Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 156–159, January 1995
Biochemistry

Steroid/thyroid hormone receptor genes in *Caenorhabditis elegans*

ZDENEK KOSTROUCH, MARTA KOSTROUCHOVA, AND J. EDWARD RALL*

Genetics and Biochemistry Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892

Contributed by J. Edward Rall, September 20, 1994

ABSTRACT **The large family of steroid/thyroid hormone receptor (STR) genes has been extensively studied in vertebrates and insects but little information is available on it in more primitive organisms. All members possess a DNA binding domain of zinc fingers of the C2, C2 type. We have used the polymerase chain reaction with degenerate oligonucleotide primers covering this region to clone three distinct members of this family from the nematode *Caenorhabditis elegans*. All three belong to the retinoic acid receptor (RAR), thyroid hormone receptor subfamily of genes. The cDNA of one of these clones shows such a high homology to DHR3, an early ecdysone response gene found in *Drosophila*, and MHR3, identified in *Manduca sexta*, that we have termed it CHR3. Furthermore, the C-terminal portion of the deduced protein sequence shows a box containing eight identical amino acids among CHR3, DHR3, and MHR3 suggesting an identical specific ligand for these proteins. CNR8 shows homology to NAK1, and CNR14 has homology to both the RAR-γ1 gene and to another ecdysone response gene, E78A. Neither of the latter two cDNAs is a clear homologue of any known gene and each is distinctive. All of these genes are expressed varyingly in both larval and adult stages of nematode development as shown by Northern blot analyses. These data demonstrate that the STR family of genes is represented in a nematode whose ancestor appeared well before the branching that gave rise to the Arthropoda and Chordata.**

More than 200 different cDNAs of the steroid/thyroid hormone receptor (STR) gene family have been cloned and sequenced from a variety of organisms. Many of these sequences are clear homologues—i.e., the thyroid hormone receptor (TR) genes from rat, chicken, mouse, human, and frog code for a highly similar protein that binds thyroid hormone. However, the term homologue can be ambiguous; in addition to the above, it can also mean evolutionarily related, as are all of these STR genes whether they bind steroids, retinoids, or thyroid hormone or have no known ligands. Our interest has been to discern the evolutionary history of this gene family. The family has been variously divided into 3–10 different groups (1–3). Sequences are available mostly from vertebrates and arthropods, although one sequence is known from an echinoderm (4). All of these genes can be divided into five or six domains denoted A–F (2, 3) (see Fig. 4 *A*). The A/B domain is most N terminal and is poorly conserved. The C domain is the zinc finger DNA binding domain and is so highly conserved as to identify the superfamily (5). The sequence of the distal part of the first zinc finger, termed the P box, distinguishes the three main subfamilies of the STR (2). The D, E, and F domains contain the activating domain, a dimerization domain, and the ligand binding domain (2, 3, 6, 7). We have searched for related genes in several invertebrate phyla and report here the identification and cDNA sequences of three genes from *Caenorhabditis elegans*.† The sequences of the P box of the cDNAs we have isolated places all of them in the TR/retinoic acid receptor (RAR) subfamily. All of these genes are expressed, and we also report the results of Northern blot analyses of different stages in *C. elegans* development.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

