Differential Activation of Lung-specific Genes by Two Forkhead Proteins, FREAC-1 and FREAC-2*

(Received for publication, November 9, 1995, and in revised form, December 12, 1995)

Marika Hellqvist, Margit Mahlapuu†, Lena Samuelsson, Sven Enerbäck, and Peter Carlsson§

From the Department of Molecular Biology, The Lundberg Laboratory, Göteborg University, Medicinaregatan 9C, S-413 90 Göteborg, Sweden

We describe the cDNA sequences for two human transcription factors, Forkhead RElated Activator (FREAC)-1 and -2, that belong to the forkhead family of eukaryotic DNA binding proteins. FREAC-1 and -2 are encoded by distinct genes, are almost identical within their DNA binding domains and in the COOH termini, but are otherwise divergent. Cotransfections with a reporter carrying FREAC binding sites showed that both proteins are transcriptional activators, and deletions located the activation domains to the COOH-terminal side of the forkhead domains. Expression of FREAC-1 and FREAC-2 is restricted to lung and placenta. We show that the promoters of genes for lung-specific proteins such as pulmonary surfactant proteins A, B, and C (SPA, SPB, and SPC) and the Clara cell 10-kDa protein (CC10) contain potential binding sites for FREAC-1 and FREAC-2. DNaseI footprinting verified that FREAC proteins bind to the predicted sites in the CC10 and SPB promoters. While an SPB promoter construct could be transactivated by both FREAC-1 and FREAC-2, CC10 was only activated by FREAC-1. Efficient activation of CC10 by FREAC-1 is shown to be specific for a lung cell line with Clara cell characteristics (H441) and to involve a region of the FREAC-1 protein unable to activate in other cell types.

Regulated gene expression depends on the concerted action of sequence-specific DNA binding proteins. Several structural motifs have been described that interact with DNA in a sequence-dependent manner and which define families of DNA binding proteins capable of regulating the initiation of transcription. Each family is defined by the structure of its DNA binding domain, but in many cases the members share other properties as well, such as the ability to heterodimerize or to convey certain intracellular signals.

The forkhead motif is a 100-amino acid DNA binding domain that defines a family of transcription factors found in metazoans and Saccharomyces. X-ray crystallography of the forkhead domain from HNF3α revealed a three-dimensional structure that is a variation on the helix-turn-helix motif (Clark et al., 1993). The forkhead domain binds DNA as a monomer and contains two loops (or wings) on the COOH-terminal side of the helix-turn-helix, which has given the structure the name “the winged helix” (Brennan, 1993; Clark et al., 1993; Lai et al., 1993). Binding of the forkhead proteins FREAC-3 and FREAC-4 to their cognate sites results in bending of the DNA at an angle of 80–90° (Pierrou et al., 1994). Selection of binding sites from random sequence oligonucleotides has shown that a number of forkhead proteins share the requirement for a RTAAAYA core sequence to bind with high affinity to DNA (Overdier et al., 1994; Pierrou et al., 1994; Kaufmann et al., 1995). Sequences flanking the core on both sides and minor variations within the core provide the specificity unique to each forkhead protein.

Many forkhead genes have been isolated based on their homology to the first identified members of this family: forkhead from Drosophila (Weigel et al., 1989; Weigel and Jäckle, 1990) and HNF3α from rat (Lai et al., 1990); little is known about their function (Bork et al., 1992; Häcker et al., 1992; Clevidence et al., 1993; Kaestner et al., 1993; Pierrou et al., 1994). Developmental mutants in Drosophila (Grossniklaus et al., 1992; Häcker et al., 1992), Caenorhabditis elegans (Miller et al., 1993), and zebrafish (Strähle et al., 1993) have been shown to be caused by mutations in genes that contain the forkhead homology, and several lines of evidence prove the importance of this gene family for the embryonic development of mammals as well. Targeted disruption of the mouse genes for HNF3α (Ang and Rossant, 1994; Weinstein et al., 1994) and BF-1 (Xuan et al., 1995) cause severe malformation of the central nervous system. The nude mice mutant, which causes defective development of the thymus and hair follicles, results from deletions within the forkhead gene whn (Nehls et al., 1994).

