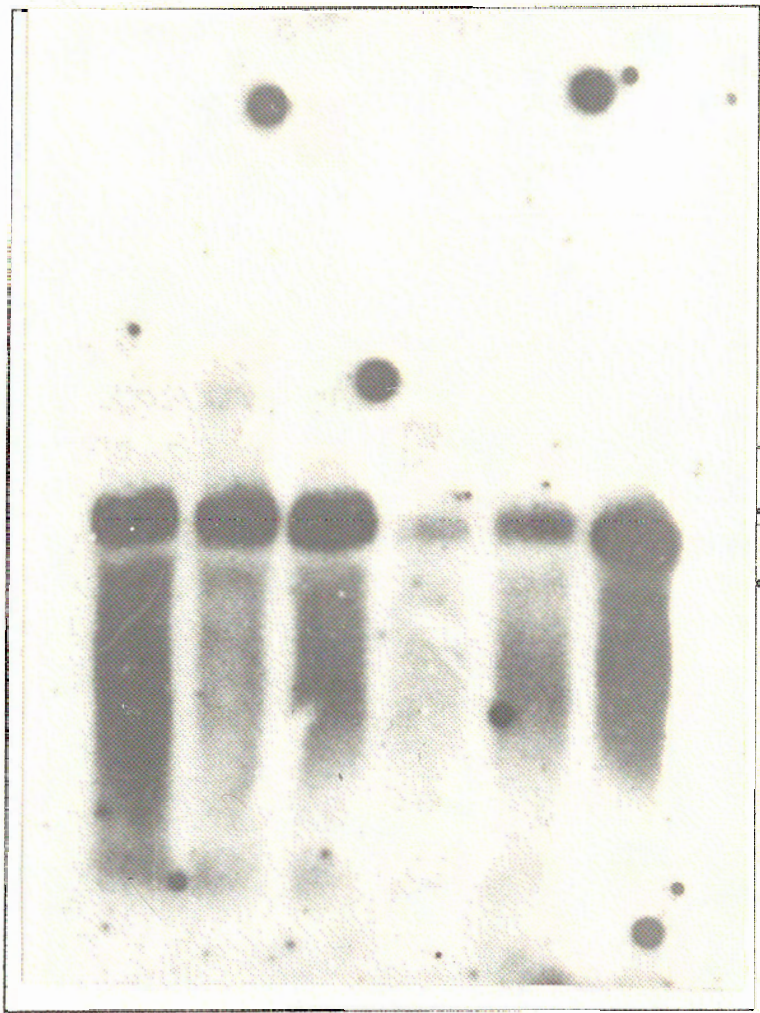
SOUTHERN ANALYSIS

A Southern blot of the EcoR1 restricted original cDNA clone called cD3A 4.5Eco was probed with either end of the EcoR1 fragment of the 2.2 kb cDNA to verify that the 2.2 kb deletion subclones were not artifacts of the deletion protocol. The blot, probed with oligonucleotide #1 located within the 3' end, hybridizes with a 2.2 kb band (Figure 12). The same blot probed with a PstI/AflIII fragment of 115 bp from the deletion clone #12 is shown is Figure 13. A 2.2 kb band is visible.

Figure 11   Developmental northern blot analysis of control hsp70A (Prasad *et al*, 1989). Blot is identical to that in Figure 10 but has been stripped and reprobed. The position of the ribosomal bands is indicated by the arrows.

