Primary Structure, Neural-specific Expression, and Chromosomal Localization of Cux-2, a Second Murine Homeobox Gene Related to Drosophila cut*

Susan E. Quaggin‡, Gregory B. Vanden Heuvel‡, Krista Golden‡, Rolf Bodmer‡, and Peter Igarashi‡

From the ‡Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06520 and §Department of Biology, University of Michigan, Ann Arbor, Michigan 48109

The cut locus of Drosophila encodes a diverged homeodomain-containing protein that is required for the development of external sensory (es) organs and other tissues. A homologous gene (Cux-1) that encodes a transcriptional repressor has been identified in the mouse and other mammals. We have identified a second murine homeobox-containing gene (designated Cux-2) that is structurally related to Drosophila cut. The murine Cux-2 homeobox was similar to Drosophila cut and encoded a homeodomain that contained a characteristic histidine residue at position 50. The predicted Cux-2 protein contained 1426 amino acids and included three internal 60-amino acid repeats (Cut repeats) that were previously found in Drosophila Cut and murine Cux-1. Unlike Cux-1, expression of Cux-2 was restricted to neural tissue. In the adult brain, Cux-2 was prominently expressed in neurons in the thalamus and autonomic system. In embryos, Cux-2 was expressed in the developing central and peripheral nervous systems, including the telencephalon and peripheral ganglia of the trigeminal and glossopharyngeal nerves. A glutathione S-transferase fusion protein containing the carboxyl-terminal Cut repeat and homeodomain of Cux-2 exhibited sequence-specific binding to oligonucleotides derived from the promoter of the Neam gene. Using an interspecific backcross panel, Cux-1 and Cux-2 were mapped to distinct loci that were genetically linked on distal chromosome 5. These results demonstrate that a family of homeobox genes related to Drosophila cut is located on chromosome 5 in the mouse. Cux-2 is expressed exclusively in the central and peripheral nervous systems, and the Cux-2 gene product binds to DNA in a sequence-specific manner. Cux-2 may encode a transcription factor that is involved in neural specification in mammals.

Homeobox-containing genes encode transcription factors that are important for tissue-specific and developmental regulation of gene expression. The homeobox is a 183-bp nucleotide sequence that encodes a DNA-binding motif called the homeodomain (1). Homeobox genes were first identified in the Antennapedia/Bithorax complexes of Drosophila melanogaster but have subsequently been found in higher metazoans including Xenopus, rodents, and humans. In general, homeobox genes can be classified into the Antennapedia (Antp)-type genes and the diverged (non-Antp-type) genes (2, 3). Antp-type homeobox genes, which include the Drosophila HOM genes and the evolutionarily related mammalian Hox genes, are clustered on the chromosome(s) and encode homeodomains that are structurally similar (60–90% identical) to the archetypal Drosophila Antp gene. Antp-type homeobox genes are primarily expressed in mesoderm and ectoderm at developmentally defined stages, and loss-of-function mutations of Antp-type genes perturb anterior-posterior pattern formation and specification of body segment identity (4). The diverged (non-Antp-type) homeobox genes, such as members of the POU, engrailed, paired, and caudal families, are not clustered on the chromosome and encode homeodomains that are less similar in sequence (<60% identical) to Antp. The spatiotemporally restricted patterns of expression of diverged homeobox genes and the phenotypes that result from alterations in their expression indicate that these genes are also important in developmental processes during embryogenesis (3, 5).

The cut locus of Drosophila contains a diverged homeobox gene that is essential for development of the peripheral nervous system and other tissues. Deletions of the centromere-proximal region of the cut locus disrupt development of the peripheral nervous system and are embryonic lethal. The peripheral nervous system of Drosophila consists of sensory organs that are of two types: external sensory neurons which comprise mechanoreceptors or chemoreceptors, and chordotonal neurons which are proprioceptive. Null mutations or deletions of the cut gene cause the homeotic transformation of external sensory organs into ectopic chordotonal organs (6). Moreover, ubiquitous expression of cut in Drosophila embryos specifically transforms chordotonal organs into ectopic external sensory organs (7). Thus, cut is required for specification of neuronal identity in external sensory organs comprising the peripheral nervous system. The cut gene product is a protein of 2,175 amino acids that contains several distinctive structural features including a diverged homeodomain near the carboxyl terminus, three internal repeats (called Cut repeats) consisting of 60 amino acids with 55–68% amino acid identity, and acidic

*dpc., days post coitus; EMSA, electrophoretic mobility-shift assays; CDP, CCAAT displacement protein; GST, glutathione S-transferase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; RFLP, restriction fragment length polymorphism; ISH, in situ hybridization; PCR, polymerase chain reaction.
domains near the amino terminus (8). Antibodies to Cut recognize 280- and 320-kDa proteins that have a restricted distribution in *Drosophila* embryos (9, 10). The antibodies label nuclei of cells comprising external sensory organs but not chordotonal organs (8). Cut protein is also present in the nuclei of epithelial cells lining Malpighian tubules, cells surrounding the anterior and posterior spiracles, and many cells in the central nervous system (9, 10). In the developing embryo, Cut protein is present in external sensory organ precursor cells and the anlagen of the Malpighian tubules in the hindgut. Null mutants of cut produce either no Cut protein or truncated proteins with an aberrant cytoplasmic localization. In mutants with absent or cytoplasmic localization of Cut, external sensory organs are transformed into chordotonal organs (9). In addition, null mutants fail to develop Malpighian tubules and exhibit incomplete air-filling of the trachea and loosening of commissural bundles in the central nervous system (10, 11). Taken together, these studies indicate that the Cut protein is required for normal development of the peripheral and central nervous systems, Malpighian tubules, and spiracles.

