MCN

Molecular and Cellular Neuroscience **20,** 404–414 (2002)
doi:10.1006/mcne.2002.1138

Cloning of a Novel *Olf-1/EBF*-like Gene, *O/E-4,* by Degenerate Oligo-based Direct Selection

Song S. Wang, Alexander G. Betz,[1] and Randall R. Reed[2]

Department of Molecular Biology and Genetics and Department of Neuroscience, The Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, PCTB 818 Baltimore, Maryland 21205

The Olf-1/EBF (O/E) family of transcription factors plays important roles in neural development and B-cell maturation. We have utilized a variation of the direct selection scheme (Degenerate Oligo-based Direct Selection, or DODS) to identify a fourth member of the O/E family (O/E-4). We demonstrated that, similar to previously identified mouse O/E proteins, O/E-4 is expressed in the neuronal and basal cell layers of mouse olfactory epithelium. However, unlike other O/E members, O/E-4 expression is absent in the vomeronasal organ. O/E-4 can partner with the other O/Es to form heterodimers that are capable of binding an Olf-1 consensus site *in vitro*. In contrast to other O/E family members, the O/E-4 protein only weakly activates transcription in a reporter assay. In summary, we have identified O/E-4 utilizing a novel homology screening method. Our results suggest that O/E-4 may interact with other O/E family members to regulate gene expression in the olfactory sensory neurons.

Key Words: **Olf-1; EBF; olfactory; VNO; transcription; direct selection; homology screening.**

INTRODUCTION

The Olf-1/EBF (O/E) transcription factors are a conserved family of helix–loop–helix (HLH) proteins present in vertebrate and invertebrate species. The first member of the O/E family, O/E-1, was identified as a protein that binds the promoters of various olfactory-enriched genes and the *mb-1* gene expressed in developing B-lymphocytes (Hagman *et al.*, 1993; Wang and Reed, 1993). Subsequently, additional family members were described in mouse (O/E-2 and 3), *C. elegans* (UNC-3), *Drosophila (Collier), Xenopus* (xCOEs), and Zebrafish (zCOEs) (Dubois and Vincent, 2001; Bally-Cuif *et al.*, 1998; Dubois *et al.*, 1998; Prasad *et al.*, 1998; Garel *et al.*, 1997; Wang *et al.*, 1997; Crozatier *et al.*, 1996). The vertebrate O/E proteins contain an unique repeated HLH motif (rHLH) not found in the invertebrate proteins, and a conserved alternative HLH domain consisting of the first helix of the rHLH motif and a second helix further upstream has been proposed. Studies in *C. elegans* indicated that the UNC-3 protein regulates aspects of motor neuron projection and sensory neuron properties, and mis-expression studies on the xCOE proteins suggest that they are involved in early neural patterning (Pozzoli *et al.*, 2001; Dubois *et al.*, 1998; Prasad *et al.*, 1998).

The high sequence similarity and overlapping patterns of expression of O/E proteins have complicated efforts to elucidate function in the mouse. For instance, a genetic disruption of the *EBF* gene results in mice that lack mature B-cells, but the olfactory system is morphologically normal and expresses olfactory-enriched genes thought to be regulated by O/E proteins (Lin and Grosschedl, 1995). The expression of O/E-2 and O/E-3 in a pattern which largely overlaps the O/E-1 protein in mouse olfactory epithelium may allow compensation by other members of the family and lead to absence of observable phenotypes (Wang *et al.*, 1997). In other areas of the nervous system where O/E expression is transient and correlates with postmitotic differentiation, the expression of O/E family members is only partially overlapping resulting in distinct functions in different neuronal cell types (Garel *et al.*, 2000, 1999, 1997). In order to elucidate the role of O/E proteins, it is crucial to characterize all the O/E members. We have utilized novel homology cloning methods in an effort to

[1] Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom.

[2] To whom correspondence and reprint requests should be addressed. Fax: (410) 614-0827. E-mail: rreed@jhmi.edu.

1044-7431/02 $35.00

obtain a complete description of O/E family of transcription factors in mouse.

The numerous model organisms used for biological research and the molecular complexity within these organisms require the identification of orthologous genes and species-specific family members in a large number of different and evolutionarily distant species. Although the products of these genes may share common biochemical properties and play similar roles in related developmental pathways, the experimental advantages of particular systems require methods for the isolation of homologous gene families in disparate species. Homology-based cloning methods, including low-stringency hybridization with cDNA probes and PCR with degenerate primers, have been successful in many systems but each have their limitations. Bioinformatics-driven identification and isolation of additional family members may be limited by the availability and the complexity of the EST-database from the tissue and species of interest, and it is unlikely that high-coverage genomic sequences will be available for many species of experimental interest.

In this report, we present the cloning of a new member of the O/E family, O/E-4, using an alternative homology screening method, Degenerate-Oligo-base Direct Selection (DODS), based on Direct Selection (Lovett *et al.*, 1991). We demonstrate that this method is capable of selecting homologous sequences from an adaptor-ligated cDNA library. Furthermore, we have characterized the O/E-4 expression pattern in mouse olfactory system and various adult tissues. Interestingly, one splice form of O/E-4 lacks the C-terminal domain that has previously been suggested to contribute to transcriptional activation. All splice variants observed appear to interact with other O/E family member and bind the Olf-1 consensus site but result in only relatively weak transcriptional activation.

RESULTS

Identification of a New O/E Family Member by Degenerate-Oligo-Based Direct Selection (DODS)

We have modified the Direct Selection procedure, originally designed to look for coding sequences within a large stretch of DNA, to identify new family members within species and orthologs present in other species. Three degenerate oligonucleotides were designed on the basis of conserved amino acid sequences at three locations in the mammalian O/E-1, -2, and -3, the *Drosophila Collier* and the *C. elegans* UNC-3 proteins (Prasad *et al.*, 1998; Wang *et al.*, 1997; Crozatier *et al.*, 1996). A mixture of these biotinylated, degenerate oligonucleotides was combined with a random-primed, adapter-flanked olfactory cDNA library and used in an initial round of selection. Additional three rounds of selection were performed and the resulting pools analyzed to determine efficiency (Fig. 1).

Southern analysis of the PCR-amplified products present after each round of selection demonstrated that representation of O/E-related cDNAs increased dramatically in the first two rounds of selection (Fig. 2A). There was only modest additional enrichment of target sequences in subsequent rounds. The selection products derived from the second and fourth rounds of selection were subcloned into plasmid vectors and the inserts amplified by PCR. These insert fragments were spotted onto nitrocellulose membranes and hybridized with ^{32}P-labeled degenerate oligonucleotide probes identical to the ones used in the initial selection (Fig. 2B). Plasmids with inserts that yielded hybridization signals higher than background were sequenced. Sequence analysis of the cloned products showed that after two rounds of selection, approximately 10% of the selected cDNAs contained O/E-like sequences, and the frequency increased to 18% after four rounds of selection. Of the 96 clones analyzed from the second round, 9 (10%) contained O/E-related sequences. After the fourth-round of selection 17 of 96 inserts (18%) contained O/E-like sequences.

