E-box motifs within the human vasopressin gene promoter contribute to a major enhancer in small-cell lung cancer

Judy M. COULSON^{*1}, Carolyn E. FISKESTRAND[†], Penella J. WOLL^{*} and John P. QUINN[†]

^{*}CRC Academic Unit of Clinical Oncology, University of Nottingham, City Hospital, Hucknall Rd, Nottingham NG5 1PB, U.K., and [†]Department of Veterinary Pathology, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, Scotland, U.K.

[Arginine]vasopressin (AVP) is a neuropeptide physiologically synthesized in the hypothalamus but pathologically expressed by small-cell lung cancer (SCLC). A minimal 65 bp AVP promoter can restrict basal activity to SCLC *in vitro*, but a 199 bp fragment directs 5-fold higher expression in SCLC [Coulson, Stanley and Woll (1999) *Br. J. Cancer* **80**, 1935–1944]. Several predicted E-box motifs occur within the 199 bp fragment, and we now describe an enhancer which contributes to AVP promoter tumour-specificity in some cell lines. The deletion of two adjacent E-boxes (–157 to –131) resulted in an approx. 70% loss of reporter gene expression in a SCLC line (Lu-165) with high endogenous AVP production. Using a series of AVP promoter deletion constructs and site-directed mutagenesis, we show that both these E-box sites were required for enhancer function, whereas mutation of an adjacent AP-1 site had no effect on the promoter activity. Electrophoretic-mobility-shift analysis indicated that, although both the predicted E-box motifs bound specific complexes, only one appeared to function as a strong E-

box which binds basic helix–loop–helix (bHLH) factors. This motif formed a complex in lung tumour-cell extracts, which was particularly strongly bound in Lu-165, and was competed for by a characterized E-box motif from the preprotachykinin A promoter. Antibody supershifts indicate that this complex is a heterodimer of upstream stimulatory factor (USF)-1 and USF-2. Non-bHLH complexes weakly bound the second potential E-box motif in a SCLC-specific manner. These complexes were not recognized by the bHLH antibodies and remain unidentified; however, they were detected in seven of eight SCLC cell lines and not in four control lines. We postulate that there is a co-operative and complex interaction between an E-box and an adjacent site constituting a SCLC-specific enhancer within the AVP proximal promoter.

Key words: lung tumour, neuropeptide, transcription, tumour-specific, upstream stimulatory factor.

INTRODUCTION

Physiological [arginine]vasopressin (AVP) expression is largely restricted to the hypothalamus and is tightly regulated by plasma osmolality. Some AVP expression is seen in other tissues [1], but it is often secreted in large amounts by small-cell lung cancer (SCLC). In fact SCLC is biochemically characterized by the production of a variety of neuropeptides, but AVP is the most common, being detected in up to 67% of clinical samples [2]. Neuropeptides are of interest in the context of tumour biology because they can cause clinical syndromes such as dilutional hyponatraemia (AVP, atrial natriuretic protein) and Cushing's syndrome [adrenocorticotrophin (ACTH)]. In addition, some neuropeptides, such as AVP and gastrin-releasing peptide, can stimulate proliferation via autocrine loops [3,4]. Elucidation of the mechanisms regulating AVP expression in SCLC may therefore provide information on establishment of the autocrine neuroendocrine phenotype and aid in the design of molecular therapies for SCLC. In addition, transcriptional regulation in these tumour-cell lines may provide an insight into physiological regulation.

Osmotic responsive elements have been identified in the AVP promoter [5,6], but it is not yet clear whether this regulation is at

the level of AVP synthesis or secretion [7]. There is, however, persuasive clinical evidence that AVP expression in SCLC is not subject to the normal osmotic regulation, as AVP production continues despite dilutional hyponatraemia. Although several alternative regulatory motifs have been identified within the proximal promoter region, including glucocorticoid and cAMP-responsive elements (CREs) [8,9], none of these has been shown to confer either normal tissue- or tumour-specificity. It therefore remains to be established which alternative elements are important in SCLC-specific expression. We have recently demonstrated that a 199 bp fragment of the human AVP promoter retains high SCLC-specific activity in a panel of cell lines [10,11]; this region contains several predicted binding motifs, including E-boxes, AP-1 and AP-2 sites [12].

Several neuronal transcription factors, which are involved in determination of cell lineage, have been identified as candidate SCLC-specific factors, including thyroid nuclear factor (TTF) [13], Pit/Oct/Unc (POU) domain factors [14–16] and the basic helix–loop–helix (bHLH) factors; the latter bind the E-box motif CANNTG. E-boxes are known to be important in the expression of several neuropeptides, including pro-opiomelanocortin (POMC) the polypeptide precursor of ACTH [17], calcitonin-gene-related peptide (CGRP) [18] and preprotachykinin A (PPT)

Abbreviations used: SCLC, small-cell lung cancer; AVP, [arginine]vasopressin; ACTH, adrenocorticotrophin; CRE, cAMP-responsive element; TTF-1, thyroid nuclear factor 1; POU, Pit/Oct/Unc; bHLH, basic helix–loop–helix; POMC, pro-opiomelanocortin; CGRP, calcitonin-gene-related peptide; PPT, preprotachykinin A; USF, upstream stimulatory factor; hASH, human achaete-scute homologue; HES, hairy-enhancer-of-split; Id, inhibitor of DNA binding; NSCLC, non-small-cell lung cancer; NRSE, neuron restrictive silencer element; CAT, chloramphenicol acetyltransferase; BCS, bovine-calf serum; HBE, human bronchial epithelium; EMSA, electrophoretic-mobility-shift assay; RT-, reverse transcription; SV40, simian virus 40.

[19,20]. E-boxes have also been characterized in the rat AVP promoter [12]. Members of the bHLH family include upstream stimulatory factor (USF)-1 and USF-2, the proto-oncogene *c-myc*, human achaete-scute homologue (hASH)-1, which has been associated with neuroendocrine tumours [21–24], hairy-enhancer-of-split (HES)-1 repressor [25] and the inhibitors of DNA binding (Ids) [26], which may play key roles in tumorigenesis [27].