A homologue of *Drosophila* Cut has been identified in the mouse, human, and other mammals (12–15). The murine homologue, named *Cux-1*, was cloned from neuroblastoma cells and identified as a transcriptional repressor of the neural cell adhesion molecule gene (*Ncam*) (13). The human homologue encodes CCAAT displacement protein (CDP), a transcriptional repressor that is involved in tissue-specific expression of cytochrome *b* heavy chain (gp91-phox) and A- 

### EXPERIMENTAL PROCEDURES

**cDNA Cloning**—Cut homologues were amplified using PCR with degenerate primers (5'-GARGARCARGARCGCTTY-3' and 5'-CATT-

**Northern Blot Analysis**—Poly(A)+ RNA was isolated from murine tissues, electrophoresed, and transferred to nylon filters (GeneScreen Plus, DuPont NEN) as described previously (20). Northern blots were hybridized with the Cux-2 cDNA clone *a9* containing nucleotides 3433–4268 in Fig. 1 which was labeled with [α-32P]dCTP using random primer extension. Hybridization was performed overnight at 42°C in Church-Gilbert medium containing 20% formamide, 10 mM magnesium acetate, salmon sperm DNA, and 2 × 10^6 cpm/ml 32P-labeled probe. Filters were washed for 30 min in 2 × SSC at room temperature, then 30 min in 0.1 × SSC containing 0.5% SDS at 65°C. Autoradiography was performed for 1 week at −70°C with a single intensifying screen. Northern blots were then stripped and rehybridized with the 1.3-kb PCRamplified fragment of pGAPD-1. 32P-labeled rat glycerolaldehyde-3-phosphate dehydrogenase (GAPDH) (21) using hybridization and washing conditions described previously (20).

**In Situ Hybridization (ISH)**—32P-Labeled riboprobes were synthesized as described previously (20). Antisense and sense riboprobes were synthesized from the linearized Cux-2 cDNA clone *a9* using T7 and T3 RNA polymerase, respectively. 10-μm thick cryosections of mouse embryos were dehydrated by means of Scott. Because our studies reveal the existence of at least two murine *cut* homologues, we refer to the homologue cloned by Valarche et al. as *Cux-1*. *Cux-2* refers to the second *cut* homologue reported herein. The corresponding gene symbols assigned by the International Committee on Standardized Nomenclature for Mice are *Cut1* and *Cut2*, respectively.

2 The murine *cut* homologue previously cloned by Valarche et al. (13) was named *Cux* in accordance with the nomenclature recommendations of Scott. Because our studies reveal the existence of at least two murine *cut* homologues, we refer to the homologue cloned by Valarche et al. as *Cux-1*. *Cux-2* refers to the second *cut* homologue reported herein. The corresponding gene symbols assigned by the International Committee on Standardized Nomenclature for Mice are *Cut1* and *Cut2*, respectively.
Murine Homeobox Gene Related to Drosophila cut

Results

CDNA Cloning of Cux-2—At the time these studies were initiated, no mammalian cut homologues had yet been cloned. Accordingly, murine cut homologues had yet been cloned.

Mice (strains CD-1 or C57BL/6J) were obtained from The Jackson Laboratories (Bar Harbor, ME) or Charles River (Wilmington, MA). To obtain embryos at defined developmental stages, mice were randomly mated, and the mating date of copulatory plugs was designated 0.5 d.p.c. At selected times, pregnant mice were killed by cervical dislocation, and the embryos were removed. Gestational age was verified according to the criteria of Theiler. Oligonucleotides were synthesized by the Yale Program for Critical Technologies and purified by OPC cartridges (Applied Biosystems, Foster City, CA). Restriction endonucleases, DNA-modifying enzymes, and polymerases were from New England Biolabs (Beverly, MA), Boehringer Mannheim, or Stratagene (La Jolla, CA). Taq DNA polymerase was from Perkin-Elmer Cetus Corp. Radionucleotides were from Amersham Corp. Digoxigenin-UTP and alkaline-phosphatase-conjugated anti-digoxigenin antibody were from Boehringer Mannheim. pBluescript II KS(−) was from Stratagene. pGEX4T3 and glutathione-agarose were from Pharmacia. Other reagents were of molecular biological grade from Sigma, Boehringer Mannheim, or U.S. Biochemical Corp.

Results

cDNA Cloning of Cux-2—At the time these studies were initiated, no mammalian cut homologues had yet been cloned. Accordingly, murine cut homologues were amplified using PCR with degenerate primers. The sequences of the primers were based on deduced amino acid sequences of Drosophila cut and mouse Cux-2.

The 5′ primer was a 256-fold degenerate mixture of 21-mers encoding the sequence WFFHNRMD, which contains four amino acids (underlined) that are conserved in all homoeodomains. Amplification of mouse genomic DNA yielded products of 300, 135, and 100 bp (not shown). Since the expected size was 135 bp based on the Drosophila sequence, the 135-bp PCR products were cloned into blue- 

sequence and sequenced. Of 26 independent transformants, six were identical to Cux-1 which was cloned independently by Valarché et al. (13). Four transformants contained a novel cut-related sequence, which we named Cux-2. 16 transformants contained inserts that were not related to cut and were not characterized further. The Cux-2 PCR product was used to screen two brain cDNA libraries, and 14 independent clones were obtained. Additional cDNA clones derived from the 5′ end of the transcript were obtained using ligation-anchored PCR. Figure 1 shows the nucleotide and deduced amino acid sequence of murine Cux-2.3 There was a single, long open reading frame that was predicted to encode a protein comprised of 1426 amino acids with an estimated molecular mass of 155 kDa, which was similar to the predicted sizes of murine Cux-1 (144 kDa) and human CDP (165 kDa). The initiation codon was identified by the presence of a Kozak (25) consensus sequence and an upstream, in-frame termination codon. At the 3′ end, we obtained several clones that were identical in the region of overlap but diverged thereafter. One clone had the sequence shown in Fig. 1 which was followed by a poly(A) tail. Two other clones contained the sequence shown in Fig. 1 followed by distinct A-rich sequences: 5′-(A)₉(CAAAATCCCCC-3′) or 5′-AAAAACCTTT(A)₉ GTTAAAAAGCCAAATCGAGGGAGAGCTGGATG(T)-3′. Since none of these sequences contained a conserved polyadenylation signal, it was likely that the cloned 3′-untranslated region was incomplete. Overall, the nucleotide sequence of murine Cux-2 was 65% identical to murine Cux-1. The deduced amino acid sequence of Cux-2 was 48% identical (63% similar) to murine Cux-1 and 29% identical (50% similar) to Drosophila Cut. In contrast, the amino acid sequence identity among the orthologues of Cux-1 in mouse, human, rat, and dog was 81–96% (86–97% similarity).

