A steroid/thyroid hormone receptor superfamily member in *Drosophila melanogaster* that shares extensive sequence similarity with a mammalian homologue

By: Vincent C.Henrich, Timothy J.Sliter, Dennis B. Lubahn, Alanna MacIntyre and Lawrence I. Gilbert

Henrich, V.C., T.J. Sliter, D.L. Lubahn, A. MacIntyre, and L.I. Gilbert (1990) A steroid/thyroid hormone receptor superfamily in *Drosophila melanogaster* that shares extensive sequence similarity with a mammalian counterpart. *Nucleic Acids Research,* 18(14): 4143-4148. DOI: 10.1093/nar/18.14.4143

Made available courtesy of Oxford University Press: http://dx.doi.org/10.1093/nar/18.14.4143

Abstract:

A gene in *Drosophila melanogaster* that maps cytologlcally to 2C1 — 3 on the distal portion of the X-chromosome encodes a member of the steroid/thyrold hormone receptor superfamily. The gene was isolated from an embryonic cDNA library using an oligonucleotide probe that specifies the consensus amino acid sequence in the DNA-binding domain of several human receptors. The conceptual amino acld sequence of 2C reveals at least four regions of homology that are shared with all identified vertebrate receptors. Reglon I includes the two cysteine-cysteine zinc fingers that comprlse a DNA-bindlng domaln which typifies all members of the superfamily. ln addition, three regions (Regions lI-IV) in the carboxy-termlnal portion of the protein that encode the putative hormone-binding domain of the 2C gene product resemble similar sequences ln vertebrate steroid/thyroid hormone receptors. The similarity suggests that this *Drosophila* receptor possesses many of the regulatory functions attributed to these regions in vertebrate counterparts. A portion of Region II also resembles part of the human c-jun oncoprotein's leucine zipper, which in turn, has been demonstrated to be the heterodimerlzation site between the *jun* and *fos* oncoproteins. The 2C receptor-llke protein most resembles the mouse H2RII binding protein, a member of the superfamily which has been implicated in the regulatlon of major histocompatibility complex (MHC) class I gene expression. These two gene products are 83% ldentical in the DNA-binding domain and 50% identical in the putative hormone-binding domain, although no ligand has been identified for either protein. The high degree of simllarity in the hormone-blnding domaln between the 2C protein and the H2RII binding protein outside regions II-IV suggests specific functional roles which are not shared by other members of the superfamily.

Article:

INTRODUCTION

Steroid hormones act on target cells by forming a complex with an intracellular receptor, that in turn, recognizes specific target DNA sequences and regulates gene expression (1). The members of the steroid/thyroid hormone receptor superfamily, which include receptors for several non-steroid hormones as well as receptor-like proteins with no known ligand, share a common 66-68 amino acid sequence comprised of two cysteine-rich zinc fingers that are necessary and sufficient for sequence recognition (2). Five members of the superfamily have been identified in *Drosophila melanogaster*, including three members of the knirps (*kni*) gene family associated with embryonic development (*kni* (3), knirps-related, *knrl* (4), and embryonic gonad, *egon* (5)), another which maps to 75B and whose transcription is induced by 20-hydroxyecdysone in the larval salivary gland (6, 7), and seven-up (*svp*), which participates in the developmental regulation of adult eye structures and strongly resembles the human COUP-transcription factor (8).

Both *svp* and 75E contain three sequences within their hormone-binding domain that show extensive similarity to vertebrate members of the receptor superfamily, while the knirps family members do not. Numerous functions besides hormone-binding, including transcription factor interactions (9) and the formation of inactive complexes with cytosol proteins (10) have been attributed to these conserved portions of the domain in various

receptors. In the case of the human retinoic and thyroid hormone receptors, one of the regions contains a heterodimerization site that resembles a leucine zipper motif (11), and in fact, other regions within the domain also resemble leucine zippers (12). This entity was originally discovered in the *jun* and *fos* nuclear oncoproteins and functions as a heterodimerization site between them by allowing the interdigitation of regularly spaced, protruding leucine side chains of an alpha-helix (13). The structural constraints of this motif remain largely unknown since substitution of single leucines within the zipper do not appreciably disrupt dimerization or gene induction, although the seven amino acid interval between leucine residues and a continuous alpha helical configuration are critically important to maintain the structure necessary to form a zipper (14, 15).