DODS Does Not Reduce the Complexity nor Introduce Bias

We next compared the sequences of each of the 26 O/E-related clones to determine whether they were derived from independent cloning events and reflected the complex population of O/E family members expressed in olfactory tissue. The sequence analysis revealed that 19 were fragments of O/E-1, -2 and -3 cDNAs (6, 7, and 6, respectively), while the remaining 7 represented a new O/E family member, O/E-4 (Fig. 2C). A further examination of the sequences revealed that each of the 26 clones was derived from an independent selection event and that four rounds of DODS did not bias the representation of their abundance in the library. The average size of the insert fragment (300 bp) resulted in most clones containing coding regions corresponding to two or three of the selection oligonucleotides, but many clones contained sequences corresponding to only a single selection oligonucleotide, suggesting that one degenerate oligonucleotide probe was sufficient for selection.

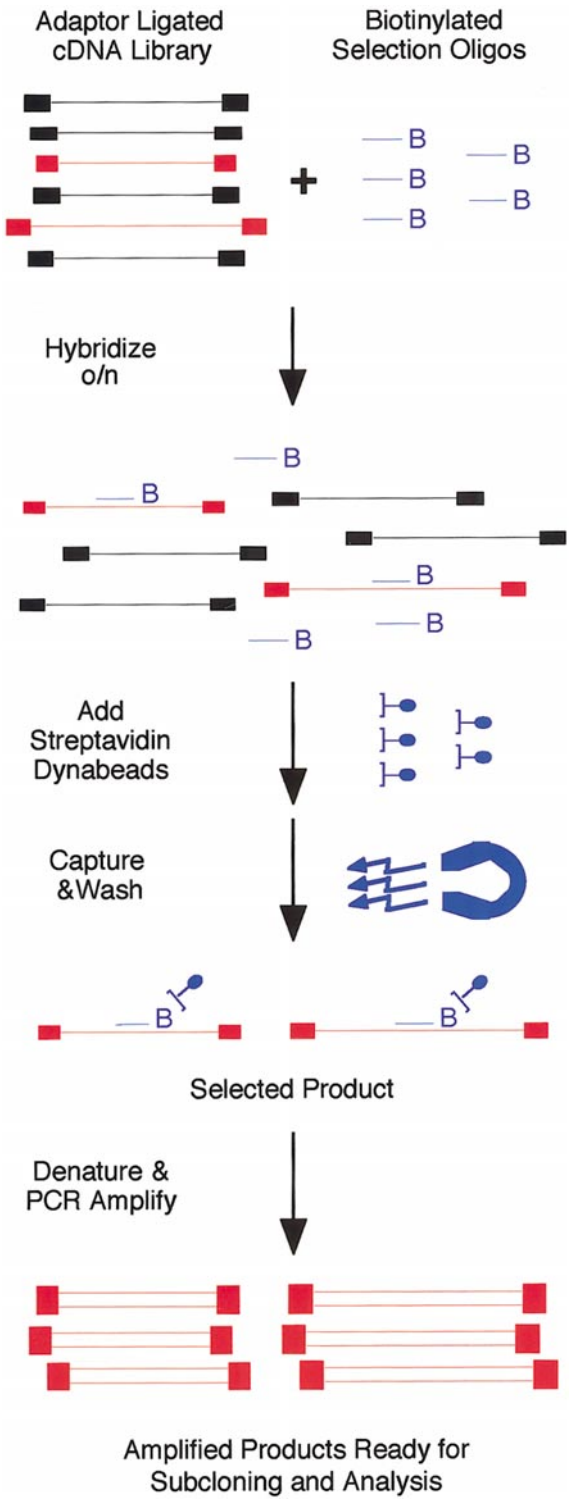


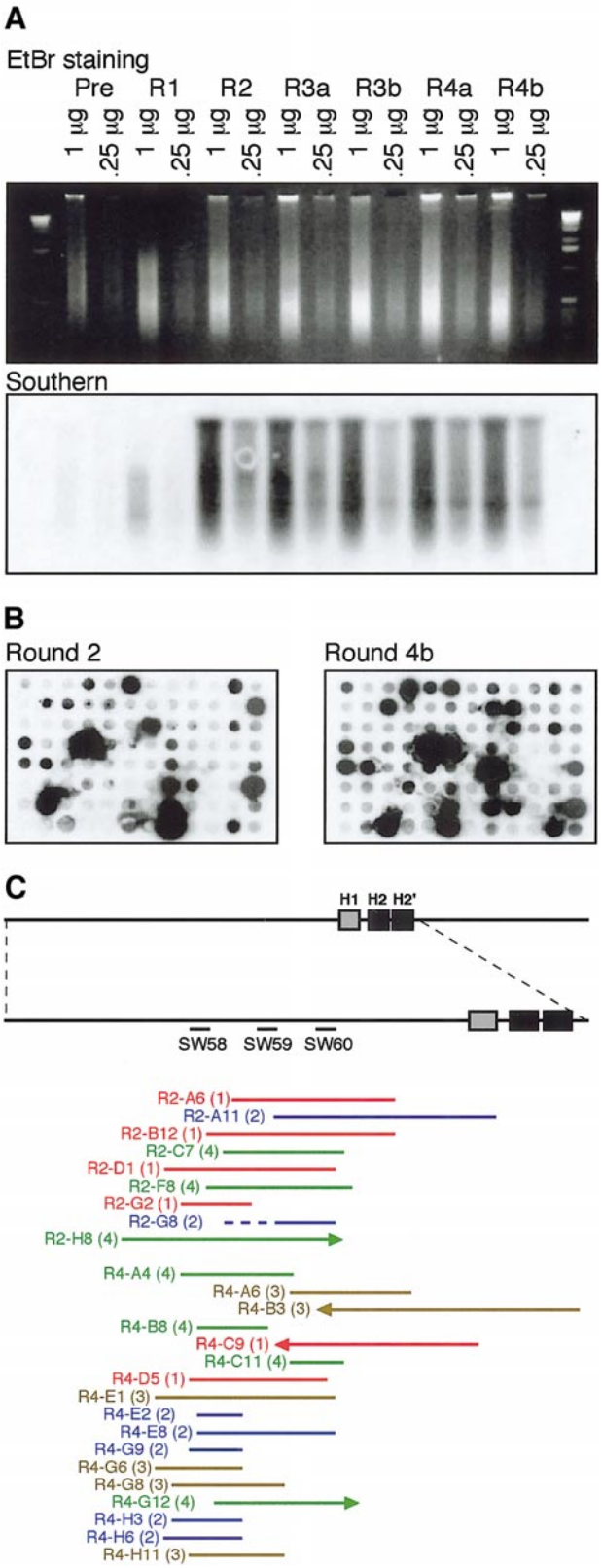
FIG. 1. Degenerate-oligo-based direct selection (DODS) scheme. The red lines represent *O/E*-related sequences and the black lines represent nonspecific sequences. The selection cycle can be repeated until significant enrichment of specific sequences is achieved.