A *C. elegans* cDNA library was purchased from Stratagene (catalog no. 9037006). The library was screened by PCR with degenerate primers devised from sequences in the DNA binding domain of the STR family (Fig. 1). Five microliters of the original library (4×10^8 phages) and 0.5 μM upstream and downstream primers were used in a final vol of 100 μl containing 200 μM each deoxynucleotide and 1.5 mM $MgCl_2$ using the Perkin–Elmer/Cetus PCR kit with AmpliTaq polymerase. The first three cycles consisted of denaturing at 94°C for 1.5 min, annealing at 65°C for 3 min, and elongating at 72°C for 5 min. Fifteen subsequent cycles were identical except for the use of 60°C for annealing and were followed by 20 cycles at an annealing temperature of 57°C. *Taq* polymerase (2.5 units per reaction) was added during the first annealing step of the first cycle. Five microliters of the PCR mixture after the first round of amplification was subjected to a similar second round of amplification with a new set of nested primers possessing a longer 3′ sequence. DNA fragments were separated by electrophoresis on 1% agarose gel in 0.5 × TBE and fragments of ≈130 bp were excised. These fragments were cloned by a CloneAmp system (GIBCO/BRL) after 8–10 cycles of PCR with primers having 5′ flanking $(CUA)_4$ or $(CAU)_4$. Clones were directly sequenced from plasmid minipreparations (Promega Magic Minipreps) by dideoxynucleotide sequencing (Sequenase, United States Biochemical). The library was then plated and screened according to the manufacturer's conditions by ^{32}P-labeled probes specific for each of three different clones. Positive clones were excised from the phage to the plasmid Bluescript using *in vivo* excision by the ExAssist system (Stratagene) supplied with the library. One clone, CNR8, which was not obtained after repeated screening, was cloned by PCR with primers homologous to the previously cloned short sequence in both upstream and downstream directions and a second primer derived from the vector. Clones of cDNA were sequenced by walking in both directions using the Prism Dye-Deoxy cycle sequencing kit and the automatic sequencer (model 373A; Applied Biosystems) and confirmed by manual sequencing with Sequenase. Worms (N2; *C. elegans* wild type, DR subclone of CB original) were generously supplied by Todd Starich of the *Caenorhabditis* Genetic Center and were cultured as hermaphrodites with weekly transfer on NGM plates fed by *Escherichia coli* OP-50, which was supplied with the worms. The different stages were obtained as described (8). RNA was prepared with an RNA-Stat 60 kit (Tel-Test, Friendswood, TX). RNA was separated on methylmercuric agarose gel or formaldehyde/agarose at 10 V/cm. RNA was transferred to a GeneScreen*Plus* membrane

Abbreviations: RAR, retinoic acid receptor; STR, steroid/thyroid hormone receptor; TR, thyroid hormone receptor.

*To whom reprint requests should be addressed.

†The sequences reported in this paper have been deposited in the GenBank data base [accession nos. U13075 (CHR3), U13076 (CNR8), and U13074 (CNR14)].

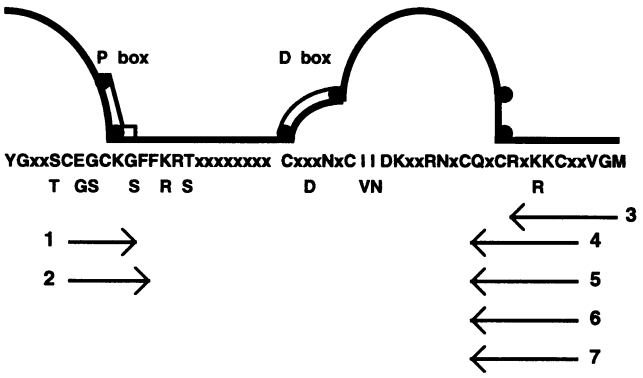

FIG. 1. Schematic representation of oligonucleotides used as primers for amplification of a part of the DNA binding domain. Consensus sequence was adapted from Laudet *et al.* (2). Sequences: 1, 5′-TGYGARGGYTGYAARRG-3′; 2, 5′-TGYGARGGYTGYAARRYTTCTT-3′; 3, 5′-CATICCIACIIIVADRCA-3′; 4, 5′-RCAYTTTCTBWRGCGRCAGKMYYSRCA-3′; 5, 5′-RCAYTTTCTBWRYCTRCAGKMYYSRCA-3′; 6, 5′-RCAYTTYTKGARGCGRCAKWRYYGRCA-3′; 7, 5′-RCAYTTYTKGAR-YCTRCAKWSYYGRCA-3′ (where R represents A or G, Y represents C or T, K represents G or T, S represents C or G, M represents A or C, W represents A or T, D represents A or G or T, V represents A or C or G, B represents G or C or T, and I is inosine). Positions of cysteines are indicated by solid circles.