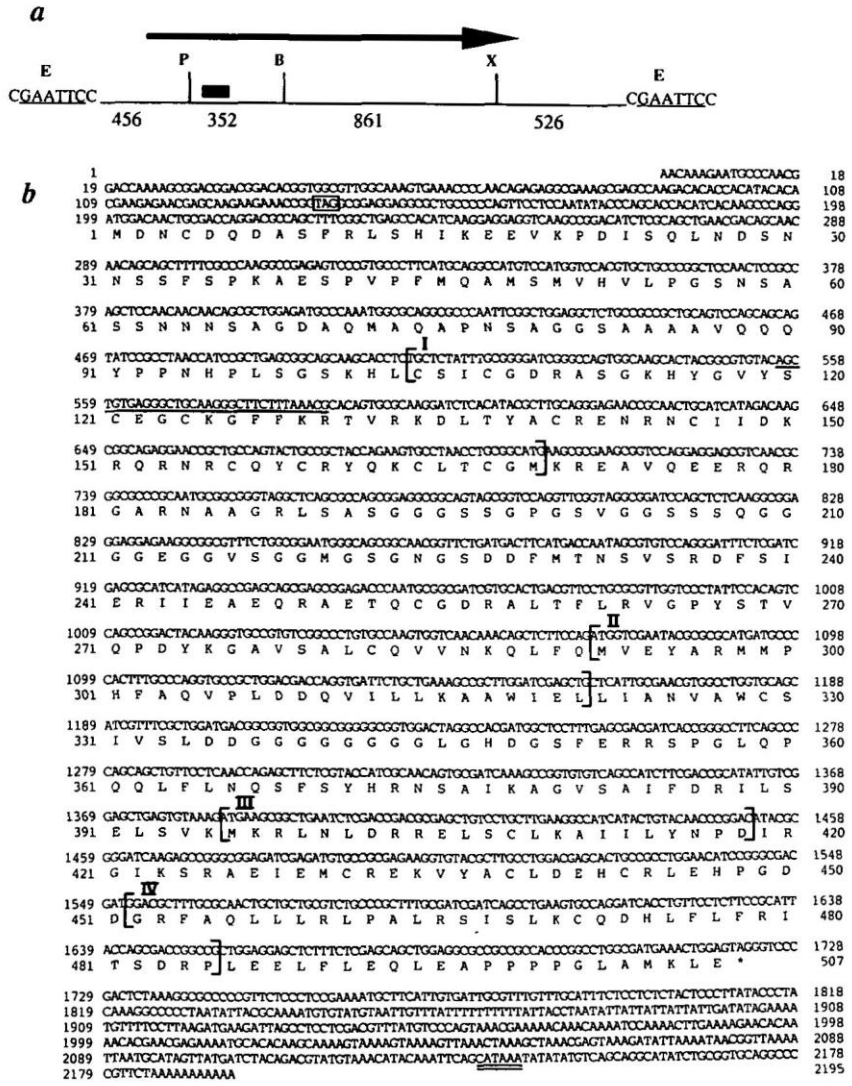

Figure 1. Structural features of the longest 2C cDNA in *Drosophila melanogaster*. (a) A restriction map of the longest 2C cDNA clone. The arrow designates the extent and direction of the longest open reading frame of the 2C gene, present in all three cDNAs whose sequence was analyzed. The small block indicates the general location of the nucleotide sequence recognized by the oligonucleotide probe. Letters designate selected internal restriction cleavage sites: P (PstI), B (BamH1), and X (XhoI). Flanking nucleotides indicate the flanking EcoR1 linkers and restriction sites (E). Number of nucleotide pairs between these sites is denoted below the restriction map. (b) The complete nucleotide and conceptualized amino acid sequence of the longest 2C cDNA. Dashed underlines denote the start and stop codons of a short open reading frame on the most 5′ portion of the cDNA. The boxed codon designates a stop codon that is in the same frame as the later start codon. A double underline indicates a polyadenylation site. The bracketed regions designate those domains compared with other receptors including the DNA-binding domain (I, Fig. 3b), and three regions in the hormone-binding domain (II, Fig. 4b), (III, Fig. 4c), and (IV, Fig. 4d). Nucleotides underscored with a solid line denote the nucleotides recognized by the 32-base oligonucleotide probe.

This study involves the characterization of a member of the steroid/thyroid hormone receptor superfamily in *Drosophila*, the 2C gene product. This receptor-like protein contains regions of extensive sequence similarity with vertebrate counterparts in both the DNA-binding and hormone-binding domains. In addition, the 2C gene product shares extensive similarity in generally nonconserved regions with the mouse transcription factor,

H2RII binding protein, which has been implicated in the regulation of expression of the major histocompatibility (MHC) complex class I gene (16).