The transcription factor AP-1 appears to be involved in the hypothalamic response to osmotic stress and may also contribute to SCLC-specific expression. Dehydration, a stimulus for increased AVP expression, induces *c-fos* expression in rats [28] and increased AP-1 DNA binding activity has also been shown in rat hypothalamus following dehydration, although no direct link to AVP expression levels was established [29]. However, it has been reported that AP-1 binding is substantially reduced in SCLC relative to non-SCLC (NSCLC) [30]. Interestingly, it has also been suggested that transcriptional regulation through E-boxes could be modulated by adjacent AP-1 sites in some promoters [31,32], and a similar arrangement of motifs occurs within the AVP proximal promoter.

We propose that the combination of repressor molecules and cell-type-specific enhancers are required to achieve tissue-specific expression of AVP, as has been hypothesized for several neuronal genes, including a variety of neuropeptides [33]. We have previously characterized such repressors in the PPT and AVP promoters [32,34–36]. Our previous data indicate that the regulatory elements required for high-level AVP expression, to which SCLC-specific activators may bind, are located between –157 and –23. We describe here analysis of potential enhancer elements involved in pathological activation of the AVP promoter. We have already identified potential transcription-factor binding sites within the proximal AVP promoter by sequence comparison, including three E-boxes, an AP-1 [11] and a motif with similarity to the neuron-restrictive silencer element (NRSE) [35]. Here we report that at least one of these putative E-boxes is involved in binding bHLH factors and that the adjacent predicted AP-1 motif does not function in lung-tumour-cell lines. Additional SCLC-specific complexes were observed binding adjacently to the strong E-box motif, which are postulated to modulate AVP promoter activity within this fragment. It appears that an interaction of factors which bind E-boxes and adjacent sites co-operatively contribute to the differential regulation observed between SCLC and normal lung or non-neuroendocrine NSCLC tumours.

MATERIALS AND METHODS

Cloning of reporter constructs

The AVP promoter/chloramphenicol acetyltransferase (CAT) reporter constructs p65, p128, p173 and p199 are based on the promoterless pBLCAT6 plasmid (pnone) as previously described [11]. p185 was generated by PCR amplification of the region –43 to +42 and ligation into PCR2.1 (Invitrogen) for subcloning into pBLCAT6, as previously described [11]. The 5' end of each construct is illustrated in relation to predicted transcription-factor-binding sites in Figure 1(a). Site-directed mutagenesis of the E-box and AP-1 sites was carried out using the forward PCR primers Amut, Bmut, (AB)mut and AP1mut, with the reverse primer VPP2. The forward primers and the specific mutations generated are illustrated in Figure 1(b). The AP-1 mutation was based on a characterized mutation known to block function [30]. These products were cloned into pBLCAT6 to produce the reporter gene constructs. The mutated sequences were checked using the signalscan program [37] available through the Human Genome Mapping Project Resource Centre (Cambridge, U.K.),

Figure 1 AVP promoter region of interest and site-directed mutagenesis of sites within the human sequence

(a) Alignment of the AVP promoter region, containing the E-box motifs of interest, between the human (H), murine (M), rat (R) and bovine (B) sequences. Mismatches (to H) are shown by lower-case letters. Transcription-factor-binding sites are boxed or underlined, the 5' end of the reporter constructs is indicated and the TATA box is at –23. (b) The forward PCR primers used to generate the mutated clones are shown aligned to the human wild-type sequence in the p199 construct (top), the position of the 5' end is indicated and the 3' end of all the cloned PCR products is at +42. The predicted transcription-factor-binding motifs are shown in **bold** *italics* and those bases mutated in each construct are underlined.

for the generation of additional known transcription-factor-binding sites. The sequence for all the inserts in the final vector was confirmed by automated sequencing.

Cell culture

The Lu-165 SCLC line was a gift from Dr. T. Terasaki (National Cancer Center, Tokyo, Japan); the NSCLC line COR-L23 and SCLC lines COR-L24 and COR-L88 were gifts from Dr. P. Twentyman (established at the MRC Clinical Oncology Research Unit, Cambridge, U.K.). NCI-H711 (SCLC) was a gift from Dr. B. Johnson (National Cancer Institute, Bethesda, MD, U.S.A.) and U2020 (SCLC) was from Dr. P. Rabbitts (MRC Laboratory of Molecular Biology, Cambridge, U.K.). Other SCLC (GLC-19, NCI-H69, NCI-H345), NSCLC (NCI-H460) and HeLa cell lines originated from the A.T.C.C. These cell lines were routinely maintained in RPMI/10% bovine-calf serum (BCS) at 5% CO₂ and 37 °C. Simian-virus-40 (SV40)-transformed human bronchial epithelial (HBE) cells were a gift from Dr. W. Franklin (University of Colorado Cancer Center, Denver, CO, U.S.A.). These were maintained [38] in fully supplemented Bronchial Epithelial Growth Medium (Clonetics; U.K. distributor TCS, Botolph Claydon, Buckingham, Bucks., U.K.) and used for experiments

at the third passage. All cells were harvested for experiments during exponential growth.

Transfection of mammalian cells

Transfections were carried out by electroporation as previously described [11]. Briefly, 7.5×10^6 cells in a volume of 400 μ l of complete medium were electroporated at 260 V and 1050 μ F, with 20 μ g of each test plasmid and 20 μ g of pSV β Gal to control for transfection efficiency. Cells were cultured in 20 ml of complete medium and harvested after an expression period of 72 h. Lysates were prepared in 500 μ l of CAT lysis buffer and assayed for protein content [ESL (exact, sensitive, low interference); Roche Diagnostics Ltd., Lewes, East Sussex, U.K.]. CAT and β -galactosidase reporter gene assays used the respective ELISA kits (Roche Diagnostics). Values were converted into pg of enzyme/ μ g of protein, and β -galactosidase levels were used to standardize the data.