The oncogenic potential of forkhead proteins was first demonstrated by the isolation of qin, a retroviral oncogene from Avian sarcoma virus 31 (Li and Vogt, 1993) with homologies to the mammalian forkhead gene BF-1 (Tao and Lai, 1992). The t(2;13)(q35;q14) translocation associated with alveolar rhabdomyosarcoma fuses part of the PAX3 gene with a forkhead gene named FKHR or ALV (Gallii et al., 1993; Shapiro et al., 1993). Fusion transcripts produced from the chimerical gene give rise to a protein where the activation domain of PAX3 is replaced by the COOH-terminal part of FKHR (Gallii et al., 1993; Fredericks et al., 1995). A similar situation exists in the t(X;11) translocation observed in a case of acute lymphocytic leukemia where the forkhead gene AFXI is fused with the gene for the putative transcription factor HTRXI (Parry et al., 1994).

Forkhead proteins are also involved in the control of genes expressed in terminally differentiated cells. The best studied examples are the HNF3 proteins, which have been found to regulate a number of genes expressed in liver or other endodermal tissues (reviewed by Costa (1994)).

We have previously described partial cDNA and genomic clones for seven human genes encoding Forkhead RElated Activator (FREAC)-1 to -7 (Pierrou et al., 1994). Two of these genes, FREAC-1 and FREAC-2, are only expressed in lung and placenta. In this paper, we report the cDNA sequences for FREAC-1 and FREAC-2. We identify binding sites for the FREAC-1 and -2 proteins in the promoter regions of several...
l lung-specific genes. Although both FREAC proteins are potent transcriptional activators and bind with high affinity to the promoter of the gene for Clara cell 10-kDa protein (CC10), only FREAC-1 activates the CC10 gene, and this activation occurs only in a lung cell line with Clara cell-like characteristics.

**EXPERIMENTAL PROCEDURES**

Isolation and Sequencing of cDNA and Genomic Clones—Isolation of the original cDNA clones for FREAC-1 and FREAC-2 have been described previously (Pierrou et al., 1994). To obtain full-length clones, three human lung cDNA libraries (Clonetech and Stratagene) were screened using probes derived from the original isolates of FREAC-1 and FREAC-2. Inserts from positive phages were subcloned and sequenced on both strands on a Pharmacia A.L.F. sequencer using T7 polymerase (Pharmacia) and either fluorescein dATP or fluorescein-labeled primer. Some regions were resequenced with α-32PdATP and Sequenase (U. S. Biochemical Corp.) or Taq polymerase (Boehringer Mannheim). One particularly difficult region was resolved with the Maxam Gilbert method.

A clone for mouse FREAC-1 was isolated by screening a genomic mouse library with a human FREAC-1 cDNA probe. Relevant fragments were identified by Southern blotting, subcloned, and sequenced.

Northern Blot—Total RNA preparation and poly(A) selection was performed as described previously (Pierrou et al., 1994). 5 μg per lane of poly(A) RNA was loaded on a 1% agarose/-formaldehyde gel, which was blotted onto Hybond filters (Amersham), hybridized at 47°C in 50% formamide, and washed at high stringency (0.1 × SSC, 0.1% SDS, 65°C). Probes were prepared from the non-considered, 3′-untranslated parts of the FREAC-1 and FREAC-2 genes.

Expression Constructs—The FREAC-1 expression plasmid was created as follows: a FREAC-1 cDNA containing the 5′-end and sequences down to position 1040 was cloned into the EcoRI site of pTZ19R (Pharmacia). To remove the 5′-untranslated sequence and insert a BamHI site in front of the initiation codon, this plasmid was digested with Ncol and Smal, filled in with Klenow enzyme, and religated. The resulting plasmid was linearized with EcoRI and filled in with Klenow enzyme; the FREAC-1 fragment was cut out with BamHI and ligated between the BamHI and Smal sites of pEV3S (Matthias et al., 1989). To add the rest of the FREAC-1 coding sequence, a NotI-EcoRI cDNA fragment spanning position 435-1482 was inserted between the NotI and the filled-in Acc5I site in the pEV3S recombinant.

The FREAC-2 expression plasmid was created by filling in an EcoRI fragment spanning nt 1-1880 with Klenow enzyme and inserting it into the Smal site of pEVRF0 (Matthias et al., 1989). All plasmids expressing truncated versions of FREAC proteins were generated through deletions of parts of the FREAC-1 and -2 genes from the full-length constructs with exonuclease Bal3I or restriction enzymes.