Figure 12 Southern blot analysis of Eco RI digested original 4.5 kb cDNA clone probed with oligo #1 located at the 3' end of the 2.2 kb cDNA clone. This was carried out to verify that the deletion clones from the 3' end were not artifacts of construction. This confirmed that the 3' sequence obtained did indeed originate from the 3' end of the 2.2 kb clone.
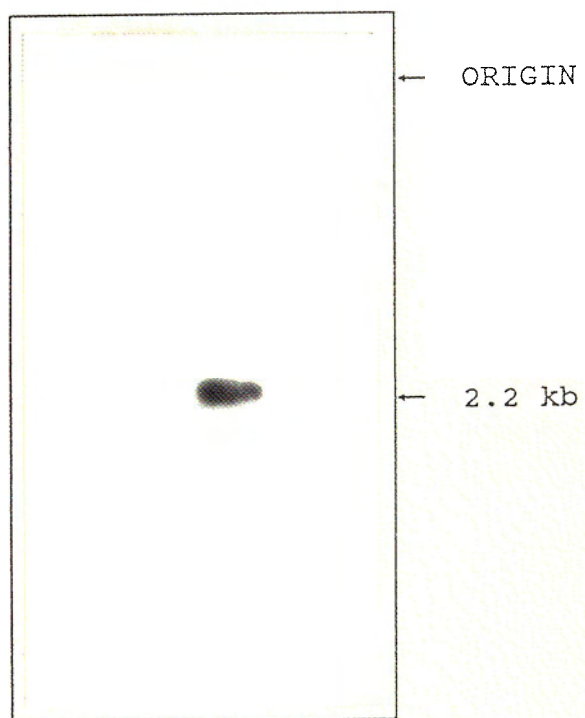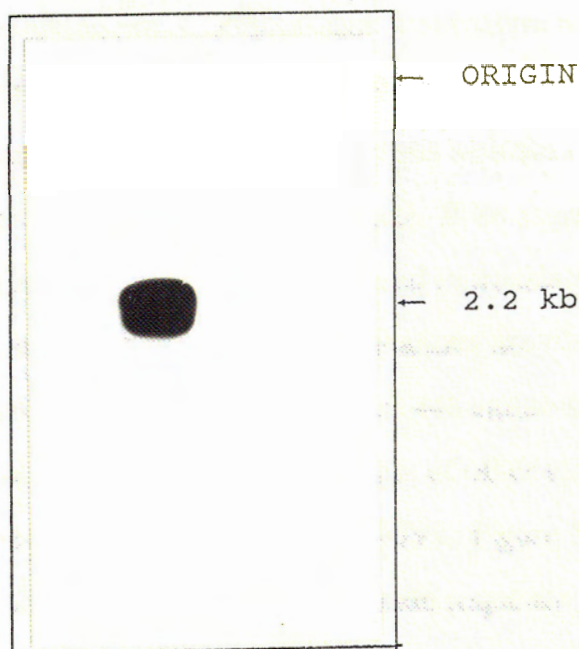
Figure 13 Southern blot analysis of Eco RI restricted original 4.5 kb cDNA clone probed with a 115 bp Pst I/Afl III fragment from deletion clone #12. This blot was also done for the reason stated in Figure 12.

DISCUSSION

I. SEQUENCE ANALYSIS OF THE *crf-1* cDNA

The sequence data obtained for the *crf-1* cDNA spans 2292 nucleotides and has a large open reading frame of 1340 nucleotides (Figure 3). It is within this ORF that peptide consensus to the DNA binding region occurs (domain C). This reading frame has several features which seem to indicate that it may code for a mRNA which could be translated into a protein. The first feature is an AUG start codon at position 32. The putative ribosome binding site has the sequence GAATTATGA and an identity occurs in eight out of ten positions with the *C. elegans* consensus sequence (M.D.Perry, personal communication). If we assume that this codon is the initiating methionine for translation and proceed to translate the remaining sequence in this reading frame several other features are observed. The size of crf-1 protein is predicted to be 50 147 Daltons for its 440 amino acids. The most obvious feature is the identity with the DNA binding site of other steroid hormone receptors. This identity spans nucleotides 269-291. Figure 5 shows an alignment of the protein sequence of this region. *Crf-1* protein sequence has been aligned with the C domain of other steroid receptors. The DNA binding region consists of two fingers which contain ten highly conserved cysteine residues that are thought to interact with zinc ions allowing the tertiary formation of two "zinc fingers" (Green and Chambon, 1988) which in turn interact with sequence specific areas of the hormone response element that allow interaction with other factors that regulate transcription (Beato, 1991). In all cases these two fingers are encoded by separate exons (Green *et al*, 1988) except for COUP-TF1 where both fingers are found in the first exon (Ritchie *et al*, 1990). This finger motif was originally detected in the Xenopus protein TFIIIA (Miller *et al*, 1985; Evans *et al*, 1988). Since the two fingers