Cloning and Expression of O/E-4 cDNA

The 120-bp fragment of O/E-4 selected in the initial DODS experiment provided a starting point for the isolation of a full-length cDNA. Initial screening of a mouse olfactory cDNA library resulted in two independent O/E-4 clones, but neither contained the putative start codon. Subsequent screening revealed four additional O/E-4 clones. The full-length O/E-4 cDNA encoding the start codon and 5′ untranslated sequences were obtained by RACE (rapid amplification of cDNA ends). The final cDNA clones used for functional studies were generated with this fragment and the previously isolated cDNAs (Fig. 3A).

Our study identified four alternative splice sites among the O/E-4 cDNA clones. All splice variants of O/E-4 encode the same N-terminal 391 amino acids, including the DNA binding domain and the first helix of the rHLH motif. The 5′ most alternative splice site determines whether the resulting cDNA encodes an intact rHLH motif. Some splice forms at this site contain additional nucleotides that introduce an in-frame stop codon between the two helices and thus result in a shortened form of O/E-4 (O/E-4S) lacking the second helix and the proline/serine-rich C-terminal domain. Four of the six cDNA clones we identified contain this premature termination and thus encode a shortened protein lacking the second helix of the rHLH motif (Figs. 3A and 3B). In contrast, the other spliced forms encode longer proteins with intact rHLH and proline/serine-rich C-terminal domain similar to the ones found in O/E-1, -2, and -3. The remaining three alternative splice sites give rise to proteins with different C-terminal ends. Although RT-PCR performed with primers spanning the splice sites showed that O/E-4/23 was the predominant form in olfactory epithelium (Fig. 4), minor forms may play a significant regulatory role. Additionally, one splice form (clone 10/14, Figs. 3A and 3B) was not detected in the olfactory tissue by RT-PCR, but it contains additional splice sites and a poly (A) tail and can not be ruled out as an immature transcript. In order to study all possible splice variants of O/E-4, we made fusion constructs of the cDNA clones that encode proteins with an intact rHLH but different C-terminal sequences (Fig. 3B).

The C-terminal domain of O/E-4/23 (the predominant form of O/E-4 in the olfactory epithelium) shares modest homology to that of other mouse O/E proteins. O/E-4 and other mouse O/E proteins share approximately 70% overall identity at the amino acid level. The N-terminal 410 amino acids of O/E-4 (from the N-terminus to the end of the rHLH motif) are 87, 88, and



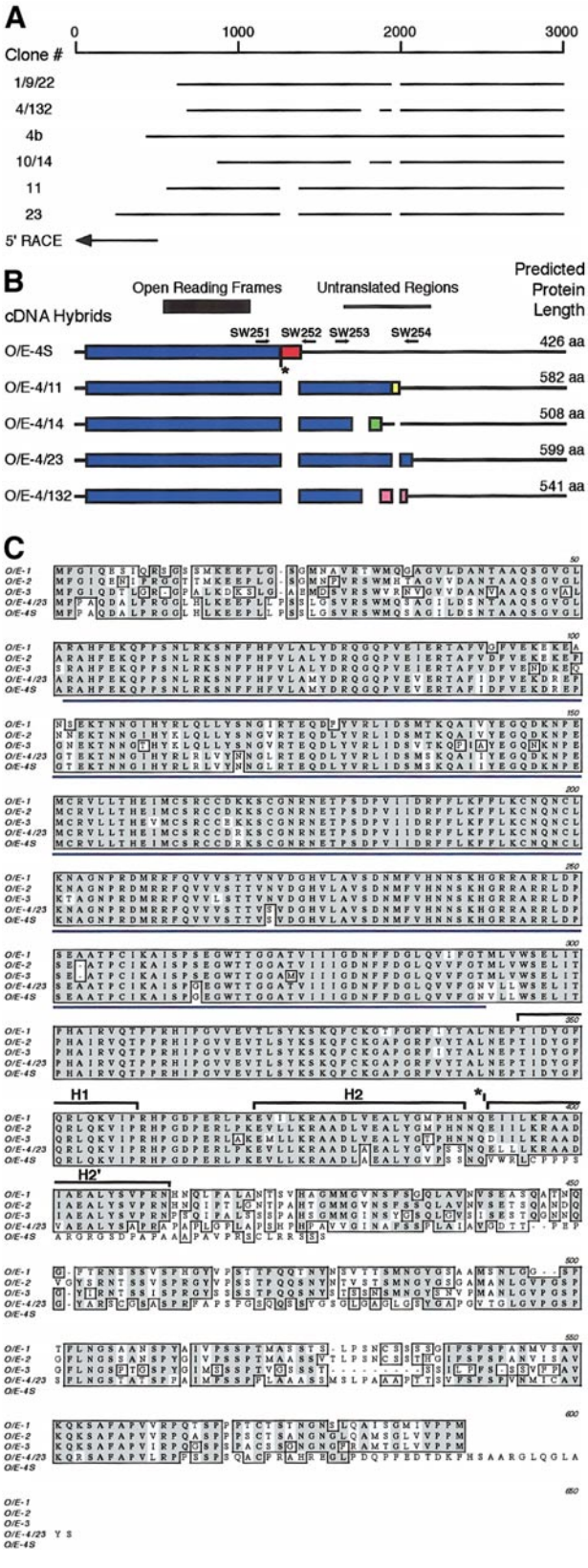
83% identical to the same regions of O/E-1, -2, and -3, respectively (Fig. 3C). Although O/E-4S and invertebrate O/E proteins all lack the second helix of the HLH domain, comparison of O/E N-terminal sequences does not reveal higher similarity between the invertebrate O/E proteins and O/E-4 than between invertebrate O/E proteins and O/E-1, -2, or -3.

O/E-4 Distribution

A semiquantitative RT-PCR was performed to determine the pattern of O/E-4 expression. RNA from 12 adult tissues was reverse transcribed, and PCR was performed with primers spanning the most 5′ alternative splice site (SW251 and SW252 in Fig. 3B) to determine O/E-4 expression and the predominant splice form in each tissue. In 25 cycles of PCR, a significant amount of O/E-4 was found only in the olfactory epithelium, and trace amounts were detectable in other neural tissues. With 30 cycles of PCR, O/E-4 was clearly present in the nervous tissues, and low amount of O/E-4 was detectable in lung, spleen, and kidney. In all tissues, the shorter splice form (encoding longer proteins with an intact rHLH) was the predominant species (Fig. 4A).