Electrophoretic-mobility-shift analysis (EMSA)

Whole-cell extracts were prepared from cell lines following three freeze-thaw cycles [39]. The samples were suspended in an equal volume of buffer (20 mM Hepes/20 % glycerol/0.42 M NaCl/0.2 mM EDTA/0.25 mM dithiothreitol/0.5 mM PMSF/1.5 mM MgCl_2) and the overall salt concentration was adjusted to 0.42 M. Cells were Dounce-homogenized, debris was pelleted and the protein concentration of extracts determined (ESL, Roche Diagnostics Ltd). The EMSA probes used are shown in Figure 4 (below). Single-stranded oligonucleotides were annealed by denaturing at 94 °C and cooling slowly to room temperature. Double-stranded oligonucleotides (100 ng) were end-labelled with [γ - 32 P]ATP and polynucleotide kinase; unincorporated label was removed using Chromaspin-10 columns (Clontech) according to the manufacturer's instructions. Reactions were carried out at 100 mM NaCl in the presence of 500 ng/ml poly(dI·dC), and cell extract was incubated with 1 ng of probe at room temperature for 10 min. Excess of cold competitor double-stranded oligonucleotides were pre-incubated with the extract for 5 min, as indicated for individual experiments. Competitors included a non-specific (NS) oligonucleotide (sense strand: 5 TCG AAT CCC TTT AAA TTT GCG AGC 3), a characterized E-box motif (E) and a double AP-1 motif (2AP) both derived from the PPT promoter sequence [19,40], and an AP-1 consensus sequence (Santa Cruz, Santa Cruz, CA, U.S.A.; see Figure 4 below). For supershifts of transcription-factor-bound probes, 1 μ l of antibody was preincubated with the extract for 15 min. Antibodies used were USF-1, USF-2 (N-terminal) and c-Myc (all TransCruz supershift antibodies, supplied by Insight Biotechnology Ltd., Wembley, Middx., U.K.). Samples were analysed on 4 % non-denaturing polyacrylamide (acrylamide/bisacrylamide ratio 29:1)/0.5 $\times$ Tris borate electrophoresis buffer gels, dried and detected by autoradiography with β max Hyperfilm (Amersham) or by PhosphorImaging (Molecular Dynamics).

RESULTS

The predicted E-boxes A and B act as an enhancer in SCLC with high AVP expression

We have previously screened a panel of SCLC and NSCLC cell lines for expression of AVP by immunocytochemistry, reverse-transcription PCR (RT-PCR) and Northern blotting [11]. A range of AVP expression was detected in the SCLC lines, which was highest in the Lu-165 cell line [41]. Control non-neuroendocrine cell lines, including the NSCLC line NCI-H460 and

SV40-transformed HBE cells, had no detectable AVP expression [11]. We have proposed that the basic level of expression of AVP in SCLC is dictated by the minimal promoter region (p65) containing an NRSE-like sequence [35]. However, inclusion of the additional 5' promoter sequence from -157 to -23 (within the construct p199) significantly increases this expression in all the SCLC cell lines tested (J. M. Coulson and P. J. Woll, unpublished work). This region is highly homologous between species (see Figure 1a). The panel of cell lines described above provides a model of high, low and undetectable AVP expression in which to elucidate the enhancer elements involved in tumour-specificity of the proximal AVP promoter.

Two E-box motifs are located at the 5' end of the 199 bp fragment, adjacent to an AP-1 site, and synergy between AP-1 and E-box sites has been suggested to contribute to tissue-specific regulation [31,32]. Using a series of 5' deletion constructs (Figure 2a), this region, containing the two potential E-box motifs annotated A and B, was tested for enhancer activity in SCLC cell lines which express different levels of endogenous AVP. Activity is shown relative to that of the p65 minimal promoter in each cell line (Figures 2b and 2c). An enhancer function was clearly attributable to the region containing the predicted E-boxes A and B in the Lu-165 cells (Figure 2b). Interestingly, while the loss of A and B decreased activity, the further deletion of the adjacent AP-1 partially restored activity, implying the loss of a repressor within this region. We have previously shown that p199 has approx. 30-fold higher activity in Lu-165 than NCI-H345 cells [11] and the region containing E-boxes A and B appeared to be less important as an enhancer in NCI-H345 cells (Figure 2c). Although the absolute activity of the minimal promoter (p65) is also higher in Lu-165 than other cell lines such as NCI-H345, the high AVP expression in Lu-165 is clearly partially reliant on the enhancer activity associated with the predicted A and B E-box sequences.

Adjacent E-boxes demonstrate co-operativity in functional assays

Only a small percentage of SCLC lines retain high AVP expression *in vitro* [42], although at least 66 % of tumours express AVP [2]. However, the Lu-165 cells are a good model for *in vivo* SCLC tumours, as this line is known to retain many characteristics of the original tumour, including AVP expression, after multiple serial passages in culture [41]. We have already shown that reporter gene expression dependent on the AVP promoter correlates to the endogenous AVP expression level in cultured lung tumour-cell lines [11]. Therefore the enhancer function of the AB region was investigated further in the Lu-165 model. To localize the E-box, or other element conveying high level promoter activation, we mutated specific bases within the motifs of interest. The primers used for site-directed mutagenesis are shown in Figure 1(b) and the constructs generated are illustrated in Figure 3(a). The mutated sequences did not bind specific protein complexes, as described below.