Near the carboxyl terminus, Cux-2 contained a homeodomain that was structurally similar to murine Cux-1 and Drosophila Cut. The homeodomain of Cux-2 was 75% identical (79% similar) to murine Cux-1 and 47% identical (61% similar) to Drosophila Cut. In contrast, the homeodomain of Cux-2 was less similar (33% identical) to the archetypal Antennapedia sequence and Hoxb-7 (not shown).

Important structural features of the 61-amino acid homeodomain include four amino acids that are invariant (Trp-48, Phe-49, Asn-51, and Arg-53) (1). These residues reside in the so-called recognition helix which interacts with the major groove of DNA (Fig. 2, helix 3). The homeodomain of Cux-2 contained these four invariant amino acids (indicated by ▼). Another eight positions are highly conserved but not invariant and include position 5 (Arg or Gly), 12 (usually Gln), 16 (Leu or Val), 20 (Phe or Tyr), 40 (Leu or Asn), 45 (Ile or Val), 55 (Lys or Arg), and 57 (Lys or Arg) (1). At these conserved positions (indicated by ▼ in Fig. 2), the sequences of Cux-2, Cux-1, and Drosophila Cut were identical or differed by a conservative amino acid substitution. In addition, Cux-2 contained a histidine at position 50 of the homeodomain, which was unique to Cut and Cux-1.

In addition to the Cut-related homeodomain, Cux-2 contained three 60-amino acid internal repeats (Cut repeats) that

3 The nucleotide and amino acid sequences described in this report have been deposited in the GenBank™ data base (accession no. U56665).
were previously identified in Cux-1 and Drosophila Cut. Fig. 2 shows that the first Cut repeat of Cux-2 was 58% identical to Drosophila Cut and 88–91% identical to Cux-1. The second Cut repeat was 68% identical to Drosophila Cut and 98% identical to Cux-1. The third Cut repeat was 68% identical to Drosophila Cut and 97–98% identical to Cux-1. Sequence similarity between Cux-2, Cux-1, and Drosophila Cut was highest near the carboxyl-terminal ends of the Cut repeats. As shown in Fig. 2, sequence similarity extended beyond the borders of the Cut repeats, as originally defined by Blochlinger et al. (8), to include the immediately adjacent regions. The relative spacing between the Cut repeats and the Cut homeodomain was also similar in Cux-2 and Cux-1. In contrast, within Cux-2 the amino acid identity between the first, second, and third Cut repeats was only 60–67%.

With the exception of the extended Cut repeats and Cut homeodomain, the amino acid sequence of murine Cux-2 was generally not very similar to murine Cux-1 (33% amino acid identity). Exceptions are indicated by dotted underlined sequence in Fig. 1 and included two short regions of 16 and 13 amino acids between the first and second Cut repeats which were 94 and 85% identical to Cux-1. A highly cationic region between the third Cut repeat and Cut homeodomain (boldface letters in Fig. 1) was also conserved (79% amino acid identity) and could represent a nuclear localization signal (26). A sequence of 24 amino acids at the carboxyl terminus of Cux-2 was 79% identical to Cux-1. However, Cux-2 lacked the highly alanine/proline-enriched carboxyl-terminal domain present in Cux-1 that structurally resembles repressor domains in some transcription factors. The amino terminus of Cux-2 was also.

**Fig. 1.** Nucleotide and deduced amino acid sequence of murine Cux-2. Nucleotide and amino acid positions are numbered on the right. Underlined nucleotides indicate an upstream, in-frame termination codon. Shaded box encloses the homeobox. Open boxes enclose the Cut repeats. Dotted underlined amino acids indicate additional regions that are similar to murine Cux-1 and human CDP. Amino acids in bold indicate a potential nuclear localization signal that is conserved in Cux-1. * indicates the termination codon.
similar (78% identical) to the amino terminus of human CDP (the sequence of murine Cux-1 is not available in this region).

Expression of Cux-2 in the Adult Mouse—The expression of Cux-2 in the adult mouse was examined using Northern blot analysis. Fig. 3 shows that the murine Cux-2 transcript was 7.2 kb and was expressed only in the adult brain. The transcript was also detected in the heads of neonatal mice, presumably reflecting expression in the neonatal brain (see below). In contrast, no Cux-2 transcripts were detected in any other tissue examined, whereas transcripts encoding GAPDH were detected in all tissues. In addition to the 7.2-kb transcript, a minor 5-kb transcript was also observed in adult brain. It was not known whether this transcript represented a splice variant, a homologous transcript, or was due to mRNA degradation.