MATERIALS AND METHODS

Isolation and sequencing

A ^{32}P, end-labelled 32 base oligonucleotide probe (5'- ACCTGTGAGGGCTGTAAGGTCTTCTTCAAAAG-3') was employed to screen an amplified lambda ZAP cDNA library (Stratagene; LaJolla, CA) prepared from poly(A)$^{+}$ RNA which had been extracted from *Drosophila melanogaster* 0-24 hour, mixed stage embryos belonging to a wild-type Canton-S strain. The cDNA inserts had been ligated to EcoRI linkers and cloned into the appropriate polylinker site. Seven positive clones were isolated from approximately 100,000 which were screened at reduced stringency (hybridization at 50°C in 6 x SSC followed by filter washes at 50°C in 2 x SSC) and four were characterized further. The cDNAs were excised *in vivo* into a Bluescript SKplasmid and subjected to preliminary restriction mapping. The dideoxy-chain termination reaction was employed for sequencing double-stranded DNA (Sequenase 2.0 kit; US Biochemical; Cleveland, OH) of three of the cDNA clones. Both strands of two entire cDNAs were sequenced with both ddGTP and ddITP, using either internal oligonucleotide primers or Bluescript primers with subclones of the original cDNAs. Reactions were labelled with 35S-ATP (Amersham; Arlington Heights, IL) and separated on a 6% polyacrylamide gel. The sequence and peptide structure were analyzed with Genetics Computer Group (GCG; Madison, WI) software on a Digital VAX system (17).

In situ hybridizations

A double-stranded cDNA insert (described in Fig. 1) was obtained by digesting the cDNA-containing recombinant Bluescript plasmid with EcoR1 (US Biochemical) and extracting the insert (GeneClean kit; Bio101; LaJolla, CA) following its separation by agarose gel electrophoresis in a Tris-acetate buffer. Digoxigenin-labelled probe DNA was synthesized (BoehringerMannheim; Mannheim, FRG) by the random priming method. Squash preparations of polytene chromosomes from salivary glands of wild-type *D. melanogaster* larvae (Samarkand BG strain) were prepared using standard procedures (18). After heat denaturation and acetylation, the chromosome preparations were hybridized with the digoxigenin-labelled DNA probe (30 pg/ul) at 58°C overnight. Following post-hybridization washes in 2 x SSC at 58°C, then 1 x SSC and 0.5x SSC at room temperature, the chromosomal site of probe hybridization was visualized immunohistochemically using an anti-digoxigeninalkaline phosphatase conjugated antibody (Boehringer-Mannheim). Chromosome preparations were mounted in water and photographed at 200-1000 x with phase contrast optics.

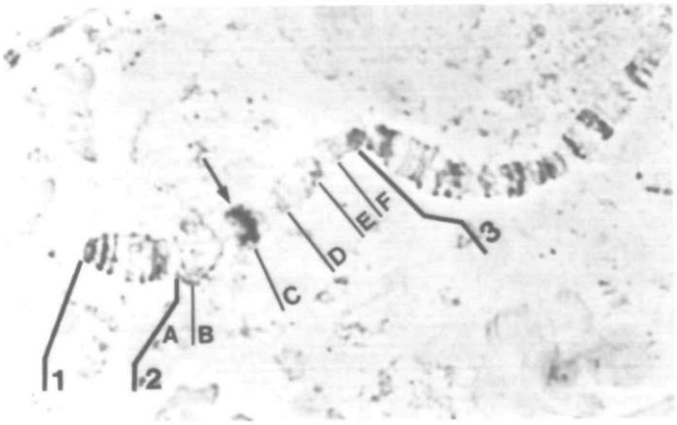

Figure 2. In situ hybridization of a 2C cDNA clone with polytene chromosomes in the salivary gland dissected from prewandering third instar larvae of *Drosophila melanogaster*.

RESULTS AND DISCUSSION

Isolation and characterization of cDNA clones

The oligonucleotide probe previously utilized to isolate the human androgen receptor gene was deduced from the consensus amino acids of the carboxy-terminal portion of the first zinc finger and the adjoining linker region of several previously cloned vertebrate receptors (19; Fig. 1). All of the four embryonic cDNA clones isolated at reduced stringency with this probe represented a single gene, based on restriction mapping patterns (Fig. la). The longest cDNA contained 2195 nucleotides, including a sequence of 169 nucleotides on the 5' end not seen in the other cDNA clones (Fig. lb). A second cDNA contained a different nucleotide sequence beginning at position 2186 and extending 113 nucleotides to an EcoR1 linker on the 3' end (sequence not shown). This variable region began with a GAAT in place of the poly(A)+ tail seen in the other cDNAs and therefore may be a damaged EcoR1 site. This was the only observed discrepancy among overlapping portions of the three cDNAs subjected to sequence analysis, and all contained an identical and complete open reading frame specifying a peptide of 507 amino acids with a predicted molecular weight of 55,245 Da. The ATG start codon that begins the longest open reading frame is preceded on the 5' side by an in-frame stop codon at nucleotide position 136 (Fig. lb). No AATAAA polyadenylation site was detected in any cDNA, although a more cryptic CATAAA polyadenylation site was found (2137; Fig lb).