The respective reporter-gene activity for these constructs in the Lu-165 cell line is shown in Figure 3(b). The loss of one or both of the E-box elements, either in the 5' deletion constructs or the site-mutated constructs, significantly reduced transcription. Although it is known that bases adjacent to the hexanucleotide core motif of E-boxes will also affect binding to these sites, it is thought that this effect is restricted to the three immediately 5' bases [43,44]. Therefore the point mutations we have introduced should be sufficiently remote from adjacent sites that the resultant loss in binding is restricted to the target E-box motif. The activity for all constructs was reduced to approx. 30 % of that seen for

Figure 2 Enhancer activity of the AB E-box region in SCLC with high AVP expression

(a) Representations showing the putative transcription-factor-binding sites in each of the 5' deletion constructs of the AVP promoter: p199 (−157 to +42), p173 (−131 to +42), p128 (−86 to +42), p65 (−23 to +42). (b) and (c) Transient transfection of SCLC cell lines with these reporter constructs; data are represented as the percentage of p65 minimal promoter activity in that cell line: (b) the Lu-165 cell line with high endogenous AVP expression ($n = 3$, ± 1 S.D.) and (c) the NCI-H345 cell line with low AVP expression ($n = 5$, ± 1 S.D.).

Figure 3 Site-directed mutagenesis of the AVP proximal promoter E-boxes and reporter-gene expression in Lu-165

(a) Illustration of the CAT reporter constructs with mutated E-boxes. The intact E-boxes and other predicted transcription-factor-binding sites are shown for each plasmid. (b) Histogram showing the relative CAT reporter activity from each of these constructs following transfection into Lu-165 cells; data are expressed as a percentage of the p199 reporter gene expression ($n = 3$, ± 1 S.D.).

AVP promoter:

A	⁵ AGGCC CACGT GTGTCC TCCGG GTGCAC ACAGG ³	-152/-137	16mer
A*	⁵ AGGCC TACGAG GTGTCC TCCGG ATGCTC ACAGG ³	-152/-137	16mer
B	⁵ GTCCC CAGAT GCCTGA CAGGG TCTAC GGACT ³	-140/-125	16mer
AB	⁵ AGGCC CACGT GTGTCC CAGAT GCCTGA TCCGG GTGCAC ACAGGG TCTAC GGACT ³	-152/-125	28mer
ABAP1	⁵ AGGCC CACGT GTGTCC CAGAT GCCTGA ATC ACTGCTGAC TCCGG GTGCAC ACAGGG TCTAC GGACT TA GTGACGACTG ³	-152/-113	40mer
AVPAP1	⁵ TGCC TGAAT CACTGC ACGG ACTTAGT GACG ³	-131/-117	15mer
AVPAP1*	⁵ TGCC CGAAG CACTGC ACGG CTTCT GACG ³	-131/-117	15mer

PPT promoter:

E	⁵ tcgaGGAGAGTGT CACGT GGCTCTCC CCTCTCACAG TGCAC CGAGAGagct ³
2AP	⁵ tcgaAGC ATCAGTCACT TCGCTCAGTTT GATGAGTA TC TCG TACTCAGTGA AGCGAGTCAA ACTACTCATT Agagct ³

Consensus:

AP1	⁵ CGCTT GATGACTCAG CCGGAA GCCAA CTACTGAGTC GGCCTT ³
------------	--

Figure 4 Probes for EMSA

Double-stranded probes for EMSA with the transcription-factor-binding sites in each shown in **bold**. The probes represent regions of the AVP promoter: A, B, AB, ABAP1, AVPAP1. In the mutated probes, A* and AVPAP1*, the base mutations are underlined. Control probes were derived from the PPT promoter: E (E-box at -60) and 2AP (double AP-1 motif) or a consensus-binding-site sequence: AP1.

p199, that is, equivalent to the activity of p173 where the whole AB region is deleted. These data imply that the predicted E-box sites interact co-operatively, so that both A and B were required for the enhancer function.

Binding of an adjacent AP-1 site is not required for AVP expression in SCLC

Site-directed mutagenesis was also used to investigate whether AP-1 binding functionally affected the E-box enhancer. The previously described AP-1 mutation [30] (see Figure 1b) was introduced into p199. In contrast with the E-boxes, mutation of the predicted AP-1 site had no impact on promoter activity in SCLC; additionally, loss of the AP-1 site did not permit reporter-gene expression in NSCLC (results not shown). The lack of a role for the AP-1 in these cells was supported by EMSA; the probes used are shown in Figure 4. The consensus AP-1 probe bound a specific protein in all the lung tumour-cell lines. However, probes corresponding to the AVP promoter AP-1 failed to show strong binding in any of the extracts tested, either in isolation (AVPAP1) or as part of the larger composite probe (ABAP1) (results not shown). Competition assays demonstrated that specific complexes on the ABAP1 probe were bound entirely via the E-boxes

and, even when this was competed out, no AP-1 shift was apparent. It therefore appeared that the AVP AP-1 motif does not have high affinity for the AP-1-binding factor present in the lung tumour lines.

E-box A binds strongly to complexes *in vitro*, especially in Lu-165

EMSA was used to investigate whether the predicted E-box motifs showed any differential transcription-factor binding in SCLC and control-cell extracts. The probes used are shown in Figure 4. A single specific band was seen with the E-box A probe in Lu-165 cell extract (Figure 5a). This complex was competed away from the E-box A probe by a 50-fold excess of homologous oligonucleotide, and by both the AB and ABAP1 composite probes, but not by a non-specific oligonucleotide (Figure 5a). Non-specific binding, as determined by competition with an irrelevant oligonucleotide, is indicated on the Figures and was attributed to the Ku antigen [45], which binds DNA termini in a sequence-independent manner. The NCI-H460 cell extract (NSCLC, no AVP) bound a specific complex of the same mobility as that in Lu-165, albeit with a lower affinity (Figure 5a).