(DuPont/NEN) using a PosiBlot system (Stratagene) or by capillary transfer. Membranes were hybridized to ^{32}P-labeled probes and examined with a bioimaging analyzer Bas-2000 (Fuji) and by autoradiography. Sequences were analyzed with the sequence analysis software package of the Genetics Computer Group. For reverse transcription, 5 μg of total RNA prepared from the different stages was treated for 15 min at 25°C by RNase free DNase I (Promega). DNase I was inactivated by a 15-min incubation at 65°C in the presence of 1.5 mM EDTA. Reverse transcription was performed with 50 units of StrataScript RNase H reverse transcriptase (Stratagene) in a final vol of 50 μl. The enzyme was then inactivated by heating to 90°C for 5 min. For synthesis of the second strand, 5 μl of this reaction mixture was subjected to PCR with primers designed for the region in the 3′ end of cDNAs with conditions as described above. The continuous clone of CNR8 was obtained from L3 cDNA by PCR with conditions as described above.

RESULTS AND DISCUSSION

We have used PCR with oligonucleotide primers patterned after the DNA binding domains, which are the most conserved region of the STR family (Fig. 1) to amplify sequences in a cDNA library of *C. elegans*. A second round of amplification was required to produce enough product of the expected size (130 bp) to be visualized by UV light on an ethidium bromide-stained agarose gel. Cloning and sequencing of these bands revealed three different products, all of which were homologous to the STR family. Two cDNA clones (CHR3 and CNR14) were obtained by screening the original library with probes based on sequences determined from the amplified bands. To obtain CNR8 we used an additional PCR with primers based on sequences from the known cDNAs and from the vector. Two separate clones with an overlap of 23 nucleotides were obtained. The total sequence of this clone was obtained by reverse transcription and PCR of RNA from larval stage L3. The three sequences we have cloned (CHR3, CNR8, and CNR14) consist of 2577, 1821, and 1907 nucleotides with open reading frames for 553, 463, and 556 amino acids. The Kozak consensus sequence around the initiating ATG is not conserved, nor is there an A nucleotide at the −3 position as frequently found in *C. elegans* (9). The 10-base sequence in the 5′ untranslated region of CNR14 is identical to the last 10 bases of the trans-splicing region described as SL1 by Krause and Hirsh (10). Alignment of amino acid sequences deduced from cDNAs indicates that all three cDNAs belong to the STR family with a fully conserved characteristic molecular signature of DNA binding domain consisting of nine cysteine residues and one methionine residue arranged in the form C-X_2-C-X_{13}-C-X_2-C-X_{15}-C-X_5-C-X_9-C-X_2-C-X_4-C-X_4-M (Fig. 2). All three cDNAs belong to the TR/RAR subfamily as evidenced by the characteristic amino acid sequence (EGCKG) in the P box located in the distal part of the first zinc finger (Fig. 3). They clearly represent related but distinct members of this class, since, in the DNA binding domain, CHR3 shows 52% amino acid identity to CNR8 and 64% identity to CNR14. Overall, the three deduced amino acid sequences show homologies to different genes. CHR3 shows 39% identity over 445 amino acids to the *Drosophila* steroid receptor homolog DHR3 (11) and 36% identity over 453 amino acids to the *Manduca sexta* homolog MHR3 (12). CNR8 shows a 32% identity over 419 amino acids with the human early response protein NAK1 (13). CNR14 is related to RAR-γ1 (14), having 29% identity over 352