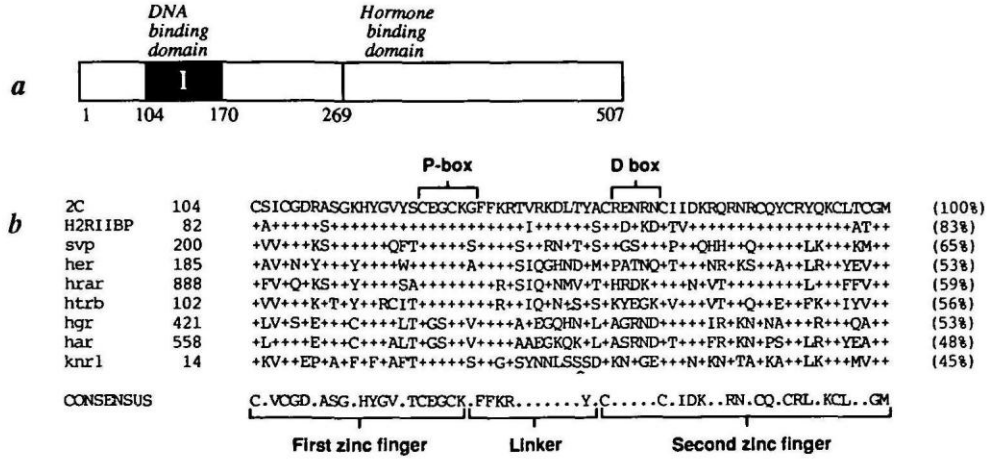

Figure 3. Comparison of the deduced amino acid sequence of the DNA binding domain between 2C and other members of the steroid hormone receptor superfamily. (a) Location of the DNA-binding domain (Region I) within the deduced amino acid sequence of the 2C gene product. Numbers denote amino acid positions that initiate each domain. Residues 1−103 denote a variable domain; 104−169 denote the DNA-binding domain; 170−268 denote a linker domain; 268−507 denote the hormone-binding domain (1). (b) Amino acid sequence comparison and structural features in the DNA-binding domain (I) of representative members of the steroid hormone receptor superfamily. Numbers at left designate the location of amino acid residues that begin the compared sequence. The receptors compared are: H2RIIBP, mouse H2RII binding protein (16); svp, *Drosophila* seven-up gene product (8); her, human estrogen receptor (24, 25); htrb, human thyroid hormone receptor (26); hgr, human glucocorticoid receptor (27); har, human androgen receptor (28); and knrl, *Drosophila* knirps-related gene product (4). Plus signs (+) designate amino acids identical to the corresponding 2C residue. Dashed lines indicate a gap in the aligned sequence (−). Dots in consensus sequence indicate no consensus amino acid residue at that position. Hats (^) note amino acids positions that formed a gap in the 2C sequence and which are not included here. In *knrl* the gap consists of a single amino acid; in htrb, the gap is two amino acids. Numbers in parentheses designate the percentage of identical residues between the 2C gene product and each member of the receptor superfamily.

Cytological mapping

Based on chromosomal *in situ* hybridizations, the gene was localized to the distal portion of the X-chromosome within the 2C1-3 cytological interval (20; Fig. 2). No signal was detected at any other chromosomal sites including those associated with other *Drosophila* members of the steroid/thyroid hormone receptor superfamily.

Sequence comparison

Based on the deduced amino acid sequence, the 2C gene product displays the same organization ascribed to other members of the steroid/thyroid hormone receptor superfamily. It includes an amino-terminal portion of 103 amino acids followed by the 66 amino acid cysteine-rich domain that defines members of the receptor superfamily and that contains the two zinc fingers responsible for DNA-binding (Fig. 3a). The 2C gene also encodes a carboxy-terminal domain that includes several regions showing similarity with analogous sequences which comprise the hormone-binding domain in vertebrate receptors (Fig. 4a). Based on this conservation, this

region will be referred to as the 2C hormone-binding domain here, although the hormone ligand associated with this receptor-like protein is not known. Overall, the 2C gene product most resembles the mouse H2RII binding protein based on a gap analysis, sharing 70% similarity including conserved amino acid substitutions, and 49 % identity (Fig. 5).