We examined the expression of Cux-2 in the adult brain in greater detail using ISH. These experiments were performed using riboprobes that were labeled with $^{35}$S to increase sensitivity of detection. The hybridization probe was derived from cDNA sequence encoding the homeodomain and carboxyl-terminal domain of Cux-2. This probe did not cross-hybridize to Cux-1 which is highly expressed in a variety of non-neural tissues including kidney and lung (18). Fig. 4B is a coronal section of adult brain showing abundant expression of Cux-2 in several regions of the brain. Highest expression of Cux-2 was observed in the thalamus and limbic system. In the thalamus Cux-2 was highly expressed in specific nuclei including the laterodorsal (ldvl), ventroposterior (vp), mediodorsal (mdl), and centromedian (cm) nuclei. Within the limbic system, Cux-2 was highly expressed in the hippocampus proper (h), dentate gyrus (dg), and amygdala (am). Lower levels of Cux-2 expression were detected in the cerebral cortex (co) and external capsule (ec). Fig. 4C shows that hybridization was specific since no appreciable signal was produced using sense Cux-2 riboprobes.

To determine the cell type in brain that expressed Cux-2, higher resolution studies were performed using digoxigenin-labeled Cux-2 riboprobes. Fig. 5A shows a coronal section through the hippocampal formation in which abundant reaction product was detected in the cytoplasm of neurons in the pyramidal cell layer (pc) of the hippocampus proper and the granule cell layer (gc) of the dentate gyrus. Labeling of pyramidal cells was seen in all three fields (CA1, CA2, and CA3) of the hippocampus proper (not shown). Hybridization signal was also observed in basket cells and interneurons. In addition, a subpopulation of neurons in the thalamus (th) was positive. In contrast, there was no hybridization signal in neuroglia, as seen most clearly in fiber bundles in the alveus (alv) and fimbria (fi) of the hippocampus. As well, no expression of Cux-2

---

**Fig. 2. Alignment of the deduced amino acid sequences of the three Cut repeats (upper panel) and homeodomains (lower panel) of Drosophila Cut, murine Cux-2, and the Cux-1 orthologues from mouse (m), human (h), dog (c), and rat (r).** The sequence of Drosophila Cut is from Blochlinger et al. (8), and the sequences of Cux-1 are from Valarche et al. (13), Neufeld et al. (12), Andres et al. (14), and Yoon and Chikaraishi (15). Amino acid positions are numbered on the right. Highlighted positions indicate amino acids that are not identical in at least three sequences. Underlined amino acids indicate the Cut repeats as originally defined by Blochlinger et al. (8). The published sequences of the canine and rat orthologues of Cux-1 are partial-length and do not include the complete sequence of the first Cut repeat. Upper panel: ●, positions that are identical in all three Cut repeats; ○, positions that are identical within a particular Cut repeat; ▲, positions that are identical except for one sequence. Lower panel: ●, positions that are invariant in homeodomains; ▲, positions that are highly conserved in homeodomains (1). Bars indicate the positions of putative $\alpha$-helices as inferred from the structure of the engrailed homeodomain. Arrow indicates the His residue that is present at position 50 in all Cut-related homeodomains.

---

**Cux-2 was observed in the thalamus and limbic system.** In the thalamus Cux-2 was highly expressed in specific nuclei including the laterodorsal (ldvl), ventroposterior (vp), lateroposterior (lpmr), mediodorsal (mdl), and centromedian (cm) nuclei. Within the limbic system, Cux-2 was highly expressed in the hippocampus proper (h), dentate gyrus (dg), and amygdala (am). Lower levels of Cux-2 expression were detected in the cerebral cortex (co) and external capsule (ec). Fig. 4C shows that hybridization was specific since no appreciable signal was produced using sense Cux-2 riboprobes. To determine the cell type in brain that expressed Cux-2, higher resolution studies were performed using digoxigenin-labeled Cux-2 riboprobes. Fig. 5A shows a coronal section through the hippocampal formation in which abundant reaction product was detected in the cytoplasm of neurons in the pyramidal cell layer (pc) of the hippocampus proper and the granule cell layer (gc) of the dentate gyrus. Labeling of pyramidal cells was seen in all three fields (CA1, CA2, and CA3) of the hippocampus proper (not shown). Hybridization signal was also observed in basket cells and interneurons. In addition, a subpopulation of neurons in the thalamus (th) was positive. In contrast, there was no hybridization signal in neuroglia, as seen most clearly in fiber bundles in the alveus (alv) and fimbria (fi) of the hippocampus. As well, no expression of Cux-2
was observed in ependymal cells lining the third ventricle (not shown). Fig. 5B shows a higher magnification image of the limbic lobe in which Cux-2 was expressed in neurons in the piriform cortex (pir), amygdaloid nuclei (am), and thalamus (th). Hybridization signal was specific since there was no signal produced using sense probes (Fig. 5C).

Expression of Cux-2 in Mouse Embryos—The expression of Cux-2 was studied in mouse embryos using ISH with 32P-labeled riboprobes. Since Cux-2 was expressed in the adult forebrain, embryos were examined at 13.5 and 16.5 d.p.c., stages at which telencephalic development has commenced. Fig. 6 shows that embryonic expression of Cux-2 was restricted to neural tissue in the developing central and peripheral nervous systems. Fig. 6B shows a sagittal section of an embryo in which arrows indicate that the highest levels of expression of Cux-2 were in the telencephalon (te) and trigeminal ganglion (V). The positive signal over the crown of the embryo represents an edge artifact. Lower levels of expression were detected in spinal cord (sc), as seen more clearly in the coronal section in Fig. 6E. Fig. 6E also confirms the high level of Cux-2 expression in trigeminal ganglia (V), lower levels of expression in the pons (p), and absence of expression in olfactory bulbs (ob). Cux-2 was also detected in peripheral ganglia of the glossopharyngeal (IX) nerve. Fig. 6H shows a higher magnification image of a coronal section through the telencephalic vesicles at 16.5 d.p.c. Highest expression of Cux-2 was in the surface of the cortical plate (arrow), a cell layer of the developing telencephalon that contains differentiating, postmitotic neurons. Much lower levels of Cux-2 were detected in the ventricular and subventricular zones (vz) which contain proliferating neuroblasts. Absence of signal using sense Cux-2 riboprobes verified the specificity of hybridization in these structures (Fig. 6, C, F, and I).

