Novel Target Sequences for Pax-6 in the Brain-specific Activating Regions of the Rat Aldolase C Gene*

Received for publication, September 12, 2002
Published, JBC Papers in Press, October 4, 2002, DOI 10.1074/jbc.M209349200

Henriette Skala-Rubinson, Joëlle Vinh†, Valérie Labas‡, Axel Kahn, and Françoise Phan Dinh Tuy§

From the Département de Génétique, Développement et Pathologie Moléculaire, Institut Cochin, INSERM, CNRS, Université René Descartes, 24, rue du faubourg Saint-Jacques, 75014 Paris and the Neurobiologie et Diversité Cellulaire, CNRS, Ecole Supérieure de Physique et de Chimie Industrielles, 10 rue Vauquelin, 75231 Paris, cedex 05, France

Upstream activating sequences of the rat aldolase C gene are shown here to confer brain-specific expression in transgenic mice. In addition to binding sites described previously for the brain-expressed POU proteins Brn-1 and Brn-2 (Skala, H., Porteu, A., Thomas, M., Szajnert, M. F., Okazawa, H., Kahn, A., and Phan-Dinh-Tuy, F. (1998) J. Biol. Chem. 273, 31806–31814), we have identified two novel DNA elements critical for an interaction with a brain-specific, high affinity DNA-binding protein. Characterization of this binding protein showed it to be sensitive to thiol oxidation and stable to heat at 100 °C. This protein was purified on the basis of its thermostability and its selective adsorption to streptavidin magnetic particles via a biotinylated multimer of its target DNA binding site. Liquid chromatography coupled to tandem mass spectrometry analysis, binding competition with consensus oligonucleotides, and antibody supershift assays led to its identification as the homeodomain paired protein Pax-6. This result suggests that the brain-specific aldolase C gene could constitute a new target for the transcription factor Pax-6, which is implicated increasingly in neurogenesis.

Regulation of gene transcription in active chromatin is dependent upon the interaction of discrete DNA motifs with regulatory proteins. These proteins either represent basal, ubiquitous elements of the transcriptional machinery or can be themselves expressed in a tissue-, development-, or stimulus-specific manner. Cell-type diversity in the different organs can be considered as the result of multiple DNA-protein and protein-protein interactions leading to the controlled expression of a very large number of genes. The central nervous system is one of the most complex structures and consequently the one for which such regulatory mechanisms are probably still the least elucidated. In light of this, identification of DNA binding sites for transcription factors involved in brain development will help to define their candidate target genes and to elucidate the complex gene regulatory network at play in the central nervous system.

The aldolase C gene encodes a brain-specific glycolytic enzyme, the fructose-1,6 diphosphate aldolase (EC 4.1.2.13). In a previous paper (1), we reported that the transcriptional activation of the brain-specific promoter (115 bp) of the rat aldolase C gene in vivo requires a cooperation between distal and proximal upstream sequences. We identified in vitro in a 0.6-kb proximal activating fragment several POU/WH motifs (1) exhibiting overlapping binding sites for the brain-restricted class III Brn-1 and Brn-2 POU proteins (2) and for the protein HNF3β, a member of the WH (winged helix) factor family, which has been shown to be involved in the early brain development (3, 4).

In this paper, we first show that the distal and proximal regulatory sequences are also able to ensure the activation of the ubiquitous thymidine kinase (tk) promoter specifically in the brain of transgenic mice. We also report that, in addition to the POU/WH motifs, these regions contain two strong in vitro binding sites for a brain-specific, thermostable protein, one located in the 0.6-kb proximal fragment and the other in the distal activating region. The sequence of these DNA motifs shared no significant homology to binding sites for known transcription factors. We therefore undertook the purification of this brain-specific binding protein. Mass spectrometry analysis of the final protein fraction identified the transcription factor Pax-6. Competition experiments with consensus binding sites and supershift experiments with specific antibodies confirmed this identification. The protein Pax-6 is a member of the Paired family. It contains a 128-amino acid paired domain described initially in the prd (paired) gene of Drosophila melanogaster (5) and a paired-type homeodomain (6). Pax-6 has been shown to play a crucial role in the development of the central nervous system and the eye and olfactory system (7–12). It is also implicated in the development of pancreatic and gastrointestinal endocrine cells (13, 14). More recently, it has been implicated in neurogenesis, with Pax-6 mutant mice showing reduced numbers of cortical neurons (15). Our data suggest that the aldolase C gene represents a new candidate target for this important transcription factor.

EXPERIMENTAL PROCEDURES

DNA Constructs—The construct 3.8/tk/CAT includes the 3.8-kb distal activating fragment of the rat aldolase C gene (1) (Fig. 1) cloned upstream of the ubiquitous promoter (−105 to +51) of the herpes simplex virus tk gene (16) ligated to the CAT coding sequence. The 3.8 + 0.6/tk/CAT construct was prepared by inserting the 0.6-kb proximal activating fragment of the rat aldolase C gene (1) (Fig. 1) upstream of the tk promoter in the plasmid 3.8/tk/CAT. Further details on the plasmid constructs are available on request.

