The Human Forkhead Protein FREAC-2 Contains Two Functionally Redundant Activation Domains and Interacts with TBP and TFIIIB*

(Received for publication, November 20, 1997, and in revised form, May 4, 1998)

Marika Hellqvist, Margit Mahlapuu, Åsa Blixt, Sven Enerbäck, and Peter Carlsson‡

Department of Molecular Biology, Göteborg University, Medicinaregatan 9C, Box 462, S-405 30 Göteborg, Sweden

Forkhead-related activator 2 (FREAC-2) is a human transcription factor expressed in lung and placenta that binds to cis-elements in several lung-specific genes. We have identified the parts of FREAC-2 responsible for trans-activation and found two functionally redundant activation domains on the C-terminal side of the DNA binding forkhead domain. Activation domain 1 consists of the most C-terminal 23 amino acids of FREAC-2 and contains a sequence motif conserved in an activation domain of another forkhead protein, FREAC-1. Activation domain 2 is built up by three synergistic subdomains in the central part of the FREAC-2 protein. FREAC-2 was shown to interact in vitro with TBP and TFIIIB. The target site for FREAC-2 on TBP was localized to the N-terminal repeat in the core domain of TBP. TFIIIB binds FREAC-2 close to the cleft between its two globular domains. The part of FREAC-2 that binds TBP was mapped to 21 amino acids in the C-terminal end of the forkhead domain. This sequence is well conserved among forkhead proteins, raising the possibility that interaction with TBP may be a general characteristic of this family of transcription factors. Overexpression of TFIIIB potentiates activation by FREAC-2 in a manner dependent on the FREAC-2 activation domains. Nuclear localization of FREAC-2 was found to depend on sequences from both