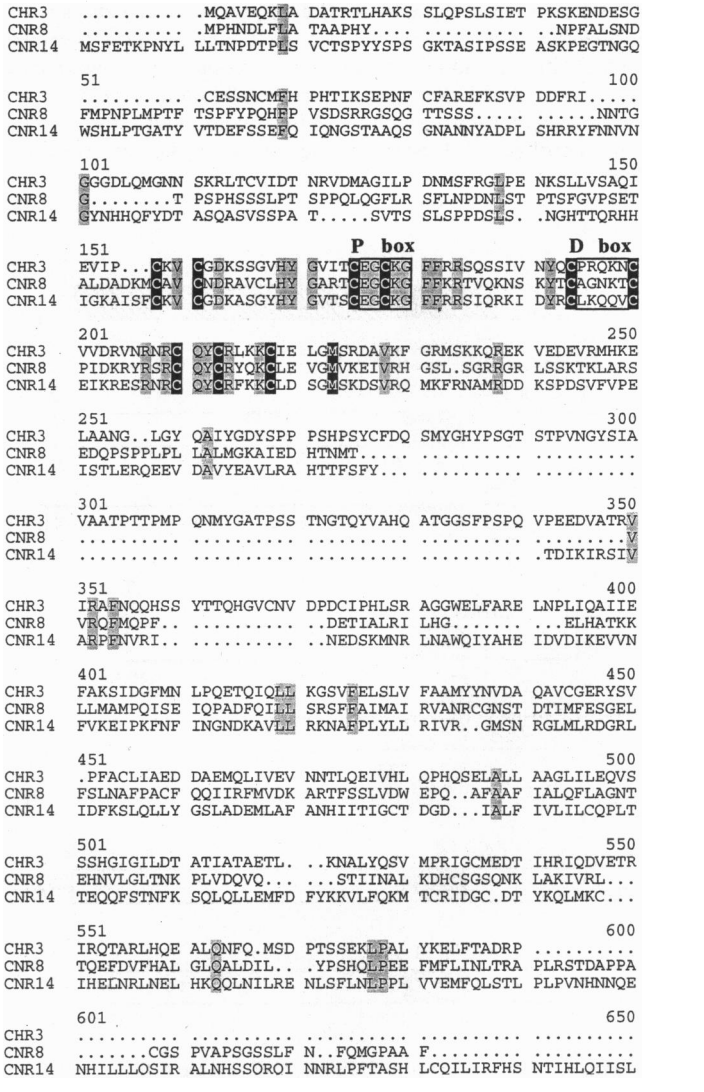

FIG. 2. Amino acid sequence alignment of *C. elegans* steroid hormone receptor family members CHR3, CNR8, and CNR14. Sequences were deduced from cDNAs and aligned by using the Genetics Computer Group program PILEUP (see *Materials and Methods*). Cysteines involved in zinc fingers and conserved methionine are shown as white symbols on a black background; other identical residues are shaded. P and D boxes are labeled.

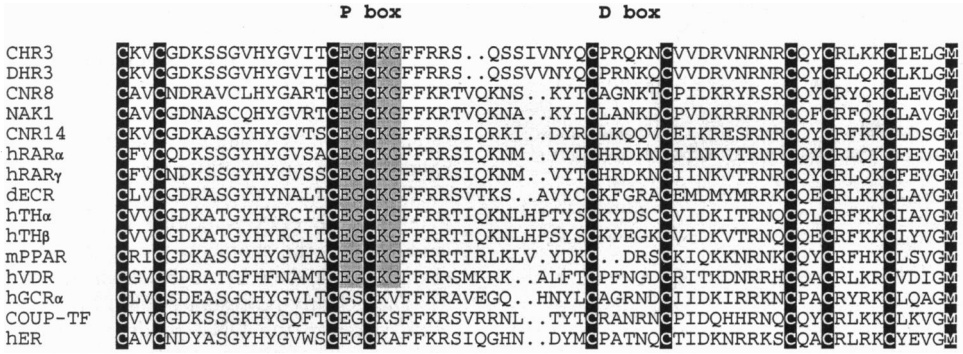

FIG. 3. Comparison of the C domain of various members of the STR family. P box sequence places the cDNAs we report here in the TR/RAR subfamily. The following sequences are shown: DHR3, *Drosophila* hormone receptor 3; NAK1, human early response gene; hRARα, human RAR-α; hRARγ, human RAR-γ; dECR, *Drosophila* ecdysone receptor; hTHα, human TR-α; hTHβ, human TR-β; mPPAR, mouse peroxisome proliferator-activated receptor; hVDR, human vitamin D receptor; hGCRα, human glucocorticoid receptor α; COUP-TF, human ovalbumin upstream promoter transcription factor; hER, human estrogen receptor. For review see Amero *et al.* (1) and Laudet *et al.* (2).

amino acids. An analysis using BLAST, however, shows greatest homology with the *Drosophila* ecdysone-inducible gene E78A (15). It is of interest that two of the three cDNAs we have isolated are closely related to ecdysone-inducible genes seen in arthropoda. Although small amounts of ecdysone and 20-hydroxyecdysone have been identified in *C. elegans* by some, a recent report failed to demonstrate synthesis of these compounds from labeled cholesterol (16, 17). Some of these relationships are shown in Fig. 4.