Luciferase Reporter Constructs—The apob-luc reporter was created by cloning the minimal apob promoter (−45 to +121) and the UMS transcriptional terminator from pBCat-45 (Carlsson and Bjursell, 1989) into pGL2-Basic (Promega) and inserting a of the ApaI linker into the SpI site between UMS and the apob promoter. Four tandem copies of the 3′- and 5′-end binding site between UMS and the apoB promoter. Four tandem copies of the transcriptional terminator from pBPcat-45 (Carlsson and Bjursell, 1991) containing 0.2% Nonidet P-40. Binding reactions (15 μl) contained 3 μl of cell extract, 1 μg of poly(dI-dC), and 20,000 cpm of the 32P-labeled, double-stranded oligonucleotide GATCCAACGTAAACAATCCGA; lower, GATCTCGATGGTTACAGTTG. 3′ and 5′-end binding sites, located 19 nt from the 5′- and 3′-end sequences, are followed by an A/T-rich context for initiation of translation (Kozak, 1989), having gyrimidines in positions 3 and +4 and a purine in position −1. The second ATG codon in the FREAC-1 cDNA is located 94 nt from the 5′-end. This codon is positioned in a near-optimal context for translational initiation, and a polypeptide translation at this codon will proceed into the forkhead homology in the correct reading frame without intervening stop codons. We have therefore assigned this codon as the start of the concept.

RESULTS

FREAC-1 and FREAC-2 are identical in the amino-terminal DNA binding domains and in the COOH termini. To isolate full-length cDNA clones for FREAC-1 and FREAC-2, we screened cDNA libraries derived from human lung. From several overlapping clones we were able to compile a cDNA sequence for FREAC-1 of 2509 nt, excluding the poly(A) tail. On Northern blots, we have estimated the size of the FREAC-1 mRNA to 2.6 kilobases (Pierrou et al., 1994). Given an average length of the poly(A) tail of 150 nt, the FREAC-1 cDNA sequence must be very close to full-length.

The first ATG codon in the FREAC-1 cDNA sequence is located 19 nt from the 5′-end. However, this codon is situated in a poor context for position translation (Kozak, 1989), having pyrimidines in positions 3 and +4 and a purine in position −1. The second ATG codon in the FREAC-1 cDNA is located 94 nt from the 5′-end. This codon is positioned in a near-optimal context for translational initiation, and a polypeptide translation at this codon will proceed into the forkhead homology in the correct reading frame without intervening stop codons. We have therefore assigned this codon as the start of the concept.

In vitro Mutagenesis—The 5′- and 3′-FREAC binding sites in the CC10 promoter were mutagenized with a three-step PCR method (Nelson and Long, 1989) using the following primers: mutagenic primers, TACCTCCATGTCAATACAGCCGCAATCTGTTTACAAAC and TGATTGGAGTCAATGAAAGCTAGCAGCATGTTTGGAGGAGGAG; chimerical primers, CCAATGCATATGCTGGTAGCTTTCGAGGCTGGTCTTCCG; and the full-length double mutant was then amplified with the two flanking primers. Mutant promoters were cloned into pGL2-Basic as described above and sequenced in their entirety.

DNAse Footprinting—The CC10 promoter fragments were end labeled with [γ-32P]ATP and polyadenyl kinase in the HindIII sites of CC10-luc and SPB-luc, respectively. The probes were released by Xhol cleavage and purified by gel electrophoresis. DNAse footprinting from the start of transcription (Pierrou et al., 1994) using up to 10 ng of FREAC-2GST fusion protein and 20,000 cpm Cerenkov of the labeled promoter fragment.

Cell Culture, Transfections, and Luciferase Assays—All cells were grown in Dulbecco's modified Eagle's medium with 10% fetal calf serum on collagen-coated plastic. Liposome-mediated transfections were performed essentially as described previously (Carlsson and Bjursell, 1989). A typical transfection contained 300 ng of luciferase reporter plasmid, a total of 300 ng of cotransfected plasmid (variable amounts of FREAC expression plasmids and complementary amounts of pEVRF expression vector), and 2 μg of Lipofectin or LipofectAmine (Life Technologies, Inc.) in 560 μl of OptiMEM (Life Technologies, Inc.). The mix was added to a subconfluent monolayer of cells in a 16-mm well. After overnight incubation, 2 ml of standard medium was added, and the incubation was continued for 24 more hours. Cell harvest and luciferase assay was performed according to Promega (Technical Bulletin No. 101).