of steroid hormone receptors are similar one might speculate that they have arisen by duplication, however there are amino acid substitutions that indicate they have since diverged significantly. The first exon called CI is closest to the amino terminal of the peptide and contains four of the conserved cysteine residues and the remainder of the residues are hydrophobic (Wahli and Martinez, 1991). In the *crf-1* protein the conserved cysteine residues lie at peptide positions 248, 257, 299 and 308 (Figure 3). Nonpolar amino acids comprise 50% of the residues in this area. The second finger called CII contains six conserved cysteines and several basic amino acids. These residues can all be found within the putative second finger of the *crf-1* protein. The splice junction following the first finger in other steroid hormone receptors seems to occur at one of three positions with respect to the last conserved cysteine residue of this finger. The *erb-A* junction is located at the second amino acid following the last conserved cysteine residue of the first finger. VD3R and NGF1-B junctions are found at the fifth amino acid following the same cysteine, while AR, PR, and ER are at the tenth amino acid. The junction as indicated by the sequence of the *crf-1* cDNA lies between the eleventh and twelfth amino acid. The remainder of an intron splice junction consensus sequence is located at nucleotides 341-346 in *crf-1* cDNA sequence. This putative splice junction contains remnant *C. elegans* 5' donor sequence RAG in juxtaposition to 3' consensus NNR (C. Fields, personal communication). Figure 14 shows the positions of the exon/intron splice junctions for several other steroid hormones. For this particular sequence, this intron remnant sequence is GAG/CAA. There is also another potential intron splice junction with the sequence AAG/CAA at position 451-457 (see Figure 3). The position of an intron in *crf-1* has been confirmed by comparison of cDNA sequence with partial genomic sequence. This was obtained from a 0.7 kb Acc1 fragment derived from the genomic phage 9aSR1 (see end of appendix section). Figure 15 shows that the genomic sequence is identical to the cDNA nucleotide

Figure 14 This is the protein sequence of the DNA binding domain of *crf-1*. The positions of the oligonucleotides SR1 and SR2 used to screen the libraries are indicated by parentheses. The arrows indicate intron and putative intron positions of the various steroid hormone receptors (Luisi *et al,* 1991)

AR, PR
ER, VD3R
NGF1-B, crf-1
erb-A, COUP

VD3R
erb-A NGF1-B crf-1

ELCAVCSDL(ATGYHYG)VASCNG(CKTFFRR)TIVSEQTFICQYNGNCDVNKNIRCACRHCRFNKCLLVGMDAK

SR1
CI
Specific DNA Binding

SR2
AR
PR
ER

CII
Non-specific DNA Binding
(dimerization?)

positions 389 to 456. The genomic sequence shares nine out of ten positions with
C. elegans consensus sequence for a 5' intron border sequence (Emmons, 1988).
This directly follows the second zinc finger and corresponds in position exactly to
the junction position of ER, AR, PR, NGF1-B, VD3R, *erb-A* and COUP (Green
and Chambon, 1988; Ritchie *et al*, 1990). This position is the seventh amino acid
following the last conserved cysteine of the second finger. This appears to be a
highly conserved location (Ritchie *et al*, 1990).

A Fasta (Wilbur and Lipman, 1983) search of the Swiss protein database was
conducted on the putative protein of *crf-1* in regions flanking the DNA binding
domain thought to correspond to A/B, D and E. Homologies to known proteins
were not revealed. This is not unusual since other putative transcription factors
have been identified that have unknown ligand activators such as COUP-TF (Power
et al, 1991).

The alignment of the protein sequence of the CI region of the DNA binding
region of *crf-1* and other steroid receptors and transcription factors is shown in
Figure 5. The CI region is most highly conserved between receptors. This
alignment served as data for the Clustal 4 program which predicts a phylogenetic
tree based on UPGMA scores (unweighted pair group method with arithmatic
mean). This is shown in figure 7. This tree shows that *crf-1* protein shares
homology with HER and *err1* and *2*. Also within this branch but diverging much
earlier are MTH, *tll*, *svp* and COUP, E75A and *crf-2*. An even earlier divergence
separates this group from HVDR, HT3R, KNIRP, KNI, *egon*, and KNRL proteins.
In a search of the Swiss protein database with the DNA binding domain of the *crf-1*
protein, the Fasta program predicted that it shared identity with *err1* and *2*. The
Fasta program uses the amino acid replacement matrix of Dayhoff *et al* (1978) while
Clustal uses a distance matrix based on operational taxonomic units (OTU's) for
sequence comparison. OTU's are pairs of sequences grouped for comparison.

The large 3' end of the *crf-1* cDNA which lacks a contiguous open reading frame from nucleotides 1350-2292 is not unique among *C. elegans* transcripts. *Fem-3*, a sex determining gene of *C. elegans*, has a 268 nucleotide 3' untranslated region (3' UTR). This 3' UTR is thought to play a role in the regulation of *fem-3* expression during development by binding a factor that inhibits translation (Ahringer and Kimble, 1991 and Lawson *et al*, 1991). This may provide an explanation for the existence of such a large 3' UTR in the *crf-1* cDNA sequence.

Trans-splicing is a phenomenon well documented in the mammalian parasite *Trypanosoma brucei* (Hannon *et al*, 1990) and nematodes such as *Ascaris lumbridicoides* (Maroney *et al*, 1990) and *C. elegans* (Blumenthal and Thomas, 1988 and Nilsen, 1989 and Thomas *et al*, 1988). The *trans*-splicing status of this *C. elegans* cDNA is unknown without more of the 5' sequence.