Production and Analysis of Transgenic Mice—The purified DNA fragments (5–10 ng) were microinjected into fertilized B6D2 mouse eggs (17). Transgenic founders and offspring were identified by Southern

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

† To whom correspondence should be addressed. Tel.: 33-1-44-41-24-12; Fax: 33-1-44-41-24-21; E-mail: tuy@cochin.insERM.fr.
blot using a random-primer CAT probe (18). CAT activities generated by transgene expression were assayed on 0.5–100 μg of extracted proteins (19).

**Oligonucleotides**—The oligonucleotides used as probes or competitors in the DNA/protein binding reactions were as follows: 28bp (nt –862 to –835 of the rat aldolase C gene, with respect to the ATG translation initiation), 5′-ATGGGTATTTAAATAAAGTGTATTTAATG-3′; site X24 (nt –824 to –822 of the rat aldolase C gene), 5′-AGTCCAAGTTTGGG-3′; Pax-6 consensus reported by Epstein et al. (20), 5′-ttgaaAGGGATTGAAGAGTCA-3′; Pax6cons (bases in uppercase letters are homologous to the human sequence), 5′-cgcACATTCAGCATTGACTGACTctcg-3′; ThBFsite (nt –834 to –829 of the rat aldolase C gene) (bases in uppercase letters are homologous to the human sequence), 5′-tgaaGGAGGATCGAATCTAATTTTGGG-3′; ThBFmut, representing the ThBFsite sequence mutated at the underlined positions –826, –825 (CC → TT), –819 (A → C), and –816 (G → A), 5′-TATGCTTTATGACCCCTATCAATTAAT-3′; ThBFdist and ThBFsite motifs/mg of beads). Magnetic beads were equilibrated by protein digestion and analysis by automated liquid chromatography-tandem mass spectrometry (LC-MS/MS). The stained gel slice corresponding to the mobility of the ThBFsite binding activity was cut out, and the proteins were reduced, alkylated, and digested by in-gel tryptic proteolysis (33). The tryptic digest was analyzed by nanoscale capillary LC/MS/MS. An UltiMate capillary liquid chromatography system (LC Packings, Amsterdam, Netherlands) was used on-line with a hybrid nanoESI quadrupole-time of flight mass spectrometer (Q-TOF; Micromass, Manchester, United Kingdom). Chromatographic separations were conducted on a reverse-phase capillary column (Pepmap C18, 75 μm inner diameter, 15-cm length; LC Packings) with a 200 nL/min flow. The gradient profile used consisted of a linear gradient from 100% A (H₂O/acetonitrile/formic acid, 96/4/0.1) (v/v) to 45% B (H₂O/acetonitrile/FA, 10/90/0.085) (v/v) in 50 min, followed by a linear gradient to 100% B in 10 min. Mass data collected during an LC-MS/MS analysis were submitted to the search software Mascot (www.matrixscience.com/). Protein identifications were obtained by comparison of experimental data to the NCBI database.

**RESULTS**

**Brain Specificity of the Aldolase C Gene Activating Regions:**

**in Vivo Analysis**—We reported previously (34) that the 115-bp promoter fragment (−199 to −84) of the rat aldolase C gene (Fig. 1) was sufficient to ensure the brain-specific expression of the CAT reporter gene in transgenic mice. However, the high level activation of this short promoter required the cooperation in vivo of a distal fragment of 3.8 kb and a more proximal fragment of 0.6 kb (each fragment alone was found unable to activate the 115-bp promoter) (1). To determine whether these activating sequences themselves confer a tissue-specificity, we cloned the distal 3.8-kb fragment alone or both activating regions upstream of the ubiquitous tk promoter fused to the CAT coding sequence. The CAT activities measured in transgenic mice are presented in Table I. The control construct tk/CAT

**FIG. 1. Scheme of the rat aldolase C gene regulatory sequences.** The 3.8-kb distal and 0.6-kb proximal activating fragments are represented upstream of the 115-bp brain-specific promoter. The already reported POU binding sites 28bp and site 824 (1) and the ThBFdist and ThBFsite motifs described in this paper are also indicated.

**Protein Digestion and Analysis by Automated Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)**—The stained gel slice corresponding to the mobility of the ThBFsite binding activity was cut out, and the proteins were reduced, alkylated, and digested by in-gel tryptic proteolysis (33). The tryptic digest was analyzed by nanoscale capillary LC/MS/MS. An UltiMate capillary liquid chromatography system (LC Packings, Amsterdam, Netherlands) was used on-line with a hybrid nanoESI quadrupole-time of flight mass spectrometer (Q-TOF; Micromass, Manchester, United Kingdom). Chromatographic separations were conducted on a reverse-phase capillary column (Pepmap C18, 75 μm inner diameter, 15-cm length; LC Packings) with a 200 nL/min flow. The gradient profile used consisted of a linear gradient from 100% A (H₂O/acetonitrile/formic acid, 96/4/0.1) (v/v) to 45% B (H₂O/acetonitrile/FA, 10/90/0.085) (v/v) in 50 min, followed by a linear gradient to 100% B in 10 min. Mass data collected during an LC-MS/MS analysis were submitted to the search software Mascot (www.matrixscience.com/). Protein identifications were obtained by comparison of experimental data to the NCBI database.