In order to investigate which O/E-4 protein is the dominant form in each of the O/E-4 positive tissue, PCR was performed with a second pair of primers spanning the remaining three alternative splice sites

FIG. 2. DODS results. (A) EtBr-stained agarose gel of pre-selection and selected cDNA and Southern hybridization of the same cDNA probed with ^{32}P-labeled degenerate oligonucleotides identical to the ones used in the selection. The selection products from the second round selection had independently undergone additional two rounds of selection, and thus the products from one selection sequence are labeled R3a and R4a, and the others were labeled R3b and R4b. (B) Selection products from Rounds 2 and 4b were subcloned into pBluescript, and the T3/T7 PCR products were dot-blotted in micro-titer-plate format and probed with ^{32}P-labeled selection oligonucleotides. (C) Individual positive clones from the dot-blot analysis shown in (B) were sequenced, and partial sequences of O/E-1, -2, -3, and -4 were obtained. The sequence of each positive clone and the locations of the three selection oligonucleotides (SW58, SW59, and SW60) with respect to a reference O/E sequence (shown on top of this panel as a straight line with solid boxes representing the helices of the HLH domain) are indicated. The helices are marked as H1, H2, and H2′ according to previous studies (Dubois and Vincent, 2001). Sequence identity of each clone is color coded with red representing O/E-1 sequences, blue for O/E-2s, brown for O/E-3s and green for O/E-4s. The corresponding location of each clone on the dot-blots is also given. For instance, R2-A6(1) indicates Round 2 selection, position A6, O/E-1 sequence.



(SW253 and SW254 in Fig. 3B) on tissues that showed any detectable O/E-4 in the RT-PCR experiment described above. The results indicated that O/E-4/23 is the predominant splice form in neuronal tissues. Interestingly, this restriction was not observed in the nonneuronal tissues and implied that O/E-4 expression is differentially regulated and may have distinct roles in neuronal and nonneuronal tissues (Fig. 4B).

The expression pattern of O/E-4 message in olfactory epithelium was examined by *in situ* hybridization with a probe derived from the 3′ coding region and part of the 3′ untranslated region of O/E-4. The O/E-4 message is restricted to the neuronal and basal cell layers, a pattern very similar to that of O/E-1, -2 and -3 (Fig. 5A). The vast majority of cells within the neuronal and basal layers hybridize with the O/E-4 probe, suggesting that individual cells are likely to coexpress O/E-4 and other O/Es. In contrast to other O/E messages, O/E-4 was not detected in the vomeronasal organ (VNO), suggesting a distinct role for O/E-4 in the expression of olfactory transduction components (Fig. 5B).

FIG. 3. Schematic drawings of alternatively spliced O/E-4 sequences and proteins they encode. (A) A partial O/E-4 sequence obtained by DODS was labeled and used to screen an oligo-dT primed mouse olfactory cDNA library. The cDNA library screening yielded 10 clones representing six independent clones, and their lengths and alternative splice sites are indicated with respect to the nucleotide numbering above. 5′ RACE was performed and the RACE product (as indicated by an arrow) contained the 5′ untranslated region (UTR) and the initiating ATG. (B) The O/E-4 mRNA is alternatively spliced at four sites. Each splicing event leads to a shift of reading frame resulting in proteins of different C-terminal sequences. The solid bars represent open reading frames in each splice variant, and each color represents an unique amino acid sequence. Full sequence of each O/E-4 splice variant was created using fragments obtained from the cDNA library screen and 5′ RACE. The positions of PCR primers used in the RT-PCR experiments shown in Fig. 4 (SW251 through SW254) are indicated. The position of the alternative splice junction between the two repeated helices is marked with an asterisk, and it's precise location is indicated in Fig. 3c. The nucleotide sequences O/E-4 cDNAs have been deposited into GenBank, and the accession numbers for O/E-4S, O/E-4/11, O/E-4/14, O/E-4/23, and O/E-4/132 are AF387630 through AF387634, respectively. (C) Alignment of O/E-1, -2, -3, and -4 protein sequences. Both O/E-4/23 and O/E-4S sequences are included. Alignment was performed using the ClustalW Alignment feature on the MacVector 6.5 sequence analysis software (Oxford Molecular Group). The DNA binding domain as defined in previous studies is underlined in blue, and the locations of the HLH domain helices are marked as H1, H2, and H2′ in which H2 and H2′ represent the two helices of the rHLH motif (Dubois and Vincent, 2001; Hagman *et al.*, 1993; Wang and Reed, 1993). The asterisk marks the alternative splice junction between the two repeated helices.

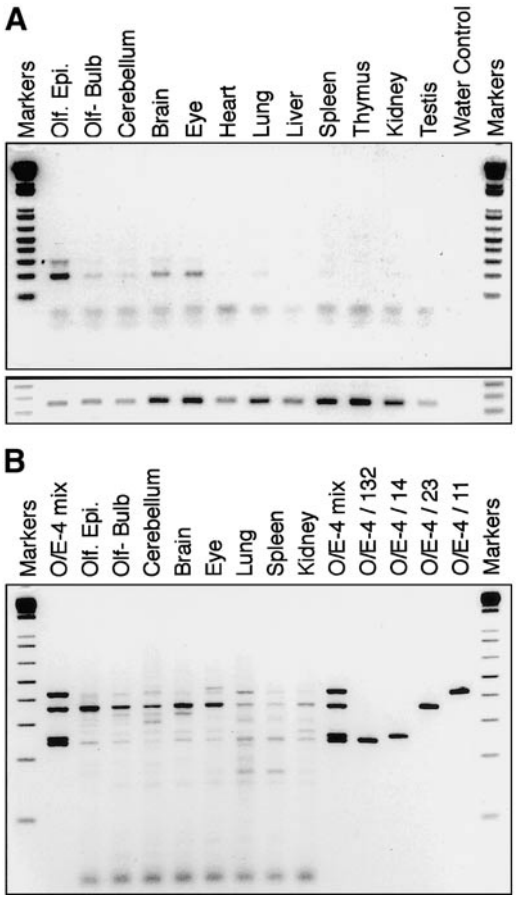


FIG. 4. O/E-4 message distribution in adult tissues. (A) The expression of O/E-4 messages in 12 adult tissues was determined by RT-PCR. The EtBr-stained agarose gel is shown as a negative image to increase clarity. Thirty cycles of PCR were performed with primers SW251 and SW252 (see Fig. 3B for primer locations) spanning the most 5′ alternative splice site. The quality of RNA from each tissue sample was determined by PCR with primers that amplify RNA polymerase II and shown at the bottom. (B) To determine the predominant splice form in each tissue sample, 35 cycles of PCR were performed with primers SW253 and SW254 spanning the remaining three alternative splice sites. PCR was also performed with isolated O/E-4 cDNAs as templates in order to provide size reference for each splice variant.