Phosphorylation of steroid hormone receptors is at present a confusing issue. There seems to be a lack of consistency with regards to site and substrate between different types of hormone receptors as well as a lack of agreement and evidence regarding the function of phosphorylation and also when this reversible post-translational modification occurs. In the *crf-1* protein there are four sites which conform to the consensus sequence of substrate requirements for protein kinase C (Prosite, PC Gene). The substrate residue is a serine or threonine. These sites are indicated in Figure 4. They span nucleotides 0-1182 and lie within the amino half of the putative protein product of *crf-1*. Two sites at positions 99 and 201 fall within the putative A/B domains while the remaining three sites occur within the putative D/E domains. These features resemble progesterone receptors in that the kinase acts on the identical substrate and known sites occur near the amino terminus (Moudgil, 1990). There is only one putative tyrosine kinase phosphorylation site at nucleotide position 126, very close to the amino terminus. This conforms to a

pattern found in estrogen receptors in that usually tyrosine kinase is involved, however the target sites usually fall near the carboxy terminus (Moudgil, 1990).

Sequence homology within the DNA binding domain tentatively indicates that *crf-1* is more closely related to the estrogen receptor family. If one examines the amino acid residues that are thought to interact with the hormone response element we see that this relationship is recapitulated. The peptide sequence of *crf-1* on the carboxy side of the first finger at residue position 91-95 is NGCKT. This sequence shares identity at three of five positions with the estrogen receptor sequence of E/DGCKX and only two of five sites with glucocorticoid receptor whose sequence is GSCKV (Danielsen *et al*, 1989; Mader *et al*, 1989). The fact that this sequence is not 100% conserved in the *crf-1* protein suggests that it may interact with unique response elements. Its function may prove to be quite interesting.

The sequence of the *crf-2* protein was obtained by A. Sluder, Harvard Medical School, Boston. She employed a similar strategy using oligonucleotides to screen two cosmid libraries to recover nuclear hormone receptors in *C. elegans*. She recovered a clone that is most closely related to mouse thyroid hormone receptor. The clone was called *crf-2* and has been localized to chromosome I. This gene has a 1.6 kb transcript and is present in embryos but not in larva and only at low levels in early adults (A. Sluder, personal communication). The *crf-2* protein product is thought to be involved in some aspect of embryogenesis.

The function of *crf-1* is unknown at present yet a steroid receptor in *C. elegans* has been characterized. *Daf-12* is a dauer inducing hormone receptor (W.H.Yeh, personal communication). This receptor is thought to transcriptionally activate genes in the dauer larva formation pathway (W.H.Yeh, personal communication). There is a variety of steroid receptor homologues in *Drosophila melanogaster* that have well defined developmentally important functions. The

product of the gene tailless (*tll*) plays a role in establishing nonmetameric domains at the anterior and posterior poles of embryos (Pignoni et al, 1990). A receptor protein product of E75A and its ligand 20-hydroxyecdysone are important in the regulation of larval molting (Feigl et al, 1989). A retinoic acid receptor homologue a product of *usp*, is also involved in pattern formation in Drosophila (Oro et al, 1990).

II. ORGANIZATION AND EXPRESSION OF *crf-1*

NORTHERN ANALYSIS

The size of mRNA was originally estimated to be 2.2 kb. However, closer examination of figures 10 and 11 indicate *crf-1* mRNA may in fact be larger based on slower mobility than the known heat shock mRNA (2.4 kb). A potential RNA band, visible in figure 10, may correspond to the expression of a *crf-1* message. This message appears to occur throughout all stages of development. The position of 18s and 28s rRNA has been indicated in figure 10. It is also possible that the signal of the 2.2 kb message may simply reflect artifactual hybridization to rRNA. Further work with a more pure source of polyA+ RNA would be required to verify that this is a real message. Equivalent amounts of RNA were run from each developmental stage and the intensity of hybridization is approximately constant. When one compares the expression profile of *crf-1* to *hsp70A* (Heschl, 1990) the differences in the intensity of hybridization between the *crf-1* probe and the *hsp70A* probe indicate that the expression of *crf-1* occurs at lower levels and the message is not as abundant as *hsp70A*. The *hsp70A* transcript is 2.4 kb in size and the abundance of the message changes throughout development (figure 11). The expression of *hsp70A* is low during L1 but increases significantly through out L2, L3/4 and adult (Prasad and Baillie, 1989).