Within the DNA-binding domain, the similarity (including conserved substitutions) between 2C and H2RII binding protein is 90%, with 83% of the amino acids being identical. Two sequences in the DNA-binding domain, the proximal (P) box and the distal (D) box (Fig. 3), are responsible for the recognition of specific target sequences and have been used to categorize members of the superfamily into at least two subfamilies (21). The P box of 2C and H2RII binding protein are identical, and place both in the estrogen receptor subfamily. Furthermore, every amino acid in the D-box of 2C is identical or a conserved substitution of the concomitant residues in the H2RII binding protein. Nevertheless, the D-box of 2C shares even more identities with svp and its human counterpart, the COUP-transcription factor.

We also compared the putative 2C hormone binding domain that begins at amino acid 269 with the corresponding carboxyterminal domain of other superfamily members in order to assess whether it possesses sequences that resemble regions in some vertebrate receptors to which specific regulatory functions have been attributed. This analysis revealed at least three intervals within the 2C hormone-binding domain that show significant similarity with other members of the superfamily. Region II (residues 292-321) resembles a portion of the glucocorticoid receptor that has been implicated in the formation of inactive heat shock protein complexes in the cytosol that dissociates in the presence of the receptor's cognate ligand (10; Fig. 4b). The carboxy-terminal portion of this subinterval includes three leucines spaced apart by seven amino acids, suggesting a leucine zipper motif similar to those observed in oncoproteins and several other receptors (11-13). Among these fifteen amino acids, the 2C protein resembles a portion of the *jun* oncoprotein leucine zipper to about the same extent that it resembles other members of the receptor superfamily (approximately 40-60% identical; Fig. 4b) with the exception of H2RII binding protein (with which 2C shares 80% identity in this region). Protein structure models utilizing Garnier-Osguthorpe-Robson (GOR) algorithms predict an alpha-helical structure in this region, as required for the formation of a leucine zipper (data not shown). A previously reported region III includes several consensus amino acids to which no specific function has been ascribed (22; Fig. 4c). Region IV corresponds to a region shown to be a heterodimerization site between the thyroid hormone and retinoic acid (but not other) receptors (11; Fig. 41). Whereas the derived 2C sequence shows some similarities in all of these intervals with other superfamily members, it consistently resembles the analogous H2RII binding protein sequence to the greatest extent, with residue identities ranging from 56% to 71 % between the two proteins within these regions. Moreover, the conceptual hormone-binding domains of the 2C and H2RII binding proteins share numerous amino acid similarities outside the generally conserved regions. In fact, the major difference between them appears to be the presence of two intervals in 2C which are not found in H2RII binding protein.

The 2C gene product is the sixth member of the steroid/thyroid hormone receptor superfamily reported in *Drosophila melanogaster* and the second which shows a particularly strong resemblance to a specific mammalian counterpart. The similarity between 2C and the mouse H2RII binding protein and svp and the human COUP-transcription factor involves extensive sequence identities in both the DNA-binding and hormone-binding domains, including portions that are not generally conserved among superfamily members. The Drosophila E75 member shares unusually high similarity with the human earl receptor but only within the DNA-binding domain (7). All of the receptor superfamily members, including these, share less similarity with each other in the amino-terminal domain and in the linker region that connects the DNA-binding and hormone-binding domains.

The degree of similarity is greatest between 2C and the H21211 binding protein in the DNA-binding domain and strongly suggests that both of these proteins recognize similar target sequences to influence gene expression, including the RII promoter site that was utilized to isolate and identify the H2RII binding protein. It is less clear whether the high similarity between 2C and H2RII binding protein in the three conserved regions

that reside in the hormone-binding domain indicates selective pressure for specific molecular interactions shared by 2C and the H2RII binding protein.