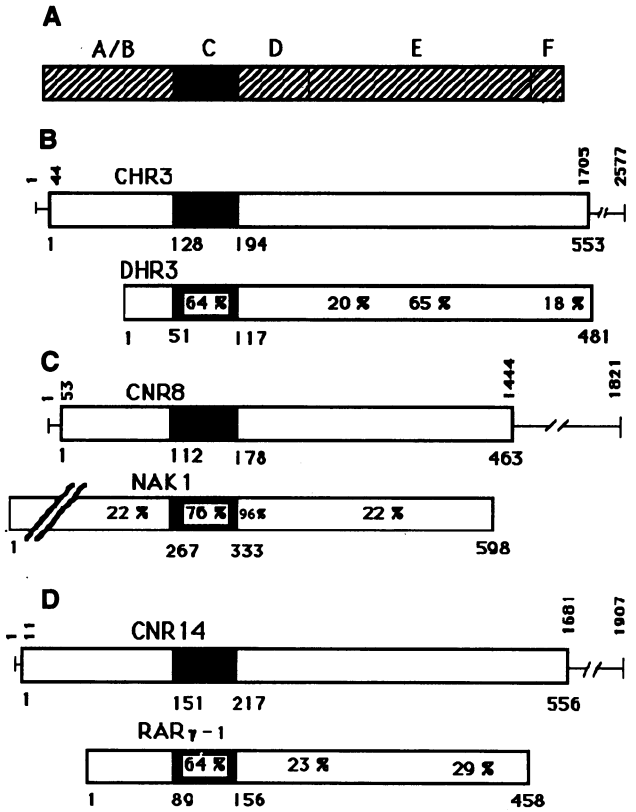

FIG. 4. Schematic representation of *C. elegans* STR family members with their closest homologues revealed by Fast A. (*A*) Domains A–F. C domain (DNA binding domain) is highlighted. Wide boxes represent open reading frames of cDNAs. Nucleotides are indicated by vertical numbers. Homology of particular regions is indicated by numbers inside boxes. Homology of <20% is not numerically indicated. Horizontally written numbers located under open reading frames represent amino acids. CHR3 is given in *B*, CNR8 is in *C*, and CNR14 is in *D*.

The A/B domain of these three cDNAs consisting of those nucleotides 5′ to the DNA binding domain, and that in general is not well conserved in the STR family, shows no obvious homology to any of the genes in the GenBank data base. The DNA binding domain (C) is homologous to all the genes of the STR family. Comparison of the D–E–F domain to all the sequences in GenBank showed that with CHR3, 35 genes possessed >20% amino acid identity over at least 100 amino acids. All of these genes belong to the STR family and the most similar gene is DHR3. A similar analysis of CNR8 disclosed 4 genes (NAK1, NGFI, Nur77, Nurr1), which were also all members of the STR family (13, 18–20). With CNR14, 5 of 7 genes displaying homology of >20% were from the STR family; one other was the human multidrug-resistance-associated protein (21) and the seventh was a heavy metal tolerance protein precursor from yeast (22). Hence, even in the ligand–dimerization–activation domain where one might expect relatively little in common, the sequences of those cDNAs we have cloned demonstrate clear similarities only to other members of this family.

The very C-terminal amino acids, which in most cases of known receptors are critical for ligand binding, show an 8-amino acid identity of CHR3 with both MHR3 and DHR3 in a sequence that we call the HR3 box (Fig. 5). This same region in human RZR-α, a human cDNA isolated from umbilical vein endothelium, and RZR-β from rat brain show a 7-amino acid identity in this region (23, 24). None of the other studied receptors contains more than 3 identical amino acids in this box. This suggests that these four genes are true homologues and that they probably all have very similar or identical ligands or possess an as yet unknown specific function.