Gel Shift Assays—COS-7 cells were transfected with plasmids encoding FREAC-1 and FREAC-2 as well as the different truncated versions of these proteins. 48-h post-transfection cells were harvested, and extracts were made by freeze-thawing in 5× elution buffer (Schmitz and Båserle, 1991) containing 0.2% Nonidet P-40. Binding reactions (15 μl) contained 3 μl of cell extract, 1 μg of poly(dI-dC), and 20,000 cpm of the 32P-labeled, double-stranded oligonucleotide GATCCAACCTGAA- CAATCGAGAG. The reaction was incubated for 15 min at room temperature and resolved on a 6% polyacrylamide (29:1 gel in Tris-glycine buffer (25 mM Tris, 190 mM glycine, 1 mM EDTA) with 5% glycerol at 15 V/cm, +4°C for 90 min.

The size of the FREAC-2 mRNA was estimated to be 2.4 kilobases on Northern blots. Despite extensive library screening, we were unable to isolate more than 1964 nt of FREAC-2 cDNA from overlapping clones. For allowing 100–200 nt of poly(A) tail leaves approximately 300 nt missing from the 5′-
end of the cDNA sequence. The reading frame defined by the forkhead homology is open from the beginning of the cDNA sequence and contains no ATG codon before the start of the forkhead motif. Thus, the 408 amino acids of deduced FREAC-2 sequence (Fig. 1B) does not represent the full-length protein but lacks the proper amino terminus. The 3'-untranslated sequence of 740 nt contains a (CA)₉ repeat in antisense orientation around nt 1580 and an AATAAA polyadenylation signal 17 nt upstream of the poly(A) addition site.

A comparison of the amino acid sequences of FREAC-1 and FREAC-2, derived from conceptual translation of the cDNAs, suggests that the two genes have evolved from a common ancestor (Fig. 1C). Within the forkhead domain and immediately adjacent sequences, the two proteins are virtually identical; three conservative amino acid substitutions, one serine/threonine and two serine/alanine, occur within 112 residues. Since the forkhead domain is responsible for DNA binding, it seems reasonable to assume that the two proteins have identical, or near identical, DNA binding specificity. FREAC-2 extends further on the amino-terminal side of the forkhead domain than FREAC-1 and has a serine-rich stretch in this region with 15 serines out of 18 residues. Also on the carboxyl-terminal side of the forkhead domain does the FREAC-2 sequence contain several homopolymeric runs of amino acids such as serine, glycine, and histidine. The central parts of the proteins are divergent, although islands of homology indicate that the sequences have a common origin. In the carboxyl-termini the similarity is again more obvious, and the eight last amino acids...
of FREAC-1 and FREAC-2 are identical. Except for the homopolymeric runs of certain amino acids, which are common among transcription factors, no conserved sequence motifs were found outside the forkhead homology when the amino acid sequences were used to search the data bases.

A comparison of the FREAC-1 and -2 DNA sequences with other known forkhead genes revealed that FREAC-1 is very similar to H-FH-8 from mouse (Cleveland et al., 1994). The cDNA sequence similarity, the matching tissue distribution of expression (Cleveland et al., 1994; Pierrou et al., 1994), and the fact that FREAC-1 and H-FH-8 are located at homologous chromosomal positions in man and mouse (Avraham et al., 1995; Larsson et al., 1995) suggested that H-FH-8 and FREAC-1 are homologous genes. However, the predicted amino acid sequences of H-FH-8 and FREAC-1 differ on both sides of the forkhead motif due to insertions or deletions in the H-FH-8 cDNA sequence, compared to that of FREAC-1. This leads to frameshifts in five different positions throughout the coding sequence. To assess whether the apparent discrepancy in use of reading frame was due to a species difference or sequencing errors, we isolated a genomic clone for the mouse homologue of FREAC-1 and sequenced the relevant regions. As shown in Fig. 1A, the human and mouse FREAC-1 sequences are indeed co-linear, and the aberrant amino acid sequence of H-FH-8 is most likely explained by sequencing errors introducing frame-shifts in five positions of the published H-FH-8 cDNA sequence (Fig. 1A).