DNA-binding Properties of Cux-2—The in vitro DNA-binding properties of Cux-2 were examined using EMSA. Previous studies indicated that bacterial fusion proteins containing the third Cut repeat and homeodomain of CDP could mediate sequence-specific binding to sites that were identical to those recognized by the native protein (27, 28). Accordingly, we examined the DNA-binding properties of glutathione S-transferase fusion proteins containing the carboxyl-terminal Cut repeat and homeodomain of Cux-2. Since the in vivo gene targets of Cux-2 were unknown, we studied binding to an upstream regulatory element of the neural cell adhesion molecule gene (Ncam), which had previously been identified as a Cux-1 target and was also expressed in neural tissue (13). Oligonucleotides were synthesized containing the sequence of element a (nucleotides –589 to –563 with respect to the translation start site) which previous studies showed was important for cell type-specific promoter activity and contained a functional Cux-1 binding site (13, 22). Fig. 7 shows that GST-Cux-2 fusion pro-
proteins could bind to the Ncam promoter sequence in a concentration-dependent manner (lanes 1 and 2). Binding was not observed in reactions containing bovine serum albumin or GST alone as negative controls (lanes 3 and 4). Binding was specific since formation of DNA-protein complexes was unaffected by excess poly(dI-dC)/poly(dI-dC) as nonspecific competitor (lanes 5 and 6) but was significantly reduced by a 5-fold molar excess of unlabeled Ncam oligonucleotide (lane 8) and was completely eliminated by a 10-fold molar excess of unlabeled oligonucleotide (lane 9). Retarded complexes appeared as a doublet of radiolabeled bands (arrowheads labeled B in Fig. 7). The etiology of the two bands was unclear, but was unlikely to be due to protein dimerization, since there was no effect of protein concentration on their relative abundance.  

To verify the specificity of the DNA-protein interaction, EMSA was repeated using an oligonucleotide containing point mutations of element a. The sequence of the mutated oligonucleotide was identical to that previously shown to abolish binding to Cux-1 (13). Fig. 8 shows that binding of the Cux-2 fusion protein to the mutated Ncam oligonucleotide was minimal compared with the wild-type sequence (lanes 1 and 2 compared with lanes 6 and 7). A faint band was seen when the Cux-2 fusion protein was incubated with the mutated Ncam oligonucleotide (lanes 1 and 2) but was easily disrupted by an equimolar amount of unlabeled wild-type Ncam oligonucleotide (lane 3). Moreover, a 50-fold molar excess of unlabeled mutated Ncam oligonucleotide did not significantly reduce binding to the wild-type Ncam sequence (lane 8 versus lanes 6 and 7). Thus, the mutated Ncam oligonucleotide was unable to strongly bind to the Cux-2 fusion protein either as labeled

**FIG. 5. Expression of Cux-2 in the limbic system.** Panel A, bright-field illumination of coronal section of the hippocampal formation following ISH with antisense digoxigenin-coupled Cux-2 riboprobes. Transcripts were detected using alkaline-phosphatase-conjugated anti-digoxigenin antibody as described under "Experimental Procedures." so, stratum oriens; pc, pyramidal cell layer; sl, stratum lucidum; sr, stratum radiatum of hippocampus proper; mo, molecular layer; gc, granule cell layer; po, polymorphic layer of the dentate gyrus; alv, alveus; fi, fimbria of the hippocampus. Arrow, basket cell; arrowhead, interneuron. Panel B, coronal section of the limbic lobe following ISH with antisense digoxigenin-coupled Cux-2 riboprobe. th, thalamus; am, amygdaloid nuclei; pir, piriform cortex. Panel C, bright-field illumination (Nomarski interference optics) of the limbic lobe following ISH with sense Cux-2 riboprobe. Bar, 100 µm.
probe or unlabeled competitor. Taken together, these results indicate that the GST-Cux-2 fusion protein recognizes and binds to an Ncam promoter element in a sequence-specific manner.

Chromosomal Localization of Cux-1 and Cux-2—Chromosomal mapping of murine Cux-1 and Cux-2 was performed using The Jackson Laboratory interspecific backcross panel BSB (23). To date, more than 600 loci that span the genome have been mapped on this panel which can, therefore, be used to identify the location of new genes anywhere in the mouse genome. We first identified restriction fragment length polymorphisms (RFLPs) between an inbred strain of Mus musculus (C57BL/6J) and the wild mouse, M. spreitus. Fig. 9A shows that a Cux-1-specific probe hybridized to TaqI restriction fragments that were 4.8 kb in C57BL/6J and 11.0 and 2.8 kb in M. spreitus. Invariant restriction fragments of 3.5 and 2.1 kb were detected in both species. A Cux-2-specific probe hybridized to PstI restriction fragments that were 3.8 kb in C57Bl/6J and 3.9 and 3.2 kb in M. spreitus. An invariant 1.0-kb fragment was detected in both species. The segregation of the TaqI and PstI RFLPs was followed in an interspecific backcross between (C57BL/6J × M. spreitus)F1 hybrid females and C57BL/6J males. 94 N2 backcross progeny were genotyped, and the results were compared to loci previously scored in this mapping panel (23, 29). Fig. 9B shows that Cux-2 (gene symbol Cutl2) was linked to D5Bir20 and D5Mit30, microsatellite markers which were previously assigned to distal chromosome 5. Two crossovers were detected among the 86 progeny genotyped for both Cutl2 and D5Bir20, and one crossover was detected among 92 progeny that were informative for Cutl2 and D5Mit30. Cux-1 (gene symbol Cutl1) was also located on distal chromosome 5 in the interval between D5Mit63 and Nfe2u (encoding the ubiquitous subunit of nuclear factor erythroid 2): among 94 meioses, one crossover was detected between Cutl1 and D5Mit63, four crossovers were detected between Cutl1 and Nfe2u, and no recombination was detected between Cutl1 and either D5Mit33 or Azgp (encoding MHC class I-like Zn-$\alpha_2$-glycoprotein). LOD scores in support of linkage to chromosome 5 were 25.3 and 28.3 for Cutl2 and Cutl1, respectively. Gene order was determined by minimization of double crossovers and verified by