O/E-4 Has Similar But Distinct in Vitro Properties as Other Mammalian O/E Proteins

All mouse O/E proteins identified to date bind a pseudopalandromic consensus sequence, the Olf-1 site. We examined the ability of O/E-4 to bind an Olf-1 site by performing an electrophoretic mobility shift assay (EMSA) with extracts prepared from HEK293 cells expressing different splice variants of O/E-4 cDNAs. Double-stranded oligonucleotide containing the olfactory cyclic nucleotide-gated channel (*OcNC*) Olf-1 site was shifted by each O/E-4 splice variant examined but not by extracts from cells transfected with pCIS empty

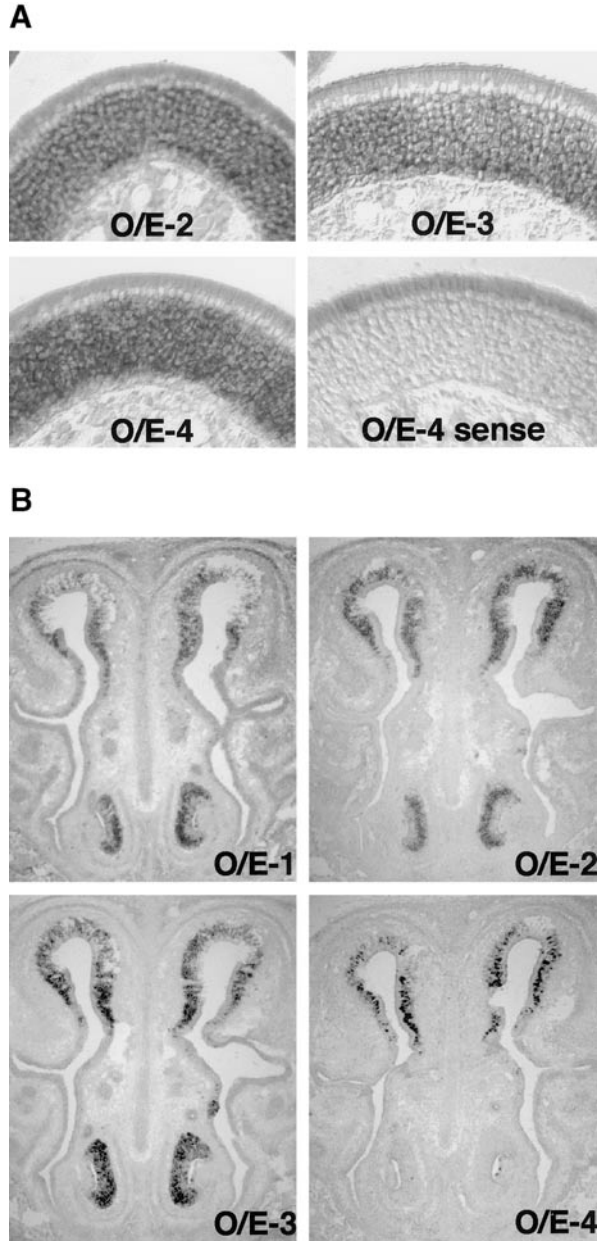


FIG. 5. *In situ* hybridization of adult and embryonic olfactory tissue. (A) The pattern of O/E-4 message in adult mouse olfactory epithelium by *in situ* hybridization. Similar to patterns of O/E-2 and O/E-3, O/E-4 message is restricted to the neuronal cell layer and the basal cell layer and is absent in the sustentacular cell layer. (B) Patterns of O/E messages in E16.5 olfactory epithelium and vomeronasal organ (VNO) by *in situ* hybridization. Unlike the other mouse O/E messages, O/E-4 is absent in the VNO.

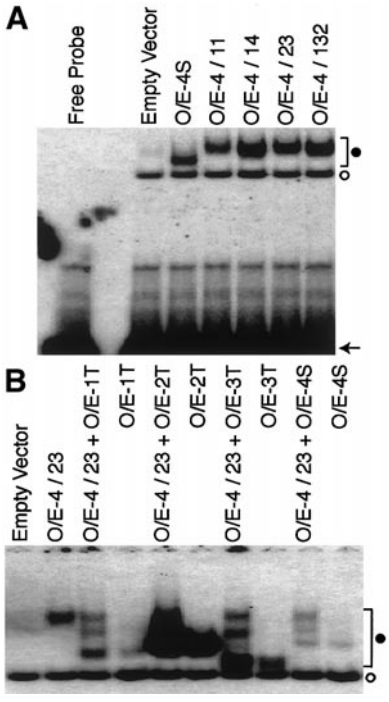


FIG. 6. O/E-4 proteins bind the O/E site as dimers. (A) The ability of the O/E-4 proteins to bind *OcNC* O/E site is shown. Extracts of HEK293 cells transfected with different O/E-4 expression constructs were used in electrophoretic mobility shift assay (EMSA). All splice variants of O/E-4 are able to bind the O/E site and retard its mobility in EMSA. The positions of O/E-DNA complexes (closed circle), nonspecific complexes (open circle), and free probes (arrow) are indicated. (B) EMSA was performed with extracts of HEK293 cells cotransfected with O/E-4/23 and truncated forms of O/Es (O/E-1T, 2T, and 3T) lacking sequences beyond the HLH domain or shortened O/E-4 (O/E-4S) lacking the second helix and sequence beyond. The coexpression of full-length and truncated forms of O/Es resulted in the formation of an additional protein-DNA complex of intermediate mobility representing a dimeric protein complex containing a full-length and a truncated O/E.

vector (Fig. 6A). Even the O/E-4S protein, which lacks the second helix of the rHLH motif found in all vertebrate O/E proteins, binds the Olf-1 consensus site as an apparent dimer.

O/E-1, -2, and -3 form homodimers and heterodimers that bind to DNA and activate transcription. The presence of a nearly identical HLH domain in the O/E-4 suggested that it might form homodimers as well as heterodimers with other O/E proteins. C-terminal-truncated forms of O/E proteins lacking amino acids downstream of the HLH domain retain the ability to bind DNA and form complexes with higher mobility than those observed with the full-length protein (Hagman *et al.*, 1993; Wang and Reed, 1993). EMSA performed with extracts from cells expressing full-length and truncated O/Es resulted in additional protein-DNA complexes of intermediate mobility representing O/E-1, -2, and -3 hetero-oligomeric protein–DNA complexes comprised of full-length and a truncated O/E proteins. Using a similar approach, we coexpressed O/E-4/23 with the truncated forms of O/E-1, -2, and -3 and showed that each was able to form heterodimers that bound an Olf-1 site (Fig. 6B). Interestingly, O/E-4S was able to interact with a full-length O/E-4/23 to form a DNA-binding complex. This result suggests that an intact rHLH motif is not necessary for dimmer formation and DNA binding.