Human Lung Cell Lines Express FREAC-2 but Not FREAC-1—Northern blots with RNA from different cell lines were hybridized with probes specific for FREAC-1 and FREAC-2. Three cell lines derived from human lung were found to express low levels of FREAC-2 mRNA (Fig. 2). A549 (Lieber et al., 1976) is a lung carcinoma cell line, and the other two are fetal lung cell lines from the third (WI-38) and fourth (IMR-90) month of pregnancy (Nichols et al., 1977). In contrast, none of the tested cell lines express FREAC-1.

FREAC-1 and FREAC-2 Have COOH-terminal Transcriptional Activation Domains—To investigate the effect on transcription from a nearby promoter of binding by FREAC proteins, we transfected cells with plasmid constructs expressing FREAC-1 and FREAC-2. To monitor the activity of FREAC proteins in the transfected cells, we cotransfected a luciferase reporter construct containing four FREAC-2 binding sites upstream of a minimal apoB promoter (Fig. 4A). The sequence of the FREAC-2 sites used in this construct was based on the consensus sequence for FREAC-2 determined by site selection from random sequence oligonucleotides (Pierrou et al., 1994, 1995).

When the luciferase activity produced by the 4× FREAC-luc reporter was compared to that of the parental apoB-luc, we found that the presence of four FREAC binding sites enhanced promoter activity, even without cotransfection with FREAC expression plasmids (data not shown). The activation varied between cell lines and indicates that endogenous transcriptional activators capable of binding to the FREAC sites are present in a variety of cell types where no expression of FREAC-1 or FREAC-2 can be detected. This is not surprising since our results with binding site selection (Pierrou et al., 1994) show that forkhead proteins are closely related with regard to sequence specificity and that the differences are often quantitative rather than qualitative. Furthermore, the large size of the forkhead gene family and the wide tissue distribution of its expression suggest that there may be forkhead proteins present in virtually every cell type.

Fig. 2. Human lung cell lines express FREAC-2. Northern blot is shown with RNA from three human lung cell lines hybridized with a probe specific for FREAC-2. No expression could be detected when a FREAC-1 probe was hybridized to an identical blot. kb, kilobases.
In the human SPB and rat CC10 promoters, the predicted binding sites are located in regions to which regulatory function has been assigned based on transfections with reporter constructs and to which nuclear proteins from lung cells have been shown to bind (Stripp et al., 1992; Bohinski et al., 1993; Bohinski et al., 1994). Therefore, we chose to investigate

**FIG. 3.** FREAC proteins bind to the promoter regions of lung-specific genes. A, DNaseI footprinting of the rat CC10 and human SPB promoters. CC10 promoter fragments with either wild type (wt) sequence or mutated in the 5', 3', or both (dbl) sites were footprinted in the absence (−) or presence of increasing amounts of FREAC-2-glutathione S-transferase (GST). The SPB promoter was footprinted with an amount of FREAC-2/GST that corresponds to the highest amount used with CC10 (approximately 10 ng). B, sequence of the two FREAC binding sites in the rat CC10 promoter. Arrows indicate the FREAC core motifs, and brackets indicate the sequences protected from DNaseI digestion. The three nucleotides in each site that were targeted with in vitro mutagenesis are marked in bold, and the actual sequences of the mutants are shown below. C, summary of putative FREAC binding sites in the promoter regions of four lung-specific genes from four mammalian species. Numbers indicate the position relative to the transcriptional start site, and rev indicates that the sequence from the antisense strand is shown.
the effect of FREAC expression on the activity of these two promoters. Fragments from the rat CC10 and human SPB promoters that had proven to be active in transient transfections were isolated by PCR. DNaseI footprinting was used to test if the predicted sites would bind FREAC-2. As shown in Fig. 3A, specific binding of FREAC-2/GST was observed for two closely positioned sites in the CC10 promoter and for one site in the SPB promoter.

To investigate if the observed binding of FREAC proteins to the promoter regions of the SPB and CC10 genes influenced transcription, we transfected luciferase reporter constructs, driven by these promoters (Fig. 4A), together with plasmids expressing FREAC-1 and FREAC-2.

The CC10 Promoter Is Activated by FREAC-1 but Not FREAC-2. Specifically in H441 Cells—Fig. 5 shows the results from cotransfection of CC10-luc with variable amounts of plasmids expressing FREAC-1, FREAC-2, or truncated versions of either gene.