SOUTHERN ANALYSIS

Southern blotting was used to verify that the deletion clones spanning the 3' UTR were not artifacts of deletion construction. In Figures 12 and 13 both the 3' and 5' ends of the sequenced cDNA clone *crf-1* hybridize to only the 2.2 EcoRI fragment of the larger 4.5 EcoRI original cDNA clone.

III. CONCLUSIONS

Crf-1 cDNA appears to code for a peptide that has an initiating methionine at position 32 and is 440 amino acid residues in length. Its predicted size is 50 147 Daltons. This protein product shares identity with the DNA binding domain of other steroid hormone receptors and transcription factors especially *err 1* and *2* and also *crf-2*, Drosophila *tll*, E75A, *svp*, human thyroid and COUP..

There is no apparent homology outside of the DNA binding region according to a FASTA protein database search. There are other proteins with unknown ligand activators. The *crf-1* protein product is most homologous to estrogen and estrogen related receptors according to a phylogenetic tree. This tree was predicted by the Clustal program using an alignment of the CI region of the DNA binding domain.

The cDNA sequence of *crf-1* has a large 3' untranslated region that lacks a contiguous reading frame. This region may be a 3' UTR and could have regulatory capabilities.

There are a variety of sites which could serve as substrates for cellular kinases in the putative protein sequence of *crf-1*. These sites are found in other steroid hormone receptors but their function and role are as yet unknown. The weak similarities do not contribute to resolving the identity of *crf-1* as a member of either family of steroid hormone receptors.

Amino acid residues of the *crf-1* protein important for HRE interaction share identity with the estrogen family. Lack of total conservation suggests that a unique response element may be involved.

The location of splice junctions following the DNA binding fingers supports identity of *crf-1* as that belonging to the superfamily of steroid/thyroid hormone receptors.

One cannot draw definite conclusions about *trans*-splicing with respect to the cDNA sequence of *crf-1*.

Northern blot analysis reveals that the *crf-1* message could be expressed during all stages of *C. elegans* development. The potential transcript is 2.2 kb in size and abundance is relatively low.

PROPOSALS FOR FUTURE RESEARCH

There are many observations surrounding *crf-1* that require further investigation and clarification. Is the 3' end really untranslated, and if so, is it involved in regulation?

The genomic clones 9aSR1 3.3Eco and 9aSR10.8Eco should be sequenced to provide information about genomic organization. They could also be used to probe a Southern blot of genomic *C. elegans*. This same blot probed with the 2.2 kb cDNA should be repeated for clarification.

One could investigate the temporal and spatial pattern of regulation of *crf-1* by *in situ* hybridization. An alternative method would be to subcloning the cDNA sequence linked to a reporter gene into a transformation vector and use it to generate transgenic worms (A. Sluder, personal communication). The expression of *crf-1* as indicated by the presence of the reporter gene product could be monitored closely since anatomy and the process of development in *C. elegans* are well characterized.

The protein product of *crf-1* could be amplified in an expression vector and characterized biochemically.

The finger swap assay could be used to identify possible ligands of *crf-1*. This could be accomplished by constructing a hybrid of the DNA binding domain of various receptors with known response elements linked to *crf-1* ligand binding domain. Subsequent trials would reveal which ligand would activate transcription.

One could also test if *crf-1* can rescue adjacent chromosomal deficiencies or mutations.

APPENDIX

ORIGIN OF THE CLONES

The original genomic clone phage 9aSR1 was isolated by A. Purac from Charon 4 genomic *C. elegans* library (Snutch, 1984). This phage was EcoRI digested to yield fragments of 9.0, 3.3, 0.95 and 0.8 kb. These were subcloned, by A. Purac, into the EcoRI site of vector pUC19 (Yanisch-Perron et al, 1985). The fragments of interest were 3.3 and 0.8 kb and were called 9aSR1 3.3Eco and 9aSR1 0.8Eco.

The 4.5 kb cDNA clone isolated from the lambda ZAP cDNA (Barstead,) was EcoRI restricted and fragments of 2.2, 1.6, 0.8 and 0.1 kb were obtained. The 2.2 kb fragment hybridized to the genomic phage 9aSR1 on a Southern blot (not shown) and was subsequently subcloned into the EcoRI site of pBluescript KS (Stratagene). This clone is referred to as cDNA 2.2Eco and nested deletions were made to allow DNA sequencing.

SUBCLONING

All phage clones were subcloned into pBluescript KS (Stratagene) and pUC19. The fragment to be cloned is cut with enzymes corresponding to unique restriction endonuclease sites within the polylinker region of the vector. The vector is cut with the same enzymes and purified by phenol/chloroform extraction and ethanol precipitation. To ligate vector to insert, reactions were set up in as small a volume as possible in a 1:3 molar ratio. The appropriate buffer and T4 DNA Ligase were added and the mixture incubated overnight at 15°C (Maniatis *et al*, 1982).