A strong E-box motif has previously been characterized at -60 in the PPT promoter (E) [19,20]; this also competed for the specific complex bound by A in either extract (Figure 5a). However, a mutated E-box probe (A*), corresponding to the sequence in the pAmut reporter constructs, failed to compete for binding to the wild-type E-box A. In addition, the labelled mutant A* probe failed to bind specific protein complexes in Lu-165 (Figure 5a) and other cell lines (results not shown), thus confirming both that the functional mutation abolished E-box binding and that no new transcription-factor-binding sites had been created which bound factors in the lung tumour cells. Whilst EMSA indicated that A functioned as a strong E-box, the predicted E-box B in fact only bound factors of the expected mobility very weakly in Lu-165 (Figure 5a). This binding was only detectable by a long exposure using the more sensitive technique of PhosphorImaging (Figure 5b) and is therefore significantly weaker than that seen for the A E-box. In addition, E-box B did not compete efficiently for specific complexes bound to the A probe at a 50-fold excess, and even at 500-fold excess of B, the binding to A was only partially competed (Figure 5c). The two E-boxes therefore bind factors with different affinities and specificity.

The PPT -60 E-box (E) is of the identical six-base motif as E-box A in the AVP promoter; however, although these probes will cross-compete at a suitable excess, the major specific complex bound by each in the Lu-165 extract differs (Figure 5d). The complex is of a different mobility, indicating that alternative specific factors may be preferentially bound, perhaps dependent upon the sequence surrounding the 6 bp consensus sequence. In contrast, similar specific complexes formed on the A E-box probe in all the cell lines tested. However, in the SCLC with low endogenous AVP expression and AVP-negative control cells, this complex appeared to be weaker than in the Lu-165 cell extract (Figure 6) as already described for NCI-H460 (Figure 5a). Although EMSA is not quantitative, the same finding was noted in replicate experiments using multiple extracts from these cell lines, and the protein content of all extracts varied by less than 20%. Therefore protein binding via the A E-box was not SCLC-specific, but may be differential according to AVP status or could be suggested to represent alternative factors in the cell lines.

The complex binding E-box A contains USF-1 and USF-2

Antibodies to USF-1, USF-2 and c-Myc were used in an attempt at identification of the bHLH factors involved in the complex

Figure 5 Protein binding to E-box motifs in SCLC and NSCLC

In each panel, specific bands are shown with arrows (→), non-specific bands with lines (—) and the free probe with a bracket (⌊). Competitors are at a 50-fold excess unless otherwise stated; NS, non-specific double-stranded oligonucleotide. **(a)** The mobility shifts are compared for the A, B and A* E-box probes in the Lu-165 cell extract and for the A probe in NCI-H460 (all 18 h exposure). **(b)** Weak bHLH binding to the predicted B E-box in Lu-165 detected by Phosphorimaging (Molecular Dynamics). **(c)** E-box B only weakly competes for binding to E-box A (24 h exposure). Competitors A and E are shown at a 50-fold excess, competitor B is shown at both 50-fold and 500-fold, and NS is shown at a 500-fold excess. **(d)** Comparison of binding to the AVP A E-box and the PPT—60 (E) E-box CACGTG probes in Lu-165 (24 h exposure).

binding to E-box A in different lung tumour-cell lines (Figure 6). Again stronger binding is evident in Lu-165 than in other cell lines expressing lower levels of vasopressin, however the antibody supershifts indicate that in each case the complex contains USF. No shift was seen with the c-Myc antibody. The USF-1 antibody

produces a weak supershift, but abolishes the original shift in all cell lines, except perhaps Lu-165. The USF-2 antibody generates a strong supershift of the majority of the complexes bound to A. This would indicate that in each case the predominant complex formed at this E-box is a heterodimer of USF-1 and USF-2.

Figure 6 Complexes binding E-box A contain USF

The complex binding to the A E-box probe was evaluated by antibody supershifts in four cell lines. Competition is shown with a 100-fold excess of double-stranded oligonucleotide or with 1 μ l of antibody for bHLH factors.

Predicted E-box B binds an additional factor in SCLC

The AB probe, containing both predicted E-boxes A and B, again bound strongly to a specific complex in Lu-165 and to a lesser extent in NCI-H460 (Figure 7a). Antibody studies also confirmed this to be USF-1/USF-2 (results not shown), which could be competed with the homologous AB probe and with A alone. Again, there was no competition with a 50-fold excess of B, implying that bHLH binding is predominantly through the A E-box, even in the context of a complex probe (Figure 7a). However, where E-box binding was competed out with excess of an unlabelled E-box motif (E), the AB probe was seen to bind an additional factor of slower mobility in the Lu-165 cell extract (Figure 7b). This extra band was not competed for by a further excess of the A probe; therefore this specific complex is unlikely to be a bHLH transcription factor. In contrast, excess of either the B or AB probes did compete, implying that this binding occurs through the B motif. However, strong binding of this factor to the B probe alone was not observed, suggesting that although the complex forms through B, specificity may be increased by interactions with sequences in the adjacent A motif. By competition studies, this appeared not to be an E-box binding protein. Most interestingly, this specific complex did not appear to be bound in AVP-negative NSCLC cell extract (NCI-H460) (Figure 7b).

To explore whether this complex was common to other SCLC extracts, a large panel of lung tumour-cell lines was screened (Figure 7c). An additional complex binding via B was seen in seven of eight SCLCs tested. The A E-box probe was retarded to the same position as in Lu-165 in four out of the seven additional SCLC extracts tested. An alternative band of faster mobility was evident in the NCI-H345 and NCI-H711 extracts, and two complexes were evident in some of the SCLC cell lines, for example NCI-H69. In contrast, no B-specific complexes were seen in any of the NSCLC, normal lung epithelium or HeLa

control extracts. The COR-L88 (SCLC) cell line was negative for a B-specific complex; however, this is known to be a morphological variant [46] which more closely resembles NSCLC. Antibodies to the common 'A type' bHLH factors failed to produce a supershift in the B-specific complex in either the Lu-165 or NCI-H345 cell lines (results not shown). These complexes are at present unidentified, but may play a major role in interacting with a bHLH factor binding E-box A, thereby constituting an enhancer in the AVP proximal promoter of lung tumour lines expressing AVP.