---

FIG. 6. Expression of Cux-2 in mouse embryos. Panels A–C, bright-field (A) and dark-field (B and C) illumination of sagittal sections of a mouse embryo at 13.5 d.p.c. following ISH with antisense (A and B) or sense (C) 32P-labeled Cux-2 riboprobes. Sections were coated with emulsion and exposed for 7 days. sc, spinal cord; cb, cerebellum; p, pons; m, midbrain; di, diencephalon; lv, lateral ventricle; te, telencephalon; ge, ganglionic eminence; ob, olfactory bulb; V, trigeminal ganglion. Bar, 1 mm. Panels D–F, bright-field (D) and dark-field (E and F) illumination of a coronal section through the head of a 13.5 d.p.c. embryo following ISH with antisense (D and E) or sense (F) 32P-labeled Cux-2 riboprobes. IX, glossopharyngeal ganglion; pt, pituitary; e, eye; ob, olfactory bulb; 3rd, third ventricle; p, pons; sc, spinal cord; V, trigeminal ganglion. Bar, 1 mm. Panels G–I, bright-field (G) and dark-field (H and I) illumination of a coronal section through the telencephalic vesicles at 16.5 d.p.c. following ISH with antisense (G and H) or sense (I) 32P-labeled Cux-2 riboprobes. 3rd, third ventricle; di, diencephalon. mZ, marginal zone; cp, cortical plate; iz, intermediate zone; vz, ventricular and subventricular zones of telencephalon. Arrow in H indicates hybridization in the surface of the cortical plate. Bar, 1 mm.
maximum likelihood analysis. Fig. 9C shows that the most likely gene order and recombination distances (cM ± standard error) were: cen . . . D5Bir20–2.3 ± 1.6–Cutl2–1.1 ± 1.1–D5Mit30–1.1 ± 1.1–D5Mit63–4.2 ± 2.1–Cutl1, Aegp, D5Mit33–4.2 ± 2.1–Nfe2u. Six crossovers were detected among 92 progeny genotyped for both Cutl2 and Cutl1, and the 95% confidence limit of the interval between these two loci was 2.4–13.7 cM.

DISCUSSION

Previous studies have identified a murine gene, named Cux-1, that was homologous to the Drosophila homeobox-containing gene cut (13). Cut homologues have also been cloned from human, dog, and rat (12, 14, 15). Sequence comparisons indicated that these Cut homologues were encoded by orthologous genes that were derived from speciation events. In this study, we now report the cloning of a second, novel murine cut homologue named Cux-2. The sequence comparisons between Cux-2 and the other mammalian cut homologues and the results of chromosomal mapping demonstrate that Cux-2 is a paralogous gene which was derived from a gene duplication event. The identification of two paralogous genes within the same species verifies the existence of a family of cut-related homeobox genes in mammals.

Cux-2 shares important structural features with Cut and the other mammalian homologues that further define this family of homeoproteins. Among members of this class, the homeodomain is highly conserved but is divergent from other classes of homeoproteins. The homeodomain of Cux-2 shares 47% amino acid sequence identity with Drosophila Cut and 75% identity with Cux-1 but only 33% identity with Antp-type homeodomains of Drosophila or mammals. The homeodomains of Cux-2, Cux-1, and Cut contain the four invariant amino acids that are found in all homeodomains but in addition contain a unique histidine residue at position 50. Based on x-ray crystallographic studies of the engraved homeodomain (30), homeodomains are believed to adopt a helix-turn-helix structure in which one α-helix, called the recognition helix (Fig. 2, helix 3), inserts in the major groove of DNA. Certain amino acid side chains protruding from helix 3 are therefore positioned to directly interact with specific bases in DNA, which confers sequence specificity. In particular, the amino acid at position 50 of the homeodomain (position 9 of the recognition helix) contacts bases that are immediately adjacent to the core ATTA sequence found in many homeodomain binding sites and appears to be primarily responsible for differences in specificity between distantly related homeodomains (31). The amino acid at position 50 is a Gln in the Antennapedia class of homeodomain proteins, a Ser in the paired family, and a Cys in the POU family. Although a His was identified at position 50 of the mammalian Cut homologues cloned previously, these were encoded by orthologous genes, so it was unclear how general this finding would be. The studies of Cux-2, a distinct mammalian Cut parologue, indicate that a His at position 50 of the homeodomain is a defining feature of the Cut family and suggest that this class may have unique DNA-binding properties that are different from other homeodomain proteins. In addition to the Cut-related homeodomain, Cux-2 contains three internal 60-amino acid repeats called Cut repeats. The Cut repeats of Cux-2 shared 88–98% amino acid identity with Cux-1 and 58–68% identity with Drosophila Cut demonstrating that a higher degree of amino acid similarity exists between the Cut repeats than between the Cut homeodomains. This high degree of evolutionary conservation supports the hypothesis that Cut repeats are functionally important domains of the protein. Indeed, recent studies indicate that Cut repeats can function as independent DNA-binding domains that recognize a sequence that contains a 5′-ATCGAT-3′ core (27, 28, 32).