Once ligation was complete, a fraction of the mixture was used to transform competent cells. To select for recombinant plasmids the bacteria was grown on NZYM plates with ampicillin (100 ug/ml) and IPTG (160 ug/ml) and XGAL (40 ug/ml). These conditions allow one to distinguish non-recombinants with functional B galactosidase (B gal) activity from recombinants which lack the functional gene due to insertional inactivation by the fragment.

TRANSFORMATION

Competent *E. coli* JM83 cells (Mandel and Higa, 1970) were incubated with 100 ng of plasmid DNA on ice for 15 minutes then heat shocked at 42°C for 2 minutes (Hanahan, 1983). The cells were then plated on NZYM plates (Maniatis *et al*, 1982) to which 100 ug/ml of ampicillin was added in order to allow for selection of transformants carrying the resistance gene on the plasmid. Positive colonies were isolated and grown up overnight at 37°C in liquid NZYM and ampicillin and a fraction of the culture was frozen at -70°C in 50% glycerol.

GENERATION OF NESTED DELETIONS FOR DNA SEQUENCING

Nested deletions for sequencing were generated according to the protocol of Henikoff (1984). This procedure requires two restriction enzyme sites within the polylinker of the vector. The cuts should result in a blunt ended or 5' overhang and in a four base 3' overhang. The enzymes used were KpnI and XhoI or SstI and XhoI on 5 ug of the 2.2 kb cDNA clone of *crf-1* in pBluescript KS vector. The restriction digest was extracted with phenol, then chloroform and precipitated in 2.5 volumes of 95% ethanol and 0.2 N sodium chloride. The DNA was washed in 70% ethanol.

This was lyophilized and resuspended in a buffer of 66 mM Tris-HCl pH 8.0 and 0.66 mM magnesium chloride. Five hundred units of the enzyme exonuclease III was added and 2.5 ul fractions of the reaction were collected every 30 seconds and blunted with S1 nuclease for 30 minutes at room temperature. These products were run on a 0.7% agarose gel to ensure that the reaction had worked. Following ethanol precipitation in 5 M ammonium acetate, the pellet was dissolved in 1X TE and subjected to Klenow DNA polymerase for 10 minutes at 37°C to ensure that blunt ended products were produced. These products were then recircularized with T4 DNA ligase at 15°C overnight. The plasmids were then transformed as previously described and the plasmid DNA was extracted in preparation for sequencing according to the Miniprep protocol. The plasmids were cut with PvuII and XhoI or XbaI and size fractionated on 0.7% agarose gels to ensure that size staggered subclones resulted.

GENOMIC SEQUENCE DERIVED FROM PHAGE 9aSRI

A 3.3 kb EcoRI subclone derived from phage 9aSRI was digested with AccI and three smaller fragments (9aSRI2.0AccI, 9aSRI0.7AccI, 9aSRI0.5AccI) were isolated and subcloned. The ends were sequenced by the dideoxy method as previously described, and sequence derived from the end of the 0.7 kb clone 9aSRI0.7AccI is shown in figure 15. Sequence from the other ends failed to match any cDNA sequences.
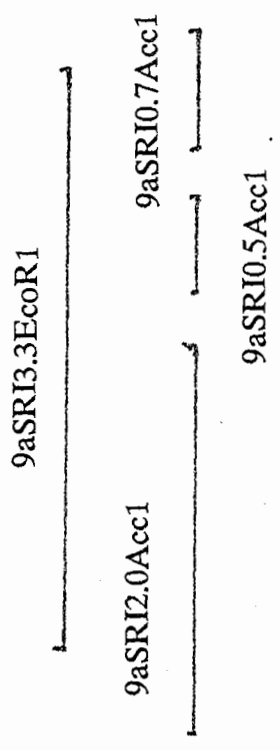
Figure 15

a. Genomic sequence of a 0.7 kb Acc1 subclone of 9aSR1. The nucleotide positions refer to the cDNA sequence in figure 3. The dots indicate identity with a 5' intron border consensus sequence.

b. Map of the Acc1 subclones of the genomic clone 9aSR13.3EcoR1. (The order is unknown).

a

389   456
                                                        ·······
AACATTCGATGTGCATGCCGTCACTGCCGCTTCAACAAATGCCTATTGGTGGGGATGGACGCAAAAGGTAAGTGA

b

9aSRI3.3EcoR1

9aSRI0.7Acc1

9aSRI0.5Acc1

9aSRI2.0Acc1

REFERENCES

Ahringer, J. and J. Kimble. 1991. Control of the sperm-oocyte switch in *Caenorhabditis elegans* hermaphrodites by the *fem-3* 3' untranslated region. Nature. **349**: 346-348.