In the lung cell line H441 is CC10-luc approximately 40-fold more active than the promoterless luciferase plasmid pGL2-Basic (data not shown). Cotransfection with a plasmid expressing full-length FREAC-1 activated the CC10 promoter in this construct up to 20-fold above the basal level. FREAC-1(1–326), which encodes a protein with 28 amino acids deleted from the COOH terminus and which repressed 4xFREAC-luc in COS-7, activated CC10-luc at least as efficiently as the full length protein (25-fold), and was much more effective when limiting concentrations of plasmid was used. No activation was ob-

---

**Fig. 4. FREAC-1 and FREAC-2 have COOH-terminal activation domains.** A, luciferase reporter constructs (top) and proteins encoded by the different FREAC expression plasmids (bottom). B, relative luciferase activity produced by COS-7 cells transfected with 4xFREAC-luc and various FREAC expression plasmids. C, gel shift assay with a FREAC probe and extracts from COS-7 cells transfected with different FREAC expression plasmids or empty pEVRF expression vector. The shifted complexes present in the "vector" lane represent endogenous COS-7 cell proteins capable of binding to the probe. In the FREAC-2, FREAC-2(1–242), FREAC-1, and FREAC-1(1–326) lanes, faster migrating bands are present in addition to the full-length complexes, which result from protease cleavage at hypersensitive sites immediately COOH-terminal of the forhead domains of both FREAC-1 and FREAC-2.
This activity was extinguished by cotransfection with plasmids expressing truncated, non-activating versions of either FREAC-1 or FREAC-2. The degree of repression depended on the amount of FREAC plasmid transfected and was equally efficient for FREAC-1(1–117) and FREAC-2(1–242). This result suggests that the truncated FREAC proteins repress transcription through competition with endogenous proteins for the same binding sites. It also shows that, in this repression of the CC10 promoter, FREAC-2 is as efficient as FREAC-1. Thus, the selective activation of the CC10 promoter by FREAC-1 in H441 cells is unlikely to reflect a difference in DNA binding between FREAC-1 and FREAC-2. Rather, it implies that other factors present in H441 cells synergize with FREAC-1 but not FREAC-2.

FREAC-1 and FREAC-2 only activated the CC10 promoter 1.8–2.5-fold in HC11 cells. This modest activation was seen using low levels of FREAC expression plasmids, and at higher levels activity again declined, possibly due to squelching. In contrast to what we observed in H441 cells, FREAC-2 was here a slightly better activator than FREAC-1. Finally, FREAC-1(1–326), which in H441 cells activated as well as, or better than, full-length FREAC-1, did not activate in HC11 cells. Instead, it repressed the CC10 promoter with approximately the same efficiency as FREAC-1(1–117) or FREAC-2(1–242).

Taken together, these results suggest that entirely different mechanisms are behind the efficient activation of the CC10 promoter by FREAC-1 seen in H441 cells and the limited activation by both FREAC-1 and FREAC-2 in HC11 cells. In H441, the CC10 promoter appears to be in a context that makes it susceptible to activation by FREAC-1 but not FREAC-2, an activation which is independent of the last 28 amino acids but requires a region between amino acids 118 and 326 in the FREAC-1 sequence. In HC11 cells, a weak activation is produced by both FREAC-1 and FREAC-2, and in the case of FREAC-1 this activation depends on the integrity of the last 28 amino acids.

To verify that all the expression constructs produced proteins that were correctly folded and able to bind DNA, we prepared extracts from transfected COS-7 cells and analyzed these for the presence of FREAC proteins with a gel shift assay. As seen in Fig. 4C, all the truncated proteins as well as the full-length proteins are expressed and bind to a FREAC site oligonucleotide in this assay. Despite the fact that it is the best activator of the CC10 promoter in H441 cells, FREAC-1(1–326) was consistently present in lower amounts than the other proteins in extracts from transfected cells.