The expression pattern of Cux-2 is markedly different from Cux-1, which is expressed in most adult tissues (13). In embryos, Cux-1 transcripts are most highly expressed in the nervous system, kidney, and lung (18), and the protein has been detected in multiple cell lineages (14). In contrast, expression of Cux-2 was restricted to the nervous system in both the developing and adult mouse. The expression of Cux-2 in the telencephalon was of interest since relatively few homeobox genes are expressed in the forebrain. For example, no Hox genes are
expressed anterior to the border between rhombomeres 2 and 3 of the hindbrain (33). Recently, diverged homeobox genes have been identified such as Dlx, Emx, and Otx (34) that are expressed in the developing forebrain. Whereas expression of these diverged homeobox genes is spatially restricted in distinct anatomical regions, Cux-2 was widely expressed in the developing central and peripheral nervous systems, which was more consistent with a role in generation or maintenance of neuronal differentiation rather than regional specification. In particular, Cux-2 was highly expressed in certain neural tissues that are involved in processing sensory information including the ventroposterior thalamic nuclei, piriform cortex, amygdala, and peripheral ganglia of the glossopharyngeal and trigeminal nerves. In comparison, cut is essential for the development of external sensory organs comprising the peripheral nervous system of Drosophila. Hence, the observations that Cux-2 was restricted to neural tissue and was highly expressed in components of the sensory nervous system raise the possibility that the function of Cut in specifying neural identity may be phylogenetically conserved in Drosophila and mammals. Other examples of diverged homeobox genes that have phylogenetically conserved functions in organisms as disparate as Drosophila and mouse have recently been described (5).

Homology to Cux-1 suggested that Cux-2 was also a transcriptional regulatory protein. However, despite the existence of full-length cDNA clones, we have been unable to detect transcriptional activation or repression by Cux-2. As an alternative, we examined the in vitro DNA-binding properties of Cux-2 using electrophoretic mobility-shift assays. A bacterial fusion protein containing the carboxyl-terminal Cut repeat and homeodomain of Cux-2 could bind to an upstream regulatory
element of the neural cell adhesion molecule (Ncam) gene, which had previously been shown to be a target for transcriptional repression by Cux-1 (13). Moreover, an element containing point mutations that abolished binding to Cux-1 was unable to bind to the Cux-2 fusion protein either as radiolabeled probe or as unlabeled competitor. These results indicated that Cux-2 encoded a sequence-specific DNA-binding protein. However, these studies do not address whether Ncam is a direct target for transcriptional regulation by Cux-2 in vivo. During late gestation (12.5 d.p.c.) Ncam is highly expressed in postmitotic neurons in the outer neural tube but is not expressed in proliferating cells in the ventricular and subventricular zones (35). At a comparable stage of development (13.5 d.p.c.), Cux-2 was highly expressed in the cortical plate of the developing telencephalon which contains postmitotic neurons; there were only low levels of expression in the ventricular and subventricular zones. The observation that the patterns of expression of Cux-2 and Ncam were overlapping and not reciprocal suggested that Ncam was not a direct target for transcriptional repression by Cux-2 in vivo or that Cux-2 functioned instead as a transcriptional activator of Ncam.

The chromosomal loci that encode Cux-1 and Cux-2 have been designated Cut-like 1 (Cutl1) and Cut-like 2 (Cutl2), respectively. In this study, we used an interspecific backcross panel to determine the localizations of Cutl1 and Cutl2 in the mouse genome. We found that Cutl1 and Cutl2 mapped to distinct loci that were genetically linked on distal chromosome 5. The mapping to different loci confirmed that Cux-1 and Cux-2 were products of distinct genes rather than, e.g., splice variants of a single gene. Previously, the human Cutl1 homologue (CUTL1) was mapped to the long arm of chromosome 7 (7q22) by analysis of somatic cell hybrids (36). The localization of Cutl1 was consistent with these previous studies since the region of mouse chromosome 5 between β-glucuronidase (Gus) and mitochondrial malate dehydrogenase (Mor1), which encompasses Cutl1, is syntenic with human chromosome 7q21-q22 (37). Cutl2 was also located near Cutl1 suggesting that distal Chr 5 may contain a cluster of homeobox genes that are related to Drosophila cut. The observation that Cutl1 and Cutl2 were genetically linked was unexpected since all previously identified members of families of diverged homeobox genes in mammals are dispersed, e.g. En-1 and En-2 on mouse chromosomes 1 and 5; Eox-1 and Eox-2 on mouse chromosomes 6 and 2; Pax-3, Pax-4, and Pax-6 on mouse chromosomes 1, 6, and 2, respectively. Cux-1 and Cux-2 may have arisen by cis duplication within a chromosome by unequal sister chromatid exchange as suggested for Hox genes (38). No other genes that are linked to Cutl1 or Cutl2 are known to be duplicated in this region. A search of the mouse genome data base did not reveal any known mutations in the region of the Cutl2 locus that could be readily associated with abnormalities of Cux-2. Although the long arm of human chromosome 7 contains a gene for familial holoprosencephaly, a developmental field defect involving the forebrain, autosomal dominant holoprosencephaly has recently been mapped more precisely to 7q36 which appears to reside outside of the region of known synteny with distal mouse chromosome 5 (39).

In summary, we have identified a second murine cut homologue named Cux-2. The pattern of expression and DNA-binding assays suggest a potential role for Cux-2 as a transcriptional factor that is important in the development of the murine nervous system. Moreover, the high degree of sequence conservation of the Cut homeomain and Cut repeats, and the existence of at least two distinct homologues within the same species, indicate that Cux-2 and Cux-1 represent a new family of homeoproteins.

Acknowledgments—We thank Jacqueline Kusnezov and Kui Li for expert technical assistance, and Michele Pucci for expert secretarial assistance. We thank Michael Brines and Susan Hoekfield for assistance with neuroanatomical localization, Lucy Rowe and Ed Birkenmeier for advice onchromosomal mapping, and Frank Ruddle and members of the Ruddle laboratory for many helpful discussions.