**RESULTS**

**Brain Specificity of the Aldolase C Gene Activating Regions:**

**in Vivo Analysis**—We reported previously (34) that the 115-bp promoter fragment (−199 to −84) of the rat aldolase C gene (Fig. 1) was sufficient to ensure the brain-specific expression of the CAT reporter gene in transgenic mice. However, the high level activation of this short promoter required the cooperation in vivo of a distal fragment of 3.8 kb and a more proximal fragment of 0.6 kb (each fragment alone was found unable to activate the 115-bp promoter) (1). To determine whether these activating sequences themselves confer a tissue-specificity, we cloned the distal 3.8-kb fragment alone or both activating regions upstream of the ubiquitous tk promoter fused to the CAT coding sequence. The CAT activities measured in transgenic mice are presented in Table I. The control construct tk/CAT
and the 3.8-kb fragment alone (transgene 3.8/tk/CAT) led to undetectable activities in all tissues tested. From the seven transgenic animals obtained for the 3.8 + 0.6/tk/CAT construct, two were negative in all tissues, and five exhibited a brain-specific CAT expression. In founders 26, 16, and 2, CAT levels were similar to those observed previously (34, 35) with the 115-bp aldolase C brain-specific 115-bp promoter without upstream sequences added whereas in founders 18 and 27, significantly higher levels were measured. Variation of the CAT values reflects the already described dependence upon the integration site of the transgene in chromatin (36). From these results, we can conclude that 1) the combination of the 3.8- and 0.6-kb upstream sequences activating the aldolase C brain-specific 115-bp promoter is also able to activate the heterologous ubiquitous tk promoter, and 2) the activation occurs only in brain, and hence these sequences behave as a tissue-specific activator. It is thus likely they contain DNA elements interacting with brain-specific proteins.

**In Vitro Analysis of the Proximal 0.6-kb Fragment**—We thus screened the proximal 0.6-kb fragment by *in vitro* DNase I footprinting of overlapping PCR fragments. Probes covering nucleotides −745 to −194 did not allow brain-specific protection from DNase I digestion. A more 5′-probe (nt −914 to −713) encompassed two already described motifs, the 28bp element, conserved between the rat and human aldolase C genes, and the site 824 (Fig. 1), both able to bind ubiquitous and brain-restricted proteins of the POU family (1). Indeed, using the recombinant GST-Brn-2 POU domain fusion protein, three protected regions (numbered 1 to 3) were detected on the coding strand (Fig. 2A, lanes 3 and 4). The footprints 1 (nt −864 to −844) and 3 (nt −820 to −802) corresponded to the POU binding sites of the 28bp element and of the site 824, respectively (Fig. 2B). The AT-rich sequence in the intermediate footprint 2 (nt −840 to −829) is also compatible with the binding of POU factors (Fig. 2B). According to our previous results, these three protections were abolished specifically by the oligonucleotide competitors 28bp (Fig. 2A, lane 2) and site 824 (not shown). Exactly the same protection and competition patterns were observed with the recombinant protein GST-Brn-1 POU domain (not shown). Using brain nuclear extracts, two further protected regions, numbered 4 and 5, were detected on the same PCR product (Fig. 2A, lanes 4 and 5). The strong footprint 5 spanned from nt −832 to nt −811 and was limited at its 5′ boundary by a hypersensitive site at nt −834. The fainter footprint 4 extended from nt −851 to nt −836. A competitor PCR fragment extending from nt −864 to nt −802 completely abolished footprint 5 and hypersensitive site but only attenuated footprint 4, which seems to represent a weak interaction (not shown). In contrast, the 28bp oligonucleotide was unable to prevent protein binding at both of these sites (lane 7). No footprint was detected with liver nuclear extracts (lanes 8 and 9). Thus, this DNase I footprinting analysis, summarized in Fig. 2B, enabled us to define in the aldolase C proximal activating region a novel DNA element, corresponding to footprint 5, overlapping with the POU binding sites but interacting strongly with a brain-specific protein different from Brn-1 and Brn-2. Because this protein was further shown to be stable at 100 °C (see below), we called it ThBF (thermostable brain factor).

**Characterization of the DNA-ThBF Interaction**—Further characterization of the footprint 5 DNA-protein interaction was performed by gel shift experiments. When the radiolabeled PCR probe (−864 to −802) was incubated with brain nuclear extracts, a single, very strong, fast-migrating complex was

| Transgene     | Founder | CAT activity |
|---------------|---------|--------------|
| tk/CAT        | 4, 10, 12, 14 | < 0.02     |
| 3.8/tk/CAT    | 1, 9, 11, 24 | < 0.02     |
| 3.8 + 0.6/tk/CAT | 19 | < 0.02     |
|               | 31 | < 0.02     |
|               | 26 | 0.13       |
|               | 16 | 0.33       |
|               | 2  | 8.2        |
|               | 18 | 170        |
|               | 27 | 670        |

**TABLE I**

**Brain specificity of the aldolase C gene activating regions**

CAT activities are expressed in cpm/min/μg of proteins.

| Transgene     | Founder | Brain | Liver, heart, lung |
|---------------|---------|-------|-------------------|
| tk/CAT        | 4, 10, 12, 14 | < 0.02 | < 0.02 |
| 3.8/tk/CAT    | 1, 9, 11, 24 | < 0.02 | < 0.02 |
| 3.8 + 0.6/tk/CAT | 19 | < 0.02 | < 0.02 |
|               | 31 | < 0.02 | < 0.02 |
|               | 26 | 0.13   | < 0.02 |
|               | 16 | 0.33   | < 0.02 |
|               | 2  | 8.2    | < 0.02 |
|               | 18 | 170    | < 0.02 |
|               | 27 | 670    | < 0.02 |