Activation of the CC10 Promoter by FREAC-1 Is Mediated by Both FREAC Binding Sites—To verify that the activation of the CC10 promoter by FREAC-1 is mediated by the two identified binding sites and to investigate each site's relative importance, we mutated the sites separately and also combined the mutations in a double mutant. DNaseI footprinting on the mutant promoters (Fig. 3A) verified that the mutations abolished binding of FREAC-2/GST. Table I summarizes the effect of the mutations on expression from CC10-luc in H441 and HC11 cells. In the absence of cotransfection, the FREAC binding sites appear to contribute little to the activity of the CC10 promoter. Mutation of the 5'-site reduced luciferase activity approximately by half, whereas knocking out binding to the 3'-site actually increased expression from CC10-luc, and the double mutant was approximately as active as the wild type promoter. Similar results of mutagenesis in this region of the CC10 promoter were reported by Sawaya and Luse (1994). Thus, no dramatic effects of the mutations were observed, which, together with the failure of truncated FREAC proteins to repress expression from the wild type CC10 promoter, suggest that
each one of the five frameshifts results from an error in the sequence of HFH-8 differs significantly from that of FREAC-1. FREAC-1 frame of the or insertions of one or two nucleotides change the reading fairly evenly distributed. In five locations, however, deletions FREAC-1: 90% homology in the coding region with differences, FREAC-1 are targets for FREAC proteins. thesurfactantproteinBgene,weverifythattheidentifiedsitesinvolved in regulation of lung-specific genes. In this paper, we showthatanumberofgenesspecificallyexpressedinthelungshowthatanumberofgenesspecificallyexpressedinthelung

| Reporter plasmid | Basal activity | -Fold activation by FREAC-1 |
|------------------|---------------|---------------------------|
|                  | H441 | HC11 | H441 | HC11 |
| CC10 wt          | 100  | 100  | 19.3 | 1.4  |
| CC10 5′          | 56   | 82   | 9.1  | 1.2  |
| CC10 3           | 130  | 213  | 11.5 | 1.3  |
| CC10 dbd         | 94   | 205  | 3.1  | 1.2  |

H441 cells lack proteins that can efficiently activate transcription through the FREAC sites. The ability of the CC10 promoter to be activated by FREAC-1 was, however, reduced by the mutations. Both single mutations were about equally effective in reducing the responsiveness to FREAC-1, and the double mutant showed the lowest level of induction. This result implies that FREAC-1 is able to activate the CC10 promoter from a single site and that no synergy exists between the two sites.

SPB Can be Activated by Both FREAC-1 and FREAC-2—In HC11 cells, both FREAC-1 and FREAC-2 activated SPB-luc (Fig. 6), and FREAC-2 is the better activator (6.5-fold). Basal activity of SPB-luc is higher in H441 than in HC11, and in this cell type FREAC-1 is more efficient than FREAC-2. FREAC-1(1–117) and FREAC-2(1–242) repressed expression to one-sixth of the basal level, which indicates that endogenous proteins in H441 cells activate the SPB promoter through the same binding site as that targeted by FREAC-1 and -2. Although cotransfection of full-length constructs of both FREAC-1 and -2 produce higher activities than the truncated, the only construct capable of giving a net increase of expression from SPB-luc in H441 cells was FREAC-1(1–326). Thus, the ability of FREAC-1(1–326) to activate appears to be a general characteristic of H441 cells rather than a phenomenon specific for the CC10 promoter.

TABLE I

| Reporter plasmid | Basal activity | -Fold activation by FREAC-1 |
|------------------|---------------|---------------------------|
|                  | H441 | HC11 | H441 | HC11 |

DISCUSSION

We have cloned the cDNA:s for two novel transcription factors, FREAC-1 and FREAC-2, which belong to the forkhead family. Previous work demonstrated expression of FREAC-1 and FREAC-2 only in lung and placenta. The restricted expression pattern suggested that FREAC-1 and FREAC-2 could be involved in regulation of lung-specific genes. In this paper, we show that a number of genes specifically expressed in the lung epithelium contain potential binding sites for FREAC proteins. For two of these genes, the Clara cell 10-kDa protein gene and the surfactant protein B gene, we verify that the identified sites are targets for FREAC proteins.

Recently, the cDNA sequence of the mouse homologue of FREAC-1, HFH-8, was published (Clevidenue et al., 1994). The nucleotide sequence of HFH-8 is very similar to that of FREAC-1: 90% homology in the coding region with differences fairly evenly distributed. In five locations, however, deletions or insertions of one or two nucleotides change the reading frame of the HFH-8 cDNA sequence compared to that of FREAC-1. As a consequence, the predicted amino acid sequence of HFH-8 differs significantly from that of FREAC-1. Sequencing of a genomic clone for mouse FREAC-1 showed that each one of the five frameshifs results from an error in the HFH-8 sequence and that FREAC-1 from mouse and man are nearly identical throughout the coding sequence.