DISCUSSION

The proximal human AVP promoter has several putative E-box sites and a NRSE-like motif adjacent to the transcriptional start site, the latter in part restricting expression to SCLC *in vitro* [35]. The two previously predicted E-boxes (–157 to –131) described here are also important in regulation of the AVP promoter, having enhancer activity in those cell lines which show high AVP expression. The corresponding E-boxes and other motifs within the rat AVP promoter have previously been demonstrated to bind specific complexes from HeLa-cell extracts [12]. In the present study we have concentrated on those motifs which lie within the p199 promoter fragment of the human promoter, as we have already shown that this is sufficient to confer high level activity in SCLC [11]. We report here the investigation of an enhancer element, corresponding to the E-boxes predicted at –157 to –131, which is involved in tumour-specific activation of the human AVP promoter in SCLC.

This sequence containing the potential E-boxes A and B acted as an enhancer of the AVP proximal promoter in the Lu-165 cell line. However, the same sequence had little enhancer activity in cell lines with low endogenous production of AVP. The enhancer function of these E-boxes in reporter-gene constructs, therefore, reflected the level of expression driven by the endogenous AVP promoter in these cell lines. In addition, we demonstrated a functional co-operativity between the A and B motifs in reporter-gene assays using mutated and 5' deleted constructs, where loss of either or both motifs reduced the activity significantly. In contrast, mutation of a postulated adjacent AP-1 site had no effect on the p199 promoter driven reporter-gene activity.

EMSA analysis showed that the individual E-box probes bound specific complexes with differing affinities. E-box A was shown to bind a bHLH factor by competition with a previously characterized E-box motif. This was seen as a single strong complex, which appeared to be up-regulated in Lu-165 relative to cell lines with lower or undetectable endogenous AVP expression. Binding to the composite AB and ABAP1 probes was preferentially due to E-box binding and, again, was predominantly via the A motif. There are a number of candidate bHLH factors which may bind the A E-box to contribute to the AVP enhancer in SCLC. hASH-1 is known to be associated with neuroendocrine tumours and, most recently, has been shown to induce tumorigenesis of pulmonary neuroendocrine cells in a transgenic model [47]. The E-box A core motif CACGTG is the preferential recognition site of both USF and c-Myc transcription factors. However, the sequence context will influence which factor binds to E-box A. USF-1 is expressed in the lung during development and is known to regulate the lung-specific surfactant-protein-A gene [48], and the *myc* gene is known to be amplified in a percentage of lung tumours [49]; these therefore represent other candidate factors. The PPT –60 E-box (E) has been well-characterized and is known to bind USF [20]. We tested for some of these candidates, and antibody studies indicated that the complex binding E-box A was a USF-1/USF-2 heterodimer.

Figure 7 Protein binding to the composite AB E-box probe

Specific bands are shown with arrows (→), and competitors are all at a 50-fold excess unless stated otherwise. **(a)** Binding of bHLH factors to the AB probe in SCLC and NSCLC extracts. **(b)** and **(c)** Binding of non-bHLH factors over the B motif. **(b)** An over-exposed autoradiogram (4 day exposure) showing a protein complex specific to the B motif binding to the AB probe, following preincubation with a 100-fold excess of competitor 1 (E) to compete out E-box binding, in Lu-165 but not NCI-H460 extract. **(c)** Screening of eight SCLC cell lines, two NSCLC cell lines and two other control cell lines for binding of B-specific complexes.

With the antibodies described here, we saw the same supershifts for the PPT -60 E-box (E) probe with USF-1 and USF-2 (results not shown). However, the complexes binding A and E in Lu-165 cells also had slightly different mobility and therefore may not be identical. In addition, the E-box A binding complex was seen in both SCLC and NSCLC cell extracts, but may function differentially in the alternative cell types. For example, the complex may contain different isoforms of USF-2 [50], which would be recognized by the N-terminal antibody, or could be involved in different multiple interactions in the context of the intact promoter.

Although the predicted E-box B was not seen to function as a strong bHLH-binding motif by EMSA, it was shown to play a role in enhancer activity by site-directed mutagenesis and affected the functionality of the A E-box in reporter-gene assays. This could result from mutation of the B motif causing the loss of a factor required in a cascade of multiple interactions, or from conformational changes in the DNA affecting the interaction of factors binding at adjacent sites. Competition EMSA using probes with multiple motifs indicated that alternative specific complexes were preferentially binding via the B motif in most SCLC cell lines, which were not recognized by antibodies to common bHLH factors. Binding at the B motif was not seen in NSCLC and therefore could potentially modulate the activity of the A E-box in different lung tumours.

Interestingly, the reverse-strand B E-box sequence (CATCTG) is identified on computer-based transcription-factor-binding-site searches as pan-2 [51]. This is the rat equivalent of human E-12, which forms heterodimers with hASH-1, and this binding site also occurs within other promoters noted to have high activity in the SCLC cell lines (J. M. Coulson and P. J. Woll, unpublished work). However, our competition and antibody data suggest that the complex binding around the B motif is either not a bHLH factor or that it shows a very discrete and specific binding. Mutational analysis of the B motif is complicated by the close proximity of other sites and predicted overlapping motifs. The SCLC-specific complexes binding through B therefore remain unidentified and await further biochemical characterization, perhaps by partial purification.

The pattern of E-box binding in lung cell lines differed somewhat from that previously reported for the equivalent motifs of the rat AVP promoter in HeLa-cell extracts [12]; however, these differences are likely to be important in promoter activation in the lung tumour model. Comparisons can be drawn between our data and the CGRP promoter, where two adjacent E-boxes, CAGCTG (-935 to -928) and CAGGTC (-914 to -907), have been studied [18]. The upstream E-box was shown to bind weakly to SCLC extracts and the other to bind strongly. The latter was a strong enhancer in SCLC, but also bound proteins in all cell lines, including those which do not express CGRP. The authors concluded that multiple interactions may be required for CGRP expression, as we suggest here for the AVP promoter. It is also interesting that CGRP expression is lost in many SCLC lines cultured *in vitro*, as is also known to be the case for AVP [4,42], suggesting that loss or reduction of a regulatory bHLH *in vitro* may affect both promoters. The specific E-box motif (CAGCTG) has previously been reported to show similar binding in SCLC and NSCLC lines, although the factors were not identified [30]. This core motif is identical with that of the A and E motifs, which bind differentially. This again implies that the A E-box alone is either not sufficient to determine activity or is reliant on adjacent sequence context, which could affect function [43,44].