The O/E-1, -2, and -3 proteins activate transcription in heterologous expression systems. The high degree of conservation between O/E-4 and other mammalian O/E proteins suggested that O/E-4 might also activate transcription. We directly addressed this hypothesis using HEK293 cells cotransfected with an O/E expression vector and a luciferase reporter construct containing ten O/E binding sites adjacent to an SV40 minimal promoter. Under our assay conditions, each of the O/E-4 splice variants tested evoked significantly less transcriptional activation than the other mammalian O/E proteins examined (Fig. 7). The strongest O/E-4 transactivator, O/E-4/23, produced less than half the luciferase reporter than observed for the other O/E proteins. These results demonstrated that O/E-4 proteins have distinct transactivation activities and may play a role in regulating other O/E proteins.

DISCUSSION

The Contribution of O/E-4 to Olfactory Gene Regulation

O/E-4 was identified in a random-primed adaptor-ligated cDNA library made from olfactory epithelium. *In situ* hybridization with an O/E-4 specific probe revealed that O/E-4 mRNA is expressed in the neuronal and basal cell layers of the olfactory epithelium, similar to the pattern reported for O/E-1, -2, and -3. The O/E-4 protein binds the consensus Olf-1 site as a homo-oligomer or as a hetero-oligomer with the other O/E proteins. O/E-4, unlike other mammalian O/E proteins, is unable to elicit high-level reporter expression in a heterologous system.

None of the O/E-4 splice variants were able to achieve the level of transcriptional activation observed for the other O/E proteins in our reporter assay system. Among O/E-4 splice variants, O/E-4/23 is the stron-

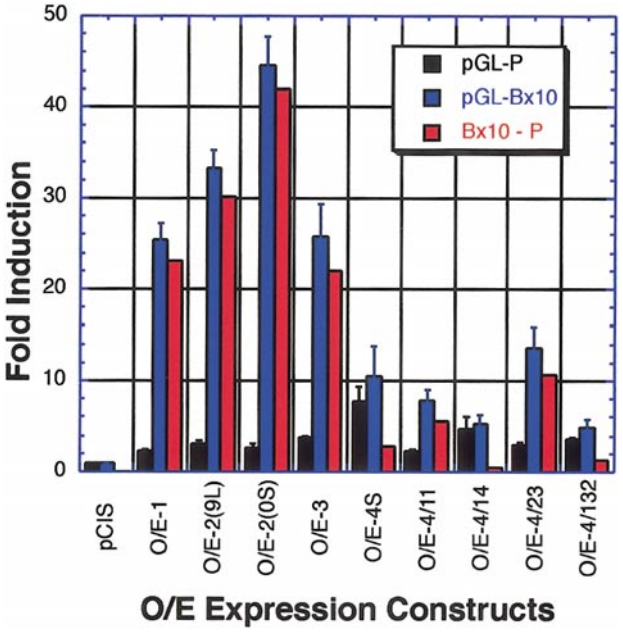


FIG. 7. O/E-4s are weak transcription activators in luciferase assays. The ability of O/E proteins to activate transcription was demonstrated with luciferase reporter assay. HEK293 cells were cotransfected with each O/E expression vector and a reporter construct containing the luciferase gene driven by either ten O/E binding sites adjacent to an SV40 minimal promoter (pGL-BX10, blue bars) or an SV40 minimal promoter alone (pGL-P, black bars), and the difference in luciferase activity (Bx10-P) was shown with red bars. The data represents the average values from three experiments, and the values from each experiment were obtained from three independent transfactions of each expression construct performed on the same day. The error bars represent the standard error. Baseline activity (onefold induction) was determined with the extract of cells cotransfected with pCIS empty expression vector and the luciferase reporter constructs.

gest transactivator, but this activity is only half of that of O/E-1 and O/E-3, and 1/3 of O/E-2. The other O/E-4 splice variants have even lower activation activity (less than 1/2 of O/E-4/23 activity), and are barely discernable above background. It is possible that assays performed using different reporters or in different cell types might give other results. However, it is interesting to speculate that O/E-4 may function by modulating the activity of other O/E proteins through the formation of heterodimers or by binding at target elements.

The role of the rHLH motif in O/E function has not been resolved in previous studies. Invertebrate O/E proteins do not have the second helix of the rHLH motif. An alternative HLH domain was proposed for O/E family of transcription factors but a functional importance for this structure has not been demonstrated (Dubois and Vincent, 2001; Crozatier *et al.*, 1996). This raised the question of whether an intact rHLH is crucial for mammalian O/E function. It was shown that an intact rHLH domain can mediate dimer formation, and the omission or disruption of the rHLH greatly reduced DNA binding and transcriptional activation in O/E-1 (Hagman *et al.*, 1995; Hagman *et al.*, 1993; Wang and Reed, 1993). The identification of an O/E-4 splice variant that lacks the second helix of the rHLH motif and retains DNA binding with similar efficiency as other O/E-4s indicates that an intact rHLH, though sufficient, is not necessary for efficient dimerization and DNA binding. This is the first evidence of a mammalian O/E protein without an intact rHLH motif. The ability of O/E-4S to partner with other O/E proteins and bind DNA further validates the importance of the HLH domain proposed for the invertebrate O/E proteins, and perhaps any two of the three proposed helices can constitute a functional dimerization domain. The identification of cDNAs encoding proteins with this novel structure will prompt further experiments to elucidate their role *in vivo.*

DODS Provides an Alternative to Traditional Homology Screening Methods

Direct Selection cloning methods were originally designed for rapid enrichment and identification of cDNAs encoded by large genomic regions. The basis of the original protocol was the hybridization of a complex cDNA library to immobilized genomic DNA. Through multiple cycles of selection/amplification process, cDNAs that strongly hybridize to the genomic sequence were preferentially enriched. In our selection scheme, DODS, degenerate oligonucleotides representing conserved regions of a gene family were used to enrich cDNA fragments derived from members of the gene family of interest. Here we demonstrate that DODS is capable of identifying new members of a conserved gene family.