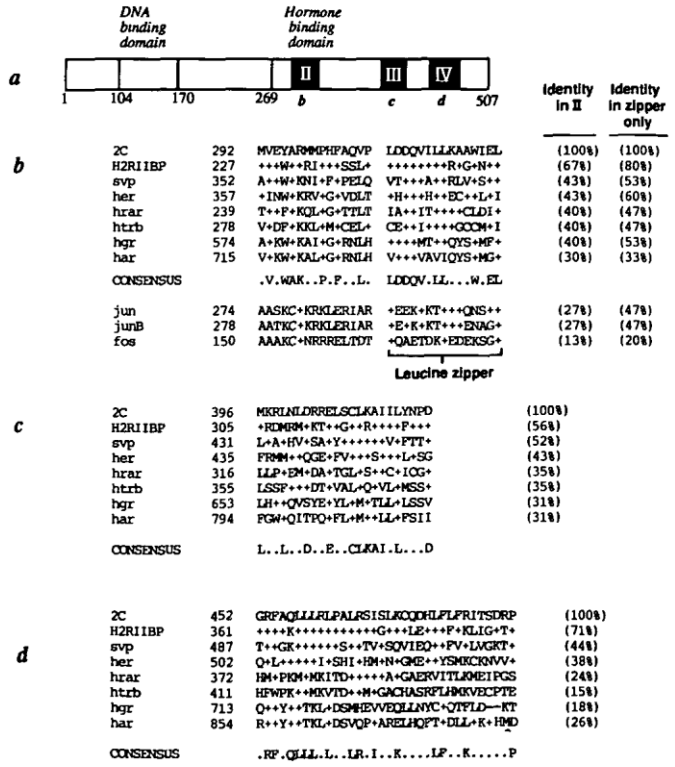

Figure 4. Comparison of the deduced amino acid sequence in the putative hormone binding domain of several members of the steroid hormone receptor superfamily. (a) Location of conserved regions in the hormone-binding domain of the 2C gene product. Numbers designate amino acid residues that initiate each domain. (b) Sequence comparison of a conserved region implicated in the formation of inactive cytosolic complexes in the human glucocorticoid receptor. Comparison includes those receptors and notations given in Figure 3 and jun, the human *c-jun* oncoprotein (29); junB, the mouse *junB* oncoprotein (30) ; fos, the human fos oncoprotein (31). No significant similarity was observed between 2C and *kni*, *knrl*, or *egon* in these regions and their sequence is not included. The numbers in parentheses designate the percentage of identical residues between the 2C gene product and each superfamily member or oncoprotein in the carboxy-terminal portion (15 residues) of this domain. Numbers at left designate location of amino acid residues that begin the compared sequence. (c) Sequence comparison between 2C and other receptors in a second conserved region for which no function has been attributed. (d) Sequence comparison of a conserved region shown to function as a dimerization site between the human retinoic acid and thyroid hormone receptors. The hat (^) designates three amino acids in har that formed a gap when compared to the 2C sequence and which are not included here.

Within region II, the 2C and H2RII proteins show as much similarity to a portion of the *jun* oncoprotein leucine zipper as they do to other members of the superfamily. Furthermore, the predicted secondary structure of this region in 2C and other receptors is alpha-helical, as required for a leucine zipper motif. The functional significance of this structural observation remains to be determined although it can be inferred that this motif may allow dimerization with the fos oncoprotein. Interestingly, the expression of *fos* oncoprotein has been correlated with the regulation of MHC gene expression (which involves H2RII binding protein), although no specific molecular mechanism has been proposed to explain this possible connection (23).

The identity between 2C and H2RII binding protein within the hormone-binding domain extends beyond the generic similarities in the noted conserved domains, and presumably indicates more specialized functions that are uniquely shared between these two receptor-like proteins. The most obvious explanation for these unique similarities is that 2C and H2RII binding protein interact with an identical or similar hormone ligand. Additionally or alternatively, these unique identities between the two proteins may indicate other specialized

forms of trans-regulation in the hormone binding domain which remain unidentified and unique to these two receptor-like proteins. The hormone ligand has not been identified for either 2C or H2RII binding protein, nor for any of the reported *Drosophila* receptor superfamily members.