E-box A (-147 to -142) has recently been reported to control AVP expression in the suprachiasmatic nucleus in response

to circadian rhythm [52]. Therefore this motif does appear to be important in regulating AVP expression in different models and is likely to be bound by defined bHLH factors in different tissues or under different stimuli, mediating alternative tissue-specific AVP expression patterns. Thus SCLC-specific activation may occur through a particular bHLH complex binding to E-box A, which appears to be a USF-1/USF-2 heterodimer. In addition, on the basis of mutagenesis analysis and reporter-gene expression, the binding of a novel factor to the adjacent B motif in SCLC acts synergistically with this E-box. Although this composite motif was only active in SCLC with high endogenous AVP expression, it is unlikely to be solely responsible for activation of transcription.

In summary, our data indicate that the E-box motif A, located between -147 and -142 in the human AVP promoter, is important in SCLC and binds a complex recognized by USF antibodies. However, this motif also interacts co-operatively with an adjacent site B, which was predicted to be an E-box by sequence analysis, but which additionally binds non-bHLH complexes specifically in SCLC. In those SCLCs which express high levels of AVP, the complex at A is more strongly bound and a complex is bound at B. In SCLCs which express lower levels of AVP, the A complex is bound less strongly and an alternative complex may be bound at B. In NSCLC, and other control cells which do not express AVP, A is bound more weakly and there is no binding at B. Therefore bHLH and other factors binding in this region interact co-operatively to constitute a major enhancer in the proximal AVP promoter. Although our data suggest AP-1 is functionally redundant in this system, the pathological expression of vasopressin in SCLC appears to depend on the combination of lost repressor function in the minimal promoter [35] and the complex E-box enhancer described here.

J.M.C. was funded by the Association for International Cancer Research, St. Andrews, Scotland (AICR) and the Cancer Research Campaign. C.F. and J.Q. were funded by the Medical Research Council and the Wellcome Trust. We thank Ms. J. Edgson and Mrs. J. Stanley for their excellent technical assistance.

REFERENCES

- Murphy, D., Funkhouser, J., Ang, H. -L., Foo, N. -C. and Carter, D. (1993) *Ann. N.Y. Acad. Sci.* **689**, 91-106
- Friedmann, A. S., Mallot, K. A., Memoli, V. A., Pai, S. I., Yu, X. -M. and North, W. G. (1994) *Br. J. Cancer* **69**, 260-263
- Sethi, T. and Rozengurt, E. (1991) *Cancer Res.* **51**, 3621-3623
- Coulson, J. M., Stanley, J., Staff, D. and Woll, P. J. (1997) *Lung Cancer* **18**, A599
- Kim, J. K., Summer, S. N., Wood, W. M. and Schrier, R. W. (1996) *Mol. Cell Endocrinol.* **123**, 179-186
- Okazaki, T., Ishikawa, T., Nishimori, S., Igarashi, T., Hata, K. and Fujita, T. (1997) *J. Biol. Chem.* **272**, 32274-32279
- Kim, J. K., Summer, S. N., Wood, W. M. and Schrier, R. W. (1997) *Neuropeptides* **31**, 423-429
- Burke, Z. D., Ho, M. Y., Morgan, H., Smith, M., Murphy, D. and Carter, D. (1997) *Neuroscience* **78**, 1177-1185
- Verbeeck, M. A. E., Sutar, W. and Burbach, J. P. H. (1991) *Mol. Endocrinol.* **5**, 795-801
- Coulson, J. M., Edgson, J., Stanley, J. and Woll, P. J. (1998) *Proc. Am. Assoc. Cancer Res.* **39**, 397
- Coulson, J. M., Stanley, J. and Woll, P. J. (1999) *Br. J. Cancer* **80**, 1935-1944
- Grace, C. O., Fink, G. and Quinn, J. P. (1999) *Neuropeptides* **33**, 81-90
- Fabbro, D., Di Loreto, C., Stamerra, O., Alberto Beltrami, C., Lonigro, R. and Damante, G. (1996) *Eur. J. Cancer* **32A**, 512-517
- Schreiber, E., Himmelman, A., Malipiero, U., Tobler, A., Stahel, R. and Fontana, A. (1992) *Cancer Res.* **52**, 6121-6124
- Nakai, S., Kawano, H., Yodate, T., Nishi, M., Kuno, J., Nagata, A., Jishage, K., Hamada, H., Fujii, H., Kawamura, K., Shiba, K. and Noda, T. (1995) *Genes Dev.* **9**, 3109-3121
- Leblond-Francillard, M., Picon, A., Bertagna, X. and de Keyser, Y. (1997) *J. Clin. Endocrinol.* **82**, 89-94
- Therrien, M. and Drouin, J. (1993) *Mol. Cell Biol.* **13**, 2342-2353