The polyadenylylation signal appears to be represented by AATATA, AATTATA, and AATAAT in clones CHR3, CNR8, and CNR14, respectively, which occur 49, 8, and 104 bases from the poly(A) tail.

The Northern blot analyses show that all genes are expressed in a developmentally regulated way. CHR3 is most highly expressed in adult worms, although it can be seen in larval stages L3 and L4. Of the two transcripts revealed, the 2.7-kb

CHR3	PTSSEK.L**PALYKELF**TADRP
DHR3	P..NVV.F**PALYKELF**SIDSQ
MHR3	P..HHV.F**PALYKELF**SLDSV
RZRα	PDIVRLHF**PPLYKELF**TSEFE
RZRβ	PDIVNTLF**PPLYKELF**NPDCA

FIG. 5. Terminal portion of ligand binding domain of CHR3, DHR3, MHR3, hRZR-α, and rRZR-β and the HR3 box.

transcript was somewhat more highly expressed in the L3 stage and the 3.2-kb transcript showed much greater expression in adult worms (Fig. 6*A*). CNR8 shows greatest expression in larval stage L3 (Fig. 6*B*). Weak bands were also detectable in stage L4 and in adults in autoradiograms exposed for long periods of time (1 week). Only a single band of 3.1 kb was seen on Northern blots except in the L4 stage where a weak 2.1-kb band is visible. CNR14 shows greatest expression in embryos followed by somewhat less expression in the adult and weakest expression in other stages (Fig. 6*C*). CNR14 has a transcript of ≈2.3 kb. A rat probe for β-actin was used to control for RNA loading on the gels (data not shown).

These genes seem likely to be involved in regulation of development since we see these stage-specific differences in expression and since two of the genes (CHR3 and CNR14) are closely related to genes known to be critical for development in insects. The cDNAs that we cloned are of interest from an evolutionary standpoint since we know of no published data on this family of genes in invertebrates other than arthropoda and a single report of a COUP-like gene in an echinoderm (4). A search of the GenBank data base revealed a short sequence of CNR14 in a summary of partial cDNA sequences (ref. 25, accession no. Z14382). A computer search immediately before submitting these sequences to GenBank revealed that the CNR8 sequence was recently reported as an open reading frame on chromosome III of *C. elegans* (similar to a steroid hormone receptor, accession no. Z36237). Two other putative STR family members that have been reported as partial cDNA clones (ref. 25, accession no. M76122; unpublished data, accession no. T01703) are unrelated to the cDNAs we have sequenced. Thus, in *C. elegans*, at least five different members of the STR family are present. We find a multiplicity of STR genes in *C. elegans*, a member of the phylum Nematoda that existed before the divergence that led to the Chordata and the Arthropoda (26).

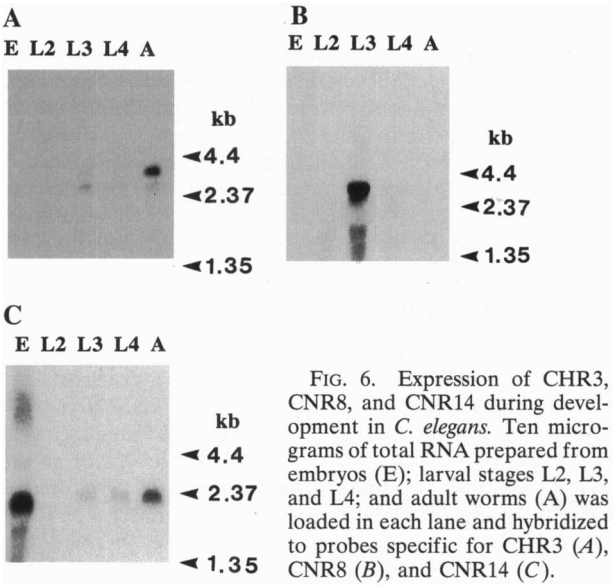

FIG. 6. Expression of CHR3, CNR8, and CNR14 during development in *C. elegans*. Ten micrograms of total RNA prepared from embryos (E); larval stages L2, L3, and L4; and adult worms (A) was loaded in each lane and hybridized to probes specific for CHR3 (*A*), CNR8 (*B*), and CNR14 (*C*).