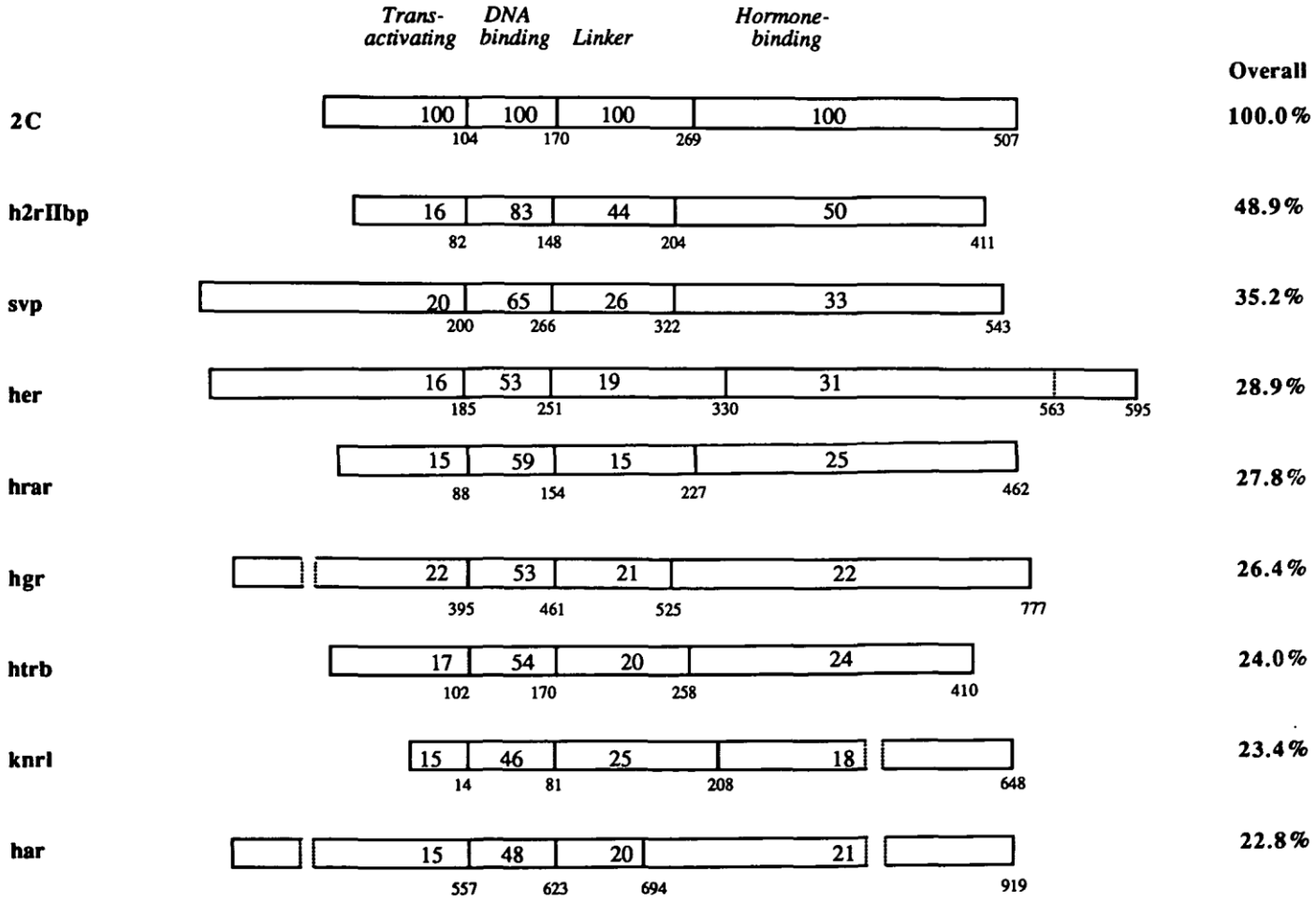

Figure 5. A comparison of individual domains within the 2C gene product with those of representative steroid/thyroid hormone receptor superfamily members. Percentages designate amino acid identities based on a gap analysis between equivalent domains of 2C and individual receptors. Numbers designate the amino acids that initiate the domain subjected to gap analysis. Abbreviations are the same as noted previously.

As more members of the superfamily are identified in *Drosophila*, it becomes increasingly apparent that many hormones, perhaps including some which also operate in vertebrates, may play an important role in insect development. Both of the major insect hormones, 20-hydroxyecdysone and juvenile hormone, presumably act on cellular targets via a receptor belonging to this superfamily.

The functional and interactive roles described here for the 2C gene product can now be assessed with both biochemical and genetic approaches. Moreover, with the tools available in *Drosophila*, it will be possible to investigate the developmental consequences of structural mutations in the 2C gene.

Acknowledgments:

We thank Frank S. French, Elizabeth M. Wilson, and David R. Joseph from the Laboratories for Reproductive Biology for providing us with the oligonucleotide probe and Lillie Searles for the cDNA library used in this study, Susan Whitfield for assistance with graphics, and David Richard and Al Baldwin for helpful discussions. The 2C gene has also been cloned by Martin Shea and Fotis Kafatos and Anthony Oro and Ronald Evans and we thank them for providing us with their unpublished data. D.B.L. is a Pew Scholar supported by the Pew Scholars Program in the Biomedical Sciences. This work was supported by NIH Grant DK 35347 and MERIT award DK30118 from the National Institutes of Health.