- 18 Ball, D. W., Compton, D., Nelkin, B. D., Baylin, S. B. and de Bustros, A. (1992) *Nucleic Acids Res.* **20**, 117–123
- 19 Paterson, J. M., Morrison, S. P., Dobson, S. P., McAllister, J. and Quinn, J. P. (1995) *Neurosci. Lett.* **191**, 185–188
- 20 Paterson, J. M., Morrison, C. F., Mendelson, S. C., McAllister, J. and Quinn, J. P. (1995) *Biochem. J.* **310**, 401–406
- 21 Borges, M., Linnoila, R. I., van de Velde, H. J. K., Chen, H., Nelkin, B. D., Mabry, M., Baylin, S. B. and Ball, D. W. (1997) *Nature (London)* **386**, 852–855
- 22 Ball, D. W., Azzoli, C. G., Baylin, S. B., Chi, D., Dou, S., Donis-Keller, H., Kumaraswamy, A., Borges, M. and Nelkin, B. D. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5648–5652
- 23 Rostomily, R. C., Bermingham-McDonogh, O., Berger, M. S., Tapscott, S. J., Reh, T. A. and Olson, J. M. (1997) *Cancer Res.* **57**, 3526–3531
- 24 Chen, H., Udelsman, R., Zeiger, M. A. and Ball, D. W. (1997) *Oncol. Rep.* **4**, 775–778
- 25 Chen, H., Thiagalingam, A., Chopra, H., Borges, M. W., Feder, J. N., Nelkin, B. D., Baylin, S. B. and Ball, D. W. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5355–5360
- 26 Sun, X. H., Copeland, N. G., Jenkins, N. A. and Baltimore, D. (1991) *Mol. Cell. Biol.* **11**, 5603–5611
- 27 Israel, M. A., Hernandez, M. C., Florio, M., Andres-Barquin, P. J., Mantani, A., Carter, J. H. and Julin, C. M. (1999) *Cancer Res.* **59**, 1726s–1730s
- 28 Ding, J. M., Carver, W. C., Terracio, L. and Buggy, J. (1994) *Mol. Brain Res.* **21**, 247–255
- 29 Ying, Z., Reisman, D. and Buggy, J. (1996) *Mol. Brain Res.* **39**, 109–116
- 30 Risse-Hackl, G., Adamkiewicz, J., Wimmel, A. and Schuermann, M. (1998) *Oncogene* **16**, 3057–3068
- 31 Yoon, S. O. and Chikaraishi, D. M. (1992) *Neuron* **9**, 55–67
- 32 Paterson, J. M., Mendelson, S. C., McAllister, J., Morrison, C. F., Dobson, S., Grace, C. and Quinn, J. P. (1995) *Neuroscience* **66**, 921–932
- 33 Quinn, J. P. (1996) *Prog. Neurobiol.* **50**, 363–379
- 34 Mendelson, S. C., Morrison, J., McAllister, J. M., Paterson, J. M., Dobson, S. P., Mulderry, P. K. and Quinn, J. P. (1995) *Neuroscience* **65**, 837–847
- 35 Coulson, J. M., Fiskerstrand, C. E., Woll, P. J. and Quinn, J. P. (1999) *Cancer Res.* **59**, 5123–5127
- 36 Fiskerstrand, C. E., Newey, P., Ebrahimi, B., Gerrard, L., Harrison, P., McGregor, G. P. and Quinn, J. P. (1999) *Biochem. J.* **342**, 847–852
- 37 Prestidge, D. S. (1996) *Comput. Appl. Biosci.* **12**, 157–160
- 38 Lechner, J. F., Haugen, A., McClendon, I. A. and Petits, E. W. (1982) *In Vitro* **18**, 633–642
- 39 Dignam, J. D., Martin, P. L., Shastry, B. S. and Roeder, R. G. (1983) *Methods Enzymol.* **101**, 582–598
- 40 Morrison, C. F., McAllister, J., Dobson, S. P., Mulderry, P. K. and Quinn, J. P. (1994) *Mol. Cell Neurosci.* **5**, 165–175
- 41 Terasaki, T., Matsuno, Y., Shimosato, Y., Yamaguchi, K., Ichinose, H., Nagatsu, T. and Kato, K. (1994) *Jpn. J. Cancer Res.* **85**, 718–722
- 42 Verbeeck, M. A. E., Elands, J. P. M., de Leij, L. F. M. H., Buys, C. H. C. M., Carney, D. N., Belper, G., Roebroek, A. J. M., Van de Ven, W. J. M. and Burbach, J. P. H. (1992) *Pathobiology* **60**, 136–142
- 43 Fisher, F. and Goding, C. R. (1992) *EMBO J.* **11**, 4103–4109
- 44 Fisher, F., Crouch, D. H., Jayaraman, P. S., Clark, W., Gillespie, D. A. F. and Goding, C. R. (1993) *EMBO J.* **12**, 5075–5082
- 45 Quinn, J. P. and Farina, A. R. (1991) *FEBS Lett.* **286**, 225–228
- 46 Baillie-Johnson, H., Twentyman, P. R., Fox, N. E., Walls, G. A., Workman, P., Watson, J. V., Johnson, N., Reeve, J. G. and Bleehen, N. M. (1985) *Br. J. Cancer* **52**, 495–504
- 47 Linnoila, R. I., Ball, D. W., Baylin, S. B., DeMayo, J. and DeMayo, F. (1999) *Proc. Am. Assoc. Cancer Res.* **40**, 1337
- 48 Gao, E., Wang, Y., Alcorn, J. I. and Mendelson, C. R. (1997) *J. Biol. Chem.* **272**, 23398–23406
- 49 Johnson, B. E., Russell, E., Simmons, A. M., Phelps, R., Steinberg, S. M., Idhe, D. C. and Gazdar, A. F. (1996) *J. Cell. Biochem. Suppl.* **24**, 210–217
- 50 Viollet, B., Lefrancois-Martinez, A. M., Henrion, A., Kahn, A., Raymondjean, M. and Martinex, A. (1996) *J. Biol. Chem.* **271**, 1405–1415
- 51 Whelan, J., Cordle, S. R., Henderson, E., Weil, P. A. and Stein, R. (1990) *Mol. Cell Biol.* **10**, 1564–1572
- 52 Jin, X., Shearman, L. P., Weaver, D. R., Zylka, M. J., De Vries, G. J. and Reppert, S. M. (1999) *Cell* **96**, 57–68

Received 17 May 1999/10 September 1999; accepted 12 October 1999