Albertson, D. G. 1985. Mapping muscle protein genes by *in situ* hybridization using biotin-labeled probes. European Molecular Biology, J. **4**: 2439-2498.

Barstead, R.J. and R. H. Waterston. 1989. The basal component of the nematode dense-body is vinculin. J. Biol. Chem. **264** (17): 10177-10185.

Beato, M. 1987. Induction of transcription by steroid hormones. Biochemie Biophys Acta. **910**: 95-102.

Beato, M. 1989. Gene regulation by steroid hormones. Cell. **56**: 335-344.

Beato, M. 1991. Transcriptional control by nuclear receptors. FASEB. **5**: 2044-2051.

Blumenthal, T and J. Thomas. 1988. *Cis* and *trans* mRNA splicing in *C.elegans*. Trends in Genetics. **4** (11): 305-308.

Brenner, S. 1974. The genetics of *Caenorhabditis elegans*. Genetics **77**: 71-94.

Chambon, P., A. Dierrich, M. Gaub, S. Jakowlev, J. Jongstra, A. Krust, J. LePennec, P. Oudet and T. Reudelhuber. 1984. Promoter elements of genes coding for proteins and modulation of transcription by estrogen and progesteron. Recent Progress in Hormone Research. **40**: 1-39

Chirgwin, J.M., A. E. Przybyla, R .J. MacDonald and W. J. Rutler. 1979. Isolation of biologically active ribonucleic acid for sources enriched in ribonuclease. Biochem. **18**: 5294-5299.

Danielsen, M., L. Hinck and G. M. Ringold. 1989. Two amino acids within the knuckle of the first zinc finger specific DNA response element activation by the glucocorticoid receptor. Cell. **57**: 1131-1138.

Dayhoff, M., R. M. Schwarty and B. C. Orcutt. 1978. "Atlas of protein sequence and structure". Vol. 5, Suppl. 3. (Ed. M. Dayhoff) pp. 345. National Biomedical Research Foundation. Silver Spring, Maryland.

Denhardt, D. T. 1966. Biochem. Biophys. Res. Commun. **23**: 641.

Emmons, S.W. 1988. The Genome in The nematode *Caenorhabditis elegans*. (Ed. W.B. Wood) pp. 56,59,61. Cold Spring Harbor Laboratory.

Evans, R.M. 1988. The steroid and thyroid hormone receptor superfamily. Science. **240**: 889-895.

Evans, R.M. and S. M. Hollenberg. 1988. Zinc fingers: guilt by association. Cell. **52**: 1-3.

Feigl, G., M. Gram and O. Pongs. 1989. A member of the steroid hormone receptor gene family is expressed in the 20-OH-ecdysone inducible puff 75B in *Drosophila melanogaster*. Nucleic Acids Research. **17** (18): 7167-7179.

Feinberg, A.P. and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem. **132**: 6-13.

Gehring, U. 1987. Steroid hormone receptors: biochemistry, genetics and molecular biology. Trends in Genetics. **12**: 399-402

Glass, C.K. and J. M. Holloway. 1990. Regulation of gene expression by the thyroid hormone receptor. Biochemie Biophys Acta. **1032**: 157-176.

Green, S. and P. Chambon. 1988. Nuclear receptors enhance our understanding of transcription regulation. Trends in Genetics. **4**: 309-314.

Green, S., V. Kumar, I. Theulaz, W. Wahli and P. Chambon. 1988. The N terminal DNA binding 'zinc finger' of the estrogen and glucocorticoid receptors determines target gene specificity. European Molecular Biology, J. **7** (10): 3037-3044.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**: 557-580.

Hannon, G.J., P. A. Maroney, J. A. Denker, and T. W. Nilsen. 1990. *Trans* splicing of nematode pre-messenger RNA *in vitro*. Cell. **61**: 1247-1255.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III in DNA sequence analysis. Gene. **28**: 351-59.

Heschl, M. F. P. 1988. Characterization of the hsp70 multigene subfamilies from *C. elegans*. Ph.D. thesis. SFU.

Kenyon, C. 1988. The nematode *Caenorhabditis elegans*. Science. **240**: 1448-1453.

Lawson, D.N., T. Rosenquist and J. Kimble. 1991. Post-transcriptional regulation of *fem-3*. in *Caenorhabditis elegans* 1991 Meeting Abstracts. pp.31.