The similarity between sequences of FREAC-1 and FREAC-2, at DNA as well as protein level, indicate a close evolutionary relationship. On the other hand, the lack of homology in the amino-terminal and central parts of the proteins shows that there has been ample time for the two genes to diverge. In other words, strong selective pressures must be behind the almost perfect conservation of the DNA binding domains and the extensive similarities in the COOH-terminal part. That the duplication of a presumed ancestral gene was not a recent event is supported by the fact that FREAC-1 and FREAC-2 are located on different chromosomes (Larsson et al., 1995).

The sequence homology in conjunction with the similarity in tissue distribution of expression suggested that FREAC-1 and FREAC-2 may be functionally redundant. However, the qualitative difference in their ability to activate the CC10 promoter shows that although both proteins are transcriptional activators with similar or identical DNA binding specificity, they are functionally distinct. It also stresses the importance of interactions other than DNA binding for the specificity of FREAC proteins. The FREAC-2 clone that we have used in the transfection experiments does not encode the full-length protein; some amino acids are missing from the amino terminus. Thus, we cannot exclude the possibility that a full-length FREAC-2 protein would exhibit other characteristics. However, this does not change the fact that the powerful COOH-terminal activation domains present in both FREAC-1 and FREAC-2 exhibit differential activation properties. Whereas both proteins potently activate a reporter construct in a heterologous cell type (COS-7), only FREAC-1 is capable of activating the CC10 promoter in a lung cell line (H441). The behavior of the FREAC-1(1–326) deletion mutant also shows that in these two contexts, the mechanisms of activation by FREAC-1 are distinct.

A comparison of the activities produced by the different deletion mutants of FREAC-1 shows that the region necessary for the cell-specific activation of CC10 is located between amino acids 326 and 356.
acid 118 and 325. This coincides with the part where FREAC-1 and FREAC-2 are most divergent. It appears that FREAC-1 and FREAC-2 have evolved to perform different biological tasks while retaining the specificity for the same DNA sites and the same organ-specific expression.

The binding sites for FREAC proteins in the CC10 promoter have been shown to bind proteins present in nuclear extracts from lung ([Stripp et al., 1992]. HNF3α and HNF3β, which are expressed in many endodermal tissues including lung ([Clevidence et al., 1993; Bingle et al., 1995], have been reported to be able to bind to these sites (Bingle and Gitlin, 1993; Sawaya et al., 1993). However, cotransfections with HNF3α and HNF3β showed no (Sawaya and Luse, 1994) or low (Bingle and Gitlin, 1993; Bingle et al., 1995) transactivation of the CC10 promoter.

The FREC site in the SPB promoter has also been shown to bind HNF3α and HNF3β (Bohinski et al., 1994) and, at least in a non-lung cell line (HepG2), does the binding of HNF3α, as well as HNF-8, mediate transcriptional activation (Clevidence et al., 1994). This picture agrees well with our observation that truncated FREAC proteins repress the SPB promoter but not the CC10 promoter in H441 cells. Thus, SPB and CC10 appear to differ not in their ability to bind the different factors but in the way they respond.

It is clear that a number of forkhead proteins are capable of binding to the FREAC sites in the CC10 promoter in vitro and in vivo. In addition to FREAC-1, FREAC-2, HNF3α, and HNF3β, two other forkhead genes, HFH-1 and HFH-4, are also expressed in the lung (Clevidence et al., 1993; Hackett et al., 1995). A more detailed analysis of temporal and spatial expression patterns will be required to understand each protein's role in regulating pulmonary genes. The example of FREAC-1 and FREAC-2 and their different effects on the CC10 promoter emphasizes the importance of context in transcription factor function. It also illustrates how interactions mediated by parts of the proteins distinct from the DNA binding domain can provide specificity. This gives us a clue to how the members of this large transcription factor family may exert their distinctive functions and how cross-talk could be avoided between proteins with overlapping DNA binding specificity.

Acknowledgments—We thank Kerstin Dahlborn for excellent technical assistance and Dr. Colin D. Bingle for communicating results prior to publication.
Differential Activation of Lung-specific Genes by Two Forkhead Proteins, FREAC-1 and FREAC-2

Marika Hellqvist, Margit Mahlapuu, Lena Samuelsson, Sven Enerbäck and Peter Carlsson

J. Biol. Chem. 1996, 271:4482-4490.
doi: 10.1074/jbc.271.8.4482

Access the most updated version of this article at http://www.jbc.org/content/271/8/4482

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 46 references, 22 of which can be accessed free at http://www.jbc.org/content/271/8/4482.full.html#ref-list-1