**FIG. 2.** DNase I footprinting pattern of the −914 to −713 fragment of the proximal activating region. A, the 3′-end-labeled upper strand was incubated with 250 ng of the recombinant GST-Brn-2 POU domain fusion protein (lanes 1–5) or 30 μg of brain nuclear extract proteins (lanes 5–7) or 60 μg of liver nuclear extract protein (lane 9), in the presence of 0.015 μg (for GST-Brn-2) or 0.08 to 0.12 μg (for nuclear extracts) of DNase I. Lanes 4 and 8 represent digestion of the probe without protein. Lane 10 is a G + A Maxam-Gilbert reaction (57). When indicated (lanes 1, 2, 6, and 7), unlabeled competitor oligonucleotides (100 ng) were included in the footprinting reactions. Protected regions are delineated by bars, and the hypersensitive site is indicated by an arrow. On the lower strand, only weak protections were observed with brain nuclear extracts (not shown). B, summary of the footprint results. The DNA sequence between nucleotides −864 and −802 from the upper strand is indicated. The protected regions and the hypersensitive site are represented by boxes and an asterisk, respectively. Limits of the oligonucleotides 28bp, site 824, and ThBFsite are indicated. Residues susceptible to methylation interference (see Fig. 3B) are also noted (circles). Solid circles represent a higher degree of interference than open circles.
detected within only 1 h of autoradiography (Fig. 3A, lanes 1 and 2). A much longer exposure allowed the detection of Oct-1 and Brn-1 binding on this probe (not shown) specifically on the 28bp element as described previously (1). The Brn-2 complex was not detectable in this experiment, because it migrated approximately at the same level as the major band. The major retarded complex was not competed by the 28bp oligonucleotide but was abolished by the unlabeled (−864 to −802) PCR fragment itself (Fig. 3A, lanes 3 and 4) and by an oligonucleotide competitor, ThBFsite, which encompasses the limits of footprint 5 (see Fig. 2B and Fig. 3A, lanes 5 and 6). When used as a labeled probe, the ThBFsite oligonucleotide interacted specifically with a protein of the same electrophoretic mobility (Fig. 3A, lanes 7 and 8), and this binding was competed reciprocally by the unlabeled PCR fragment (lane 9). In DNase I experiments, the oligonucleotide ThBFsite abolished specifically the footprint 5 and the hypersensitive site produced by brain nuclear extracts (Fig. 2A, lane 6) but did not suppress the GST-Brn-2 footprints 1–3 (Fig. 2A, lane 1). It thus seems likely that the strong DNA binding activity observed by gel shift experiments is ThBF, responsible for footprint 5.

To delineate more precisely the DNA bases involved in this interaction, we performed a methylation interference experiment using the PCR (−864 to −802) probe. Methylation at nucleotides −819 and −816 on the upper strand and −827, −826, −825, and −822 on the lower strand (Fig. 2B) interfered with protein binding (Fig. 3B). A mutated version of the ThBF-site oligonucleotide, ThBFmut, where nucleotides −816, −819, −825, and −826 were substituted, was indeed unable to compete for ThBF binding (Fig. 3A, lane 10). Fortuitously, we found that ThBF was also able to interact in vitro with a DNA motif including a sequence conserved perfectly between the rat and human aldolase C genes over 27 nucleotides (nt −4424 to nt −4390 of the rat gene), located in the subfragment A of the 3.8-kb distal activating region (Fig. 1), and termed ThBFdist. In competition experiments, the protein ThBF exhibited comparable in vitro affinities for the probes ThBFsite and ThBFdist (Fig. 3A, lanes 11–17).

We next tested for ThBF binding in different tissue and cell extracts by further gel shift assays. It was found present approximately at the same level in fetal and adult brain extracts (not shown). In extracts from several other tissues (liver, kidney, muscle, spleen, pancreas) and cell lines (HeLa, B cell, adipocyte 3T3L1), the ThBF band was undetectable (not shown). Thus, ThBF appears to be an essentially brain-restricted protein interacting strongly with at least two binding sites in the aldolase C gene activating sequences. Computer analysis of ThBFsite and ThBFdist sequences (MatInspector, TESS, TFSsearch, and PatSearch programs) failed to reveal overt homologies with binding sites for known transcription factors, and screening of a rat brain cDNA expression library using the multimerized ThBFsite as a probe did not identify positive clones. We therefore decided to characterize the biochemical properties of the protein ThBF and to undertake its purification.

**Molecular Mass, Heat Stability, and Redox Sensitivity of ThBF**—The size of ThBF was estimated by elution/renaturation of brain nuclear extract proteins separated by SDS-PAGE. Gel shift assays of the eluted fractions showed that ThBF binding was recovered from a gel slice corresponding to apparent molecular masses of 48 to 50 kDa (not shown). Testing ThBF thermostability, we found that significant binding was retained after a 7-min incubation of crude brain nuclear extracts at 100 ºC (Fig. 4A, lanes 1 and 2). In addition, after a differential ammonium sulfate precipitation (not shown) or one round of DNA affinity chromatography (see below) to remove protease(s), the binding of ThBF was found to be completely unaffected by this heating step (Fig. 4A, lanes 3 and 4).