References:

1. L Green, S., and P. Chambon (1988) Trends in Genet. 4, 309-313.
2. Evans, R. (1988) Science 240, 889-895.
3. Nauber, U., M.J. Pankratz, A. Kienlin, E. Seifert, U. Klemm, and H. Jackie (1988) Nature 336, 489-492.
4. Oro, A.E., E.S. Ong, J.S. Margolis, J.W. Posakony, M. McKeown, and R.M. Evans (1988) Nature 336, 493-496.
5. Rothe, M., U. Nauber, and H. Jackie (1989) EMBO J. 8, 3087-3094.
6. Feigl, G., M. Gram, and 0. Pongs (1989) Nuc. Acids Res. 17, 7167-7178.
7. Segraves, W., and D. Hogness (1990) Genes and Devel. 4, 204-219.
8. Mlodzik, M., Y. Hiromi, U. Weber, C.S. Goodman, and G.M. Rubin (1990) Cett 60, 211-224.
9. Beato, M. (1989) Cell 56, 335-344.
10. Pratt,W.B., D. Jolly, D. Pratt, S.M. Hollenberg, V. Giguere, F.M. Cadepond, G. Schweizer-Groyer, M-G. Catelli, R.M. Evans, and E.-E. Baulieu (1988) J. Biol. Chem. 263. 267-273.
11. *Glass*, C.K., S.M. Lipkin, 0.V. Devary, and M.G. Rosenfeld (1989) Cell 59, 697-708.
12. Forman,B.M., C.-r. Yang, M. Au, J. Casanova, J. Ghysdael, and H.H. Samuels (1989) Mol. Endocrinol. 3, 1610-1626.
13. Landschulz, W.H., P.F. Johnson, and S.L. McKnight (1988) Science 240, 1759-1764.
14. Kouzarides, T., and E. Ziff (1988) Nature 336, 646-65L
15. Schuermann, M., M. Neuberg, J.B. Hunter, T. Jenowein, R-P. Ryseck, R. Bravo, and R. Muller (1989) Cell 56, 507-516.
16. Hamada, K., S.L. Gleason, B-Z Levi, S. Hirschfeld, E. Appella, and K. Ozato (1989) Proc. Natl. Acad. Sci. USA 86, 8289-8293.
17. Devereux, J., P. Haeberli, and 0. Smithies (1984) Nuc. Acids. Res. 12, 387 -395.
18. Pardue, M.L. (1986) Drosophila: A Practical Approach, Ed. by D.B. Roberts; 1RL Press, Washington, D.C., pp. 111-138.
19. Lubahn, D., D.R. Joseph, P.M. Sullivan, J.F. Willard, F.S. French, and E.M. Witson (1988) Science 240, 327-330.
20. Sorsa, V. (1988) Chromosome MapS of Drosophita, Volume 1I. CRC Press. Boca Raton, FL.
21. Umesono, K., and R.M. Evans (1989) Cell 57, 1139-1147.
22. Wang, L-H, S.Y. Tsai, R.G. Cook, W.G. Beanie, M-1 Tsai, and B.W. O'Malley (1989) Nature 340, 163-166.
23. Feldmann, M., and L. Eisenbach (1988) Sci. Amer. 259, 60-68.
24. Greene, G, P. Gilna, M. Waterfield, A. Baker, Y. Hort, and J. Shine (1986) Science 231, 1150-1154.
25. Green, S, P. Walter, V. Kumar, A. Must, 1.-M. Bornert, P. ArgoS and P. Chambon (1986) Nature 320, 134 -139.
26. Weinberger, C, C. Thompson, E. Ong, R. Lebo, D. Gruol, and R.M EvanS (1986) Nature 324, 641-646.
27. Hollenberg, S., C. Weinberg, E. Ong, G. Cerelli, A. Oro, R. Lebo, E.B. Thompson, M. Rosenfeld, and R.M. Evans (1985) Nature 318, 635-641
28. Lubahn, D.L., D.R. JoSeph, M. Sar, J.A. Tan, H.N. Higgs, R.E. Larson, F.S. French, and E.M. Wilson (1988) Mol. Endocrinol. 2, 1265-1275.
29. Bohmann, D., T.J. Bos, A. Admon, T. Nishinura, R.K Vogt, and R. Tjian (1988) Science 238, 1386-1392.
30. Ryder, K., L.F. Lau, and D. Nathan (1988) Proc. Natl. Acad. Sci. USA 85, 1487-1491.
31. VanStraaten, F., R. Muller, T. Curran, C. van Bevern, and I.M. Verna (1983) Proc. Natl. Acad Sci. USA 80, 3183-3187.