We are grateful to Dr. Vera Nikodem for critical advice throughout the course of this work, to Drs. Joram Piatigorsky and John Wootton for advice on the manuscript, and to the *Caenorhabditis* Genetics Center, which is funded by the National Institutes of Health National Center for Research Resources for supplying us with *C. elegans* worms.

1. Amero, S. A., Kretsinger, R. H., Moncrief, N. D., Yamamoto, K. R. & Pearson, W. R. (1992) *Mol. Endocrinol.* **6,** 3–7.
2. Laudet, V., Hanni, C., Coll, J., Catzeflis, F. & Stehelin, D. (1992) *EMBO J.* **11,** 1003–1013.
3. Forman, B. M. & Samuels, H. H. (1990) *Mol. Endocrinol.* **4,** 1293–1301.
4. Chan, S.-M., Xu, N., Niemeyer, C. C., Bone, J. R. & Flytzanis, C. N. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10568–10572.
5. Freedman, L. P. (1992) *Endocr. Rev.* **13,** 129–145.
6. Green, S., Kumar, V., Theulaz, I., Wahli, W. & Chambon, P. (1988) *EMBO J.* **7,** 3037–3044.
7. Truss, M. & Beato, M. (1993) *Endocr. Rev.* **14,** 459–479.
8. Sulston, J. & Hodgkin, J. (1988) in *The Nematode Caenorhabditis elegans*, ed. Wood, W. B. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 587–606.
9. Krause, M. W. (1995) *Methods Cell Biol.* **48,** in press.
10. Krause, M. W. & Hirsh, D. (1987) *Cell* **49,** 753–761.
11. Koelle, M. R., Segraves, W. A. & Hogness, D. S. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 6167–6171.
12. Palli, S. R., Hiruma, K. & Riddiford, L. M. (1992) *Dev. Biol.* **150,** 306–318.
13. Nakai, A., Kartha, S., Sakurai, A., Toback, F. G. & DeGroot, L. J. (1990) *Mol. Endocrinol.* **4,** 1438–1443.
14. Krust, A., Kastner, P., Petkovich, M., Zelent, A. & Chambon, P. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5310–5314.
15. Stone, B. L. & Thummel, C. S. (1993) *Cell* **75,** 307–320.
16. Mercer, J. G., Munn, A. E. & Rees, H. H. (1988) *Comp. Biochem. Biophys. B Comp. Biochem.* **90,** 261–267.
17. Chitwood, D. J. & Feldlaufer, M. F. (1990) *J. Nemat.* **22,** 598–607.
18. Milbrandt, J. (1988) *Neuron* **1,** 183–188.
19. Davis, I. J., Hazel, T. G. & Lau, L. F. (1991) *Mol. Endocrinol.* **5,** 854–859.
20. Law, S. W., Conneely, O. M., DeMayo, F. J. & O'Malley, B. W. (1992) *Mol. Endocrinol.* **6,** 2129–2135.
21. Cole, S. P. C., Bhardwaj, G., Gerlach, J. H., Mackie, J. E., Grant, C. E., Almquist, K. C., Steward, A. J., Kurz, E. U., Duncan, A. M. V. & Deeley, R. G. (1992) *Science* **258,** 1650–1654.
22. Ortiz, D. F., Kreppel, L., Speiser, D. M., Scheel, G., McDonald, G. & Ow, D. W. (1992) *EMBO J.* **11,** 3491–3499.
23. Becker-Andre, M., Andre, E. & DeLamarter, J. F. (1993) *Biochem. Biophys. Res. Commun.* **194,** 1371–1379.
24. Carlberg, C., van Huijsduijnen, R. H., Staple, J. K., DeLamarter, J. F. & Becker-Andre, M. (1994) *Mol. Endocrinol.* **8,** 757–770.
25. Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J. & Sulston, J. (1992) *Nat. Genet.* **1,** 114–123.
26. Erwin, D. H. (1991) *Trends Ecol. Evol.* **6,** 131–134.