The strong binding of ThBF to its DNA sequence target was detected in the presence of 1 mM of the chelating agent EDTA in brain nuclear extracts and was insensitive to a further EDTA concentration increase up to 25 mM (not shown). Therefore, ThBF does not resemble a zinc finger domain DNA-binding protein (37). Nuclear extracts were prepared in the presence of 1 mM dithiothreitol, most likely rendering the ThBF molecules in a reduced state. We thus examined the sensitivity of ThBF to redox conditions by assaying the DNA binding activity of a semipurified fraction (obtained after one cycle of DNA affinity chromatography) in the presence of various thiol-reactive metal ions. A strong inhibition of DNA binding was observed after preincubation in the presence of Zn(OAc)2 or CdCl2 (Fig. 4B, lanes 1–7). The retarded complex was abolished almost completely by 500 µM Zn2+ or 50 µM Cd2+ ions. In contrast, ThBF binding to DNA was found unaffected by the presence of 2 mM MnCl2 (Fig. 4B, lane 8) and 5 mM MgCl2 (in the standard binding reaction), two metal ions with lower oxidation potentials. Finally, preincubation of the protein fraction in the presence of 2 mM N-ethylmaleimide, a sulfhydryl alkylating agent, resulted in a complete inhibition of ThBF binding (Fig. 4B, lane 9). These results suggest that at least one critical thiol group is required for ThBF binding to DNA.

**Purification of the Protein ThBF**—ThBF was purified from adult rat brains as described under “Experimental Procedures.” Two cycles of DNA affinity chromatography were performed using streptavidin-magnetic particles coated with biotinylated multimers of the ThBFsite motif. Taking advantage of the thermostability of ThBF, we included a 100 ºC incubation step after the first elution from the DNA beads. Streptavidin-binding proteins were removed by the addition of non-DNA-coated beads in the preincubation step. As multimerization of the ThBFsite oligonucleotide resulted in the creation of AT-rich junction motifs able to bind POU proteins such as Oct-1, Brn-1, and Brn-2 (not shown), we added a 12 molar excess of the oligonucleotide 28bp, which represents a good binding site for these factors (1), in the preincubation mixtures. A 12 molar excess of the oligonucleotide ThBFmut was also included to remove nonspecific DNA-binding proteins.

The second eluate from DNA-coated beads displayed a strong binding activity toward the ThBFsite probe, which was abolished specifically by the unlabeled ThBFsite competitor but not by the mutated ThBFmut oligonucleotide (not shown). Binding activities toward the 28bp probe or toward an unrelated AP-1 consensus binding site were no longer detectable (not shown). Quantification of ThBF binding by gel shift assays was not reliable, because the amount of probe-ThBF complex depended not only on the amount of active ThBF but also on parameters varying along the course of purification such as the concentration of poly(dA-dT) and of other DNA-interacting molecules in the assay. Therefore, accurate determination of the % recovery and purification fold was not possible. The last step of purification involved SDS-polyacrylamide gel electrophoresis where a crude brain nuclear extract and the purified ThBF fraction were run in parallel. The ThBFsite binding activity of the crude extract was recovered from the unstained gel slice corresponding to the apparent molecular mass of 48 to 50 kDa as described above. A silver nitrate-stained gel slice from the purified fraction lane, corresponding to the same electrophoretic mobility, was cut out and subjected to mass spectrometry analysis.

**Identification of ThBF as Pax-6**—Peptides resulting from the in-gel tryptic proteolysis of the 48–50 kDa gel slice were extracted and analyzed by nanoscale LC-MS/MS to generate associated information about mass and amino acid sequence for
Each detected peptide. The set of data was submitted for database searching using the NCBI database. Ten peptides were shown to match to the paired protein Pax-6 from *Rattus norvegicus*, which has a predicted molecular mass of 46,725 Da. To confirm that ThBF is Pax-6, we generated the oligonucleotide pax6cons respecting a consensus sequence selected for the binding of the Pax-6 paired domain (20). When pax6cons was radiolabeled and incubated with brain nuclear extracts, a retarded band with similar intensity and electrophoretic mobility as the ThBFsite complex was observed (Fig. 4C, lanes 1 and 2). This retarded band was competed efficiently by the unlabeled pax6cons oligonucleotide itself (Fig. 4C, lanes 3 and 4) and by the ThBFsite and ThBFdist but not ThBFmut oligonucleotides (Fig. 4C, lanes 3 and 5–7). Reciprocally, the pax6cons oligonucleotide was able to abolish the complex formed with the ThBFsite probe, and competition comparisons showed that the ThBFsite and the pax6cons motifs displayed similar affinities for the protein ThBF (Fig. 4C, lanes 8–14). Taken together, these results indicate that pax6cons and ThBFsite oligonucleotides were able to bind the same protein. Another Pax-6 consensus binding site (38) yielded similar results, with a lower affinity for ThBF than pax6cons (not shown). Indeed, sequence alignments revealed that ThBFsite and ThBFdist match with the consensus sequence for Pax-6 binding at 13 and 16 positions out of 20, respectively, and that some of the homologous residues correspond to the DNA/ThBF binding site of ThBF. A, gel shift experiments. The partially methylated end-labeled upper or lower strands of the (−864 to −802) PCR fragment were incubated with brain nuclear extracts (lanes 2–17). When indicated, unlabeled competitors (20 ng in lanes 3, 4, 6, 8–10, 14, and 17; 10 ng in lanes 13 and 16; or 4 ng in lanes 12 and 15) were included in the binding mixture. B, methylation interference analysis. The partially methylated end-labeled upper or lower strands of the (−864 to −802) PCR fragment were incubated with brain nuclear extracts. The bound and free probes (see A, lane 2) were then eluted and submitted to chemical cleavage. Residues at which methylation interferes with protein binding are indicated by circles. Solid circles represent a higher degree of interference than open circles. G+A, Maxam-Gilbert reaction. These data are summarized in Fig. 2B.
**Pax-6 Binding Sites in the Aldolase C Gene**