DODS retains several advantages of conventional homology screening methods while avoiding several potential problems. Low-stringency hybridization screening often yields ambiguous hybridization signals that can easily be missed in the background of stronger "perfectly matched" signals. Methods based on PCR amplification with degenerate pairs of oligonucleotides rely heavily on the knowledge of conserved domains with favorable spacing. In addition, successful identification of two different oligonucleotides and accurate prediction of the 3′ end sequence for each oligonucleotide is essential. Finally, PCR-based methods are highly

susceptible to representational bias of their abundance in the tissue. In the DODS procedure, the subcloned selection products are individually analyzed to reduce the possibility of losing weaker hybridization signals to the background. DODS does not rely on primer-extension in the selection process and the 3′ regions of the selection oligonucleotides are not essential. This allows longer oligonucleotides to be designed and selection to be achieved at any region along the degenerate sequence. Products are amplified with a single oligonucleotide between selections so low abundance targets are not likely to be lost. As demonstrated in our study, the complexity of the selected products remained diverse even after more than 100 cycles of PCR amplification (both during library amplification and selection product amplification).

The flexibility in oligonucleotide design is perhaps the greatest advantage of DODS. We demonstrated that one selection oligonucleotide of 30 bases was sufficient for enrichment and thus eliminate the requirement for multiple or one large homologous region in the gene family of interest. Furthermore, DODS may also allow greater degeneracy in the oligonucleotides than typical PCR applications. The selection oligonucleotides used in this study had degeneracies of 2^{11} to 2^{13}. When there are multiple homologous regions available in a gene family, the oligonucleotides with lowest degeneracy in each homologous region can be used together for DODS, a strategy that would optimize the identification of all homologous sequences. In addition to homology screening, DODS may also be used to identify cDNAs with limited amino acid sequence information (i.e., from peptide sequencing or phage display selection).

EXPERIMENTAL METHODS

Library Construction

The selection library was constructed using the Marathon cDNA Amplification kit (Clontech, Palo Alto, CA) according to the manufacturer's instructions but with an alternate adaptor. First strand cDNA was synthesized with random hexamers and poly(A)+ olfactory epithelium RNA from C57BL6 mice, and the subsequent second strand cDNA synthesized according to the kit protocol. The adaptor was prepared by annealing 5′-phosphorylated SW144 oligonucleotide (5′-AGGTTCTCGAGCGGCCGCCCGGG-3′) to SW143 (5′-AAGGTCCCGGGCGGCCGCTCGAGAACCT-3′), and the resulting double stranded adaptor was ligated to the synthesized cDNA. The ligation product was purified by gel filtration over a NICK column (Amersham Pharmacia, Piscataway, NJ), and 20 100-μl fractions were collected. Five microliters of each fraction was PCR amplified (94°C for 30 s, 65°C for 45 s, 72°C for 2 min, 25 cycles) with SW143 in the presence of Pfu DNA polymerase (Stratagene, La Jolla, CA). After amplification, the PCR products were pooled, phenol/chloroform extracted, ethanol precipitated, and resuspended in water. This material was used as the library for DODS.

Degenerate-Oligo-Based Direct Selection (DODS)

The amino acid sequences of O/E-1, -2 and -3, UNC-3 and *Collier* were aligned and the homologous domains were reverse translated. Three biotinylated degenerate-oligonucleotides: SW58 (5′-TAYGARGGNCARGAYAARAAYCCNGARATGTG-3′), (aa141–151, 2048 degeneracy); SW59 (5′-AARTTYTTYYTNAARTGYAAYCARAAYTG-3′), (aa189–198, 2048 degeneracy); and SW60 (5′-AAYATGTTYGTNCAYAAYAAYWSNAARCAYGG-3′), (aa231–241, 8192 degeneracy) were designed to contain all possible codons at conserved amino acid regions that would produce the lowest degeneracy. The selection oligonucleotides were mixed in equal-weight concentration and used for the DODS procedure.

The selection library and the selection oligonucleotide (1.5 μg each) of were mixed in minimum volume ($<$5 μl), and appropriate amount of ExpressHyb (Clontech) was added to a final volume of 25 μl. The selection reaction was topped with mineral oil, denatured in boiling water for 5 min and immediately placed at 50°C overnight. Strepavidin–Dynabeads (200 μl) (Dynal, Oslo, Norway) were washed three times and resuspended in 200 μl of binding buffer (5 mM Tris, 0.5 mM EDTA, and 1 M NaCl, pH 7.5). The selection mixture was added to the prepared Dynabeads and binding continued for 1/2 h at room temperature with agitation. In 500-μl volume, the oligonucleotide-bound Dynabeads were washed five times in 5$\times$ SSC, 1% SDS at room temperature, six times in 2$\times$ SSC, 1% SDS at 50°C for 10 min each, and rinsed three additional times in the wash buffer at room temperature. The captured cDNA/selection oligonucleotide/Dynabead complex was denatured in 50 μl of 100 mM NaOH at 65°C for 10 min. After separating the solution containing captured cDNAs from the selection oligo/Dynabead complex, the cDNAs were neutralized in 25 μl of 1 M Tris (pH 7) and applied to a NICK column. Fractions were PCR amplified as the DODS library, and the pooled PCR products were subjected to three additional rounds of selection.

Southern blot analysis was performed on PCR-amplified products from each selection to determine the efficiency of DODS selection. The second and fourth rounds selection were digested with *Not*I and cloned into pBluescript (Stratagene). Ninety-six clones were picked from each selection, and inserts were PCR-amplified with T3 and T7 primers and dot-blotted onto nitrocellulose membranes. The selection oligonucleotides were labeled with [γ-^{32}P]ATP and hybridized to the gel-fractionated PCR amplification products or dot-blotted inserts in ExpressHyb for 2 h at 50°C. The Southern blots were washed twice in 2× SSC, 1% SDS at 55°C for 15 min each.

Phage Library Screening

Full-length O/E-4 was obtained from an oligo-dT-primed adult mouse olfactory epithelium cDNA library and 5′ RACE (rapid amplification of cDNA ends). The cDNA library was first screened using a probe derived from the initial DODS clone and resulted in two independent clones, the longer of which still was approximately 180 amino acid residues downstream from the location of the N-terminus of other family members. The library was rescreened with a 300-bp fragment corresponding to cDNA sequence from the 5′ end of the longer cDNA clone to the first *Sac*I site, and four additional independent clones were obtained. The longest isolate was still 60 amino acids short of the putative N-terminus. In total, 2.5×10^6 phage clones were screened and six independent O/E-4 sequences were isolated. We next performed 5′ RACE with a Marathon cDNA amplification kit and obtained the initiating methionine and 5′-untranslated region. The RACE product was joined to the longest cDNA clone to generate a cDNA that contained the entire coding sequence.