REFERENCES

1. Scott, M. P., Tamkun, J. W., and Hartzell, G. W., III (1989) Biochim. Biophys. Acta 989, 24–48
2. Kappen, C., Schughart, K., and Ruddle, F. H. (1993) Genomics 18, 54–70
3. Duboule, D. (1994) Guidebook to the Homeobox Genes, Oxford University Press, Oxford
4. McGinnis, W., and Krumlauf, R. (1992) Cell 68, 283–302
5. Sefton, B., and Scott, M. P. (1994) Dev. Suppl., 61–71
6. Bodmer, R., Barbel, S., Sperber, S., Jack, J. W., Jan, L. Y., and Jan, Y. N. (1987) Cell 51, 293–307
7. Blochlinger, K., Jan, L. Y., and Jan, Y. N. (1991) Genes Dev. 5, 1124–1135
8. Blochlinger, K., Bodmer, R., Jack, J. W., Jan, L. Y., and Jan, Y. N. (1988) Nature 333, 629–635
9. Blochlinger, K., Bodmer, R., Jan, L. Y., and Jan, Y. N. (1990) Genes Dev. 4, 1329–1331
10. Liu, S., McLeod, E., and Jack, J. (1991) Genetics 127, 151–159
11. Blochlinger, K., Jan, L. Y., and Jan, Y. N. (1995) Development 117, 441–450
12. Neufeld, E. J., Skalnik, D. G., Lievens, P. M. J., and Orkin, S. H. (1992) Nat. Genet. 1, 50–55
13. Valarche, I., Tissier-Seta, J.-P., Hirsch, M.-R., Martinez, S., Goridis, C., and Brunet, J.-F. (1995) Development 119, 881–896
14. Andres, V., Nadal-Ginard, B., and Mahdavi, V. (1992) Development 116, 321–334
15. Yoon, S. O., and Chikaraishi, D. K. (1994) J. Biol. Chem. 269, 18453–18462
16. Suguira-Purga, C., Barberis, A., Schaffner, G., and Busslinger, M. (1998) EMBO J. 7, 3099–3107
17. Skalnik, D. G., Strauss, E. C., and Orkin, S. H. (1991) J. Biol. Chem. 266, 16736–16744
18. Vanden Heuvel, G. B., Bodmer, R., McConnell, K. R., Nagami, G. T., and Igarashi, P. (1996) Kidney Int., 50, 453–461
19. Quaggin, S. E., Bodmer, R., and Igarashi, P. (1994) Mol. Biol. Cell 5, 85a
20. Igarashi, P., Vanden Heuvel, G. B., Payne, J. A., and Forbush, B., III (1995) Am. J. Physiol. 269, F405–F418
21. Fort, P., Marty, L., Piechaczyk, M., El Sabrouty, S., Dani, C., Jeanteur, P., and Blanchard, J. M. (1995) Nucleic Acids Res. 23, 1431–1442
22. Hirsch, M. R., Mauger, L., Deagostini-Bazin, H., Bally-Cuif, L., and Goridis, C. (1990) Mol. Cell Biol. 10, 1959–1968
23. Fort, P., Marty, L., Piechaczyk, M., En-Sabrouty, S., Dani, C., Jeanteur, P., and Blanchard, J. M. (1995) Mol. Cell Biol. 15, 129–140
24. Zambra, L., Bierroé, G., Tampin, O. J., Denis-Larose, C., and Nepveu, A. (1995) Mol. Cell Biol. 15, 129–140
25. Auduero, B., Neufeld, E. J., and Orkin, S. H. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 7577–7581
26. Peters, L. L., and Eicher, E. M. (1994) Genes Dev. 209, 490–491
27. Kessenich, C. K., Nikriesser, D., and Ruddle, F. H. (1993) Genes Dev. 8, 245–257
28. Hunt, P., Whiting, J., Muchamore, I., Marshall, H., and Krumlauf, R. (1991) Dev. Suppl., 187–196
29. Finkelstein, R., and Boncinelli, E. (1994) Trends Genet. 10, 310–315
30. Bally-Cuif, L., Goridis, C., and Santoni, M. J. (1992) Development 117, 543–552
31. Scherer, S. W., Neufeld, E. J., Lievens, P. M., Orkin, S. H., Kim, J., and Tsui, L. C. (1993) Genomics 15, 695–696
32. Kozak, C. A., Goffinet, A., and Stephenson, D. A. (1994) Mamm. Genome 5, 865–878
33. Kappen, C., Schughart, K., and Ruddle, F. H. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 5459–5463
34. Muenke, M., Gorgiourri, F., Bay, C., Yi, D. H., Collins, A. L., Johnson, V. P., Hengst, R. C., M. M., Schaffner, G. B., Wekh, L., Lubinsky, M. S., Daack-Hirsch, S., Moore, C. A., Dobyns, W. B., Murray, J. C., and Price, R. A. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 8102–8106

Downloaded from http://www.jbc.org/ by guest on July 24, 2018
Primary Structure, Neural-specific Expression, and Chromosomal Localization of 
Cux-2, a Second Murine Homeobox Gene Related to Drosophila cut 
Susan E. Quaggin, Gregory B. Vanden Heuvel, Krista Golden, Rolf Bodmer and Peter Igarashi 

J. Biol. Chem. 1996, 271:22624-22634. 
doi: 10.1074/jbc.271.37.22624

Access the most updated version of this article at http://www.jbc.org/content/271/37/22624

Alerts: 
- When this article is cited 
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 36 references, 15 of which can be accessed free at 
http://www.jbc.org/content/271/37/22624.full.html#ref-list-1