**A**

**Fig. 5. The Pax-6 binding sites of the aldolase C gene.** A, sequence homologies between the aldolase C gene Pax-6 binding sites and a Pax-6 consensus motif. The ThBFsite (upper strand), ThBFdist (lower strand), and pax6cons oligonucleotide sequences are aligned with the consensus reported by Epstein et al. (20) for the recognition of the Pax-6 paired domain. ThBFsite positions where methylation interferes with protein binding (see Fig. 2B and Fig. 3B) are noted by asterisks. Upper-case letters in the ThBFdist sequence represent bases, which are conserved between the rat and human aldolase C genes. Lowercase letters in the pax6cons sequence represent hanging bases flanking the Pax-6 consensus (W = A or T, S = G or C, K = G or T, M = A or C, Y = T or C). B, gel supershift experiments. The proteins and probes used and the presence or absence of the anti-Pax-6 antibody are indicated at the top of the gels.

Contact points determined by methylation interference experiments (Fig. 5A).

To establish definitively that the protein interacting with ThBFsite and ThBFdist was actually Pax-6, a specific polyclonal antibody was incubated with preformed DNA-protein complexes. The ThBF complexes observed with crude nuclear extracts from adult (Fig. 5B, lanes 1–4) and fetal (lanes 7 and 8) rat brain and with a purified fraction (lanes 5 and 6) were supershifted. This supershift was specific, because the binding of an unrelated transcription factor, USF, to the MLP site (21) interfered with protein binding both in the aldolase C sequence (lanes 9 and 10) and in the Pax protein binding sites of several candidate target genes. Epstein et al. (20) suggested that, because of the rather large DNA recognition sequences, no single nucleotide position is essential for the binding of paired domains (20). Indeed, the extended 27-bp footprint produced by Pax-6 at ThBFsite in the aldolase C gene is consistent with the 25- to 28-bp footprints produced by the Pax-6 paired domain (20) and by protein Pax-5 (45) in candidate target genes. Moreover, methylation of guanines at positions 5 (opposite strand) and 14 of the consensus interfered with protein binding both in the aldolase C sequence and in a representative sequence selected for binding of the Pax-6 paired domain (20) (Fig. 5A). A thymine at position 13, which is located in the selected sequences for binding of Pax-6 and in the Pax protein binding sites of several candidate target genes (20), is also present in the aldolase C ThBFsite and ThBFdist motifs (Fig. 5A).

**DISCUSSION**

We demonstrated previously that two upstream regions were necessary to activate the brain-specific aldolase C gene promoter, a distal fragment of 3.8 kb and a more proximal fragment of 0.6 kb (1). We show in this paper that the combination of these genomic sequences constitute a tissue-specific activator able to activate the ubiquitous tk promoter specifically in the brain of transgenic mice. We already identified in the 0.6-kb proximal fragment several POU motifs able to bind in vitro the proteins Brn-1 and Brn-2 expressed specifically in embryonic and adult nervous system. Progressive 5' deletions removing these binding sites additionally reduced transgene expression (1). Further analysis of the 0.6-kb fragment by in vitro DNase I foot printing allowed us to identify here a new protein binding site in the aldolase C gene regulatory regions. This new DNA element, extending from nucleotides −832 to −811, was located in close proximity downstream of the reported 28bp POU protein target site (1). It was shown to interact strongly with a brain-specific, thermostable protein referred to here as ThBF. This protein was also able to bind with the same affinity to a more distal target site, ThBFdist, extending from nucleotides −4424 to −4390. Moreover, ThBFdist is located in the subfragment A of the 3.8-kb distal activating region, which has been shown to be absolutely required for transcriptional activation (1). We undertook the purification of the ThBF protein from adult rat brain using a protocol including two cycles of specific adsorption onto streptavidin-magnetic particles coated with a biotinylated multimer of ThBFsite, an incubation step at 100 °C, and SDS-PAGE. The final polycrylamide fraction containing ThBF was analyzed by mass spectrometry, and ten peptides were found to match exactly with the transcription factor Pax-6. Several lines of evidence confirm the identification of ThBF as Pax-6: 1) gel shift competition experiments established that oligonucleotides representing consensus Pax-6 binding sites, ThBFsite and ThBFdist motifs, interacted with the same protein, 2) the 48–50-kDa apparent molecular mass of ThBF, as determined by renaturation from SDS-polyacylamide gel slices, is consistent with the molecular mass of 46,725 Da predicted for the R. norvegicus Pax-6 protein, 3) like Pax-6, ThBF is a brain-specific protein, and 4) finally, observation of a supershift with polyclonal anti-Pax-6 antibodies brought the direct proof that the protein interacting with both ThBFsite and ThBFdist was actually Pax-6.