RT-PCR (Reverse Transcription-Polymerase Chain Reaction)

Total RNA from 12 adult tissues was extracted with RNAzol B (Tel-Test, Inc., Friendswood, TX) according to manufacture's instruction. RNA (2 μg) was reverse-transcribed with M-MLV reverse transcriptase (Ambion, Austin, TX) and random hexamer in 40-μl volume, and 1 μl was used for each PCR reaction. Two primer pairs were used to determine the composition of O/E-4 splice variants. SW251 (5′-GCTTTGTCTACACAGCCCT-3′) and SW252 (5′-ACGTCCGCCGCACGTTTCA-3′) spanned the most 5′ alternative splice site, and SW253 (5′-GCTATGGCAGTGGCCTTGGA-3′) and SW254 (5′-GCAGAGTGGAACTTGTCAGTAT-3′) spanned the remaining three alternative splice sites. All PCR were carried out with 94°C denaturing, 60°C annealing, and 72°C extension (1 min each) cycles.

In Situ Hybridization

Olfactory tissue was fixed in Bouins solution as described (Davis and Reed, 1996). Mouse embryonic tissue was fixed with 4% paraformaldehyde before sectioning. The RNA *in situ* hybridization was carried out as described (Wang *et al.*, 1997) with the following modifications. The probe hybridization of 14- to 20-μm tissue sections was carried out with 0.5 μg/ml of digoxigenin-labeled riboprobe at 65–70°C overnight. Posthybridization wash was carried out twice in 0.2× SSC at 70°C, and antibody incubation was performed in the presence of 5% heat-inactivated normal goat serum.

Digoxigenin-labeled riboprobes for *in situ* hybridization (DIG RNA Labeling Mix, Roche, Indianapolis, IN) were derived from the divergent sequences corresponding to the C-terminal coding and the 3′-untranslated regions of the O/E cDNAs. The O/E-1 (*Hin*dIII/*Ava*II), O/E-2 (*Xmn*I/*Hin*dIII), O/E-3 (*Pst*I/*Bsp*HI), and the O/E-4 (3′ most *Sac*I/*Xba*I) probes were first subcloned into pBluescript KS II (Stratagene), and antisense and sense (control) probes were synthesized with T3 or T7 RNA polymerase (Stratagene), respectively. Previous Southern hybridization analysis confirmed that cross-hybridization was eliminated among the three O/E genes even in the most conserved regions at the high stringency wash condition (0.2× SSC, 65°C) used in these experiments.

Electrophoretic Mobility Shift Assay

Electrophoretic mobility shift assays (EMSA) were performed with whole cell extracts (Cell Culture Lysis Reagent, Promega, Madison, WI) from HEK293 cells transiently transfected with a CMV promoter-driven mammalian expression vector (pCIS) (Wang and Reed, 1993) directing expression of the O/E proteins. The full-length and truncated O/E-1, 2 and 3 expression constructs were previously described (Wang *et al.*, 1997). The splice variants of O/E-4 were cloned into pCIS as *Not*I/*Xho*I fragments.

Luciferase Reporter Assay

The constructions of the reporter plasmids and the expression plasmids for the O/E family members has been described (Wang *et al.*, 1997). For expression experiments, one 60-mm plate of HEK293 cells was tran-

siently transfected with 1 μg of the indicated pGL-based reporter plasmid along with indicated amount of pCIS-O/E plasmid. All transfections were adjusted to 5 μg total DNA with empty pCIS expression vector DNA. The luciferase reporter activity was measured from equivalent amount of protein lysate of each sample using a luciferase assay system (Promega) and Monolight 2010 luminometer (Analytical Luminescence Laboratory). The relative luciferase activity was calculated using the reference indicated in each experiment.

ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute and the National Institute on Deafness and Other Communication Disorders.

REFERENCES

Bally-Cuif, L., Dubois, L., and Vincent, A. (1998). Molecular cloning of Zcoe2, the zebrafish homolog of Xenopus Xcoe2 and mouse EBF-2, and its expression during primary neurogenesis. *Mech. Dev.* **77:** 85–90.

Crozatier, M., Valle, D., Dubois, L., Ibnsouda, S., and Vincent, A. (1996). Collier, a novel regulator of Drosophila head development, is expressed in a single mitotic domain. *Curr. Biol.* **6:** 707–718.

Davis, J. A., and Reed, R. R. (1996). Role of Olf-1 and Pax-6 transcription factors in neurodevelopment. *J. Neurosci.* **16:** 5082–5094.

Dubois, L., and Vincent, A. (2001). The COE—Collier/Olf1/EBF—transcription factors: Structural conservation and diversity of developmental functions. *Mech. Dev.* **108:** 3–12.

Dubois, L., Bally-Cuif, L., Crozatier, M., Moreau, J., Paquereau, L., and Vincent, A. (1998). XCoe2, a transcription factor of the Col/Olf-1/EBF family involved in the specification of primary neurons in Xenopus. *Curr. Biol.* **8:** 199–209.

Garel, S., Garcia-Dominguez, M., and Charnay, P. (2000). Control of the migratory pathway of facial branchiomotor neurones. *Development* **127:** 5297–5307.

Garel, S., Marin, F., Grosschedl, R., and Charnay, P. (1999). Ebf1 controls early cell differentiation in the embryonic striatum. *Development* **126:** 5285–5294.

Garel, S., Marin, F., Mattei, M. G., Vesque, C., Vincent, A., and Charnay, P. (1997). Family of Ebf/Olf-1-related genes potentially involved in neuronal differentiation and regional specification in the central nervous system. *Dev. Dynam.* **210:** 191–205.

Hagman, J., Gutch, M. J., Lin, H., and Grosschedl, R. (1995). EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains. *EMBO J.* **14:** 2907–2916.

Hagman, J., Belanger, C., Travis, A., Turck, C. W., and Grosschedl, R. (1993). Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. *Genes Dev.* **7:** 760–773.

Lin, H., and Grosschedl, R. (1995). Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376:** 263–267.

Lovett, M., Kere, J., and Hinton, L. M. (1991). Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* **88:** 9628–9632.

Pozzoli, O., Bosetti, A., Croci, L., Consalez, G. G., and Vetter, M. L. (2001). Xebf3 is a regulator of neuronal differentiation during primary neurogenesis in xenopus. *Dev. Biol.* **233:** 495–512.

Prasad, B. C., Ye, B., Zackhary, R., Schrader, K., Seydoux, G., and Reed, R. R. (1998). unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors. *Development* **125:** 1561–1568.

Wang, M. M., and Reed, R. R. (1993). Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature* **364:** 121–126.

Wang, S. S., Tsai, R. Y., and Reed, R. R. (1997). The characterization of the Olf-1/EBF-like HLH transcription factor family: Implications in olfactory gene regulation and neuronal development. *J. Neurosci.* **17:** 4149–4158.

Received October 30, 2001
Revised March 20, 2002
Accepted March 25, 2002