Luisi, B. F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto, P. B. Sigler. 1991. Crystallographic analysis of the interactions of the glucocorticoid receptor with DNA. Nature. **352**: 497-505.

Mader, S., V. Kumar, H. deVerneuil and P. Chambon. 1989. Three amino acids of the estrogen receptor are essential to its ability to distinguish an estrogen response element from a glucocorticoid response element. Nature. **338**: 271-273.

Mandel, M and A. Higa. 1970. Calcium dependent bacteriophage DNA infection. J. Mol. Biol. **53**: 154-155.

Maniatis, T., E. F. Fritsch and J. Sambrook. 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, New York.

Maroney, P. A., G. J. Hannon, J. A. Denker, and T. W. Nilsen. 1990. The nematode spliced leader RNA participates in *trans*-splicing as an Sm snRNP. European Molecular Biology, J. 9 (11): 3667-3673.

Miller, J., A. D. McLachlan and A. Klug. 1985. Repetitive zinc binding domains in the protein transcription factor IIIA from Xenopus oocytes. European Molecular Biology, J. 4: 1609-1614.

Moudgil, V.K. 1990. Phosphorylation of steroid hormone receptors. Biochemie Biophys Acta. 1055: 243-258.

Nilsen, T.W. 1989. *Trans*-splicing in nematodes. Experimental Parasitology. 69: 413-416.

Oro, A. E., M. McKeown and R .M. Evans. 1990. Relationship between the product of the Drosphila ultraspiracle locus and the vertebrate retinoid X receptor. Nature. 347: 298-301.

Pignoni, F., R. M. Baldarelli, E. Steingrimsson, R. J. Diaz, A. Patapoutian, J. R. Merriam and J. A. Lengyel. 1990. The Drosophila gene tailless is expressed at the embryonic termini and is a member of the steroid receptor superfamily. Cell. 62: 151-163.

Power, R. F., J. P. Lydon, O. M. Connelly, B. W. O'Malley. 1991. Dopamine activation of an orphan of the steroid receptor superfamily. Science. 252: 1546-1548.

Prasad, S. and D. L. Baillie. 1989. Evolutionary conserved coding sequences in the *dpy-20-unc-22* region of *C. elegans*. Genomics. 5: 185-198.

PC Gene @1988-1990 Amos Bairoch. Dept. Biochimie Medicale, University of Geneva, Switzerland.

Ritchie, H. H., L. Wang, S. Tsai, B. W. O'Malley and M. Tsai. 1990. Coup TF gene: a structure unique for steroid/thyroid receptor superfamily. Nucleic Acids Research. 181 (23): 6857-6862.

Rose, A. M., D. L. Baillie, E. P. M. Candido, K. A. Beckenbach and D. Nelson. 1982. The linkage mapping of cloned restriction fragment length differences in *Caenorhabditis elegans*. Mol. Gen. Genet. 188: 286-291.

Rothe, M., U. Nauber and H. Jackle. 1989. Three hormone receptor-like Drosphila genes encode an identical DNA-binding finger. European Molecular Biology, J, 8 (10): 3087-3094.

Sanger, R., S. Nicklen and A. R. Coulson, 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74: 5463-5467.

Schwabe, J. W. R., and D. Rhodes. 1991. Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. Trends in Biochemistry. 16: 291-296.

Snutch, T. 1984. A molecular and genetic analysis of the heat shock response of *C. elegans*. Ph.D. thesis. SFU.

Tabor, S. and C. C. Richardson. 1984. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA. **84** (14): 4767-4771.

Thomas, J. D., R. C. Conrad and T. Blumenthal. 1988. The *C. elegans trans*-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. Cell. **54**: 533-539.

Wang, L., S. Y. Tsai, R. G. Cook, W. G. Beattie, M.Tsai and B. W. O"Malley. 1989. COUP transcription factor is a member of steroid receptor superfamily. Nature. **340**: 163-166.

Wahli, W. and E. Martinez. 1991. Superfamily of steroid nuclear receptors: positive and negative regulators of gene expression. FASEB. **5**: 2243-2249.

Wilbur, W. J. and D. J. Lipman. 1983. Fasta computer based sequence data searches. Proc. Natn. Acad. Sci. USA. **80**: 726.

Wood, W. B. 1988. The genome of *Caenorhabditis elegans*. (ed. W.B.Wood) pp.8-12. Cold Spring Harbour Laboratories.

Yamamoto, K.R. 1985. Steroid receptor regulated transcription of specific genes and gene networks. Ann. Rev. Genet. **19**: 209-252.