Two pieces of information concerning Pax-6 were brought about our experiments: 1) Pax-6 is a thermostable protein, and 2) redox conditions modify the DNA binding activity of Pax-6. The 100% identical human and mouse Pax-6 proteins contain six cysteine residues, four of which are located in the paired domain (39). Inhibition of the DNA binding activity of Pax-6 by thiol-reactive metals and the alkylating agent N-ethylmaleimide suggests that at least one of them is critical, involved either directly in the contact with DNA, as it has been demonstrated for the transcription factors NFkB and AP-1 (40–42), or in the maintenance of the tridimensional structure of the protein. Redox sensitivity has already been reported for the proteins Pax-5 and Pax-8 and attributed to the reversible formation of an intramolecular disulfide bond involving conserved cysteines present in the paired domain (43). In B lymphocytes, the redox status of Pax-5 was shown to depend on the nuclear translocation of the enzyme Ref-1 (44).

It is noteworthy that, despite their similar affinity for Pax-6 demonstrated by competition experiments, the sequence homology between the oligonucleotides ThBFsite and Pax6cons is rather low (10 matches out of 20, with 2 of them being located outside the 16-nucleotide core; see Fig. 5A). In fact, no well defined optimal consensus sequence has been described for the binding of the different Pax proteins, and the initially reported core motif GTTCC is absent from several Pax protein binding sites. Epstein et al. (20) suggested that, because of the rather large DNA recognition sequences, no single nucleotide position is essential for the binding of paired domains (20). Indeed, the extended 27-bp footprint produced by Pax-6 at ThBFsite in the aldolase C gene is consistent with the 25- to 28-bp footprints produced by the Pax-6 paired domain (20) and by protein Pax-5 (45) in candidate target genes. Moreover, methylation of guanines at positions 5 (opposite strand) and 14 of the consensus interfered with protein binding both in the aldolase C sequence and in a representative sequence selected for binding of the Pax-6 paired domain (20) (Fig. 5A). A thymine at position 13, which is found in the selected sequences for binding of Pax-6 and in the Pax protein binding sites of several candidate target genes (20), is also present in the aldolase C ThBFsite and ThBFdist motifs (Fig. 5A).

Pax-6 belongs to a subclass of Pax proteins that possess a paired-type homeodomain in addition to their paired domain (6). Several AT-rich motifs able to bind homeodomains are present both overlapping and immediately upstream of the
ThBFsite motif in the aldolase C gene and have been shown already to interact with POU homeoproteins (see this paper and Ref. 1). This could allow either exclusive binding of Pax-6 or other homeodomain proteins or cooperative interactions between the paired domain of Pax-6 and the homeodomains of various other classes, as has been reported previously (46–48).

Identification of target genes for Pax-6 is an important approach to understanding the pivotal function of this transcription factor (for review, see Ref. 12). Binding sites for Pax-6 were found within its own promoter (49), in the promoters and enhancers of several crystalline genes (50), in the promoters of the genes encoding the transcription factors c-Maf and SIX3, also involved in lens development (51, 52), and in the promoter and enhancer of the glucagon, insulin, and somatostatin genes (14, 47). Other genes containing a Pax-6 binding site have been identified after selection of genomic sites (53), and potential lens target genes were selected recently using cDNA microarrays (54). For some of them, mutation of the Pax-6 binding site is conferred by a multiplicity of DNA motifs that are conserved perfectly between rat and human, makes it highly likely to be an additional target for Pax-6. The aldolase C protein is detectable at a low level in all embryonic tissues (35) whereas a strong, brain-specific expression is observed at the adult stage. Although few data are available concerning the Pax-6 expression in adult (56), it is interesting to note that we purified Pax-6 from adult brain.

In conclusion, the brain-specific transcriptional activation of the aldolase C gene is conferred by a multiplicity of DNA motifs scattered along distal and proximal sequences and binding brain-specific transcription factors, in particular POU factors (35, 47), which are conserved perfectly between rat and human, makes it highly likely to be an additional target for Pax-6. The aldolase C protein is detectable at a low level in all embryonic tissues whereas a strong, brain-specific expression is observed at the adult stage. Although few data are available concerning the Pax-6 expression in adult, it is interesting to note that we purified Pax-6 from adult brain.

In conclusion, the brain-specific transcriptional activation of the aldolase C gene is conferred by a multiplicity of DNA motifs scattered along distal and proximal sequences and binding brain-specific transcription factors, in particular POU factors and Pax-6. Previous functional data (1) showed that cooperation between these elements is required to assure a high level expression. The protein Pax-6, in addition to its functions during development, may also participate directly in controlling genes expressed specifically in adult brain.

REFERENCES

1. Skala, H., Porteu, A., Thomas, M., Stajniert, M. F., Okazawa, H., Kahn, A., and Phan-Dinh-Tuy, F. (1998) J. Biol. Chem. 273, 31806–31814
2. Ryan, A., and Rosenfeld, M. (1997) Genes Dev. 11, 1207–1225
3. Ang, S.-L., and Rossant, J. (2002) Cell 111, 714–722
4. Wagner, C., and Gruss, P. (1991) Development 113, 1435–1449
5. Hogg, B., Horshagh, G., Cohen, J., Hetherington, C., Fisher, G., and Lyon, M. (1986) J. Embryol. Exp. Morphol. 97, 95–110
6. Ashery-Padan, R., and Gruss, P. (2001)Curr. Opin. Cell Biol. 13, 706–714
7. Stoykova, A., Treichel, D., Hallonet, M., and Gruss, P. (2000) J. Neurosci. 20, 8042–8050
8. O’Leary, D. D., and Nakagawa, Y. (2002)Curr. Opin. Neurobiol. 12, 14–25
9. Heyningen, V. V., and Williamson, K. (2002)Hum. Mol. Genet. 11, 1161–1167
10. Larsson, L., St-Onge, L., Hougaard, D., Sosa-Pineda, B., and Gruss, P. (1998)Mech. Dev. 79, 153–159
11. Sander, M., Neubauer, A., Kalamaras, J., He, H., Martin, G., and German, M. (1997)Genes Dev. 11, 1662–1673
12. Heins, N., Malatseta, P., Cecconi, P., Nakayasu, T., Tuck, K., Hack, M., Chapoult, P., Barde, Y., and Got, M. (2002) Nat. Neurosci. 5, 308–315
13. McKechnie, S., and Kingsbury, R. (1982)Science 217, 316–324
14. Hogen, B., Bedington, R., Costantini, F., and Lacy, E. (1989)Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
15. Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M. A., Jaenisch, R., and Berns, A. (1991) Nucleic Acids Res. 19, 4293
16. German, C., Moffit, L., and Howard, B. (1982) Mol. Cell. Biol. 2, 1044–1051
17. Epstein, J., Cai, J., Glaser, T., Jepeal, L., and Maas, R. (1994) J. Biol. Chem. 269, 8355–8361
18. Travers, M., and Roeder, R. G. (1985) Cell 43, 165–175
19. Thomas, M., Makeh, I., Briand, P., Kahn, A., and Skala, H. (1993) Eur. J. Biochem. 218, 143–151
20. Toledano, M., Ghosh, D., Trinh, F., and Leonard, W. (1993)Mol. Cell. Biol. 714, 4194–4200
21. Wilson, W., Bellen, H., and Gehring, W. (1990) Annu. Rev. Cell Biol. 6, 679–714
22. Berg, J., and Yigong, S. (1996) Science 271, 1081–1085
23. Meech, R., Kallunki, P., Edelman, G., and Jones, F. (1999) Proc. Natl Acad. Sci. USA 96, 2420–2425
24. Glaser, T., Walton, D., and Maas, R. (1992) Nat. Genet. 2, 232–238
25. Abate, C., Patel, L., Rauscher, F., III, and Curran, T. (1996)Science 271, 1157–1161
26. Toledano, M., Ghosh, D., Trinh, F., and Leonard, W. (1993)Mol. Cell. Biol. 13, 159–166
27. Tell, G., Scalani, A., Pellizzari, L., Formisano, S., Puccio, C., and Damante, G. (1996) J. Biol. Chem. 271, 25062–25072
28. Tell, G., Zecch, A., Pellizzari, L., Spessotto, P., Colombati, A., Kelley, M., Damante, G., and Puccio, C. (2000) Nucleic Acids Res. 28, 1099–1105
29. Kosmik, Z., Wang, S., Do, M., Edelman, G., and Jones, F. (2001) J. Biol. Chem. 276, 4124–4132
30. Jun, S., and Deplanck, C. (1996) Development 2639–2650
31. Fritz-Laser, B., Estreich, A., Klages, N., Stieve, S., and Philipp, J. (1999) J. Biol. Chem. 274, 4131–4132
32. Mikkola, I., Bresser, M., and Johansen, T. (2001) J. Biol. Chem. 276, 4109–4118
33. Piazza, S., Donzis, B., and Saule, S. (1999)Cell. Growth Differ. 4, 1041–1050
34. Cevid, A., and Piatigorsky, J. (1996)Bioessays 18, 621–630
35. Sakai, M., Serria, M., Boda, H., Yoshida, K., Imaki, J., and Nishi, S. (2001)Nucleic Acids Res. 29, 1228–1237
36. Langel, J., and Gray, J. (2001)Biochem. Biophys. Res. Commun. 287, 372–376
37. Zhou, Y., Zheng, J., Gu, X., Li, W., and Saunders, G. (2000)Gene 245, 319–328
38. Chauhan, B., Reel, Z., Zhang, W., Duncan, M., Kilimann, M., and Cvekl, A. (2002) J. Biol. Chem. 277, 11539–11544
39. Holst, B., Wang, Y., Jones, F., and Edelman, G. (1997) Proc. Natl Acad. Sci. 94, 1445–1450
40. Stoykova, A., and Gruss, P. (1994)J. Neurosci. 14, 1385–1412
41. Maxam, A. M., and Gilbert, W. (1980)Methods Enzymol. 65, 499–560
Novel Target Sequences for Pax-6 in the Brain-specific Activating Regions of the Rat Aldolase C Gene

Henriette Skala-Rubinson, Joëlle Vinh, Valérie Labas, Axel Kahn and Françoise Phan
Dinh Tuy

J. Biol. Chem. 2002, 277:47190-47196.
doi: 10.1074/jbc.M209349200 originally published online October 4, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M209349200

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 56 references, 23 of which can be accessed free at
http://www.jbc.org/content/277/49/47190.full.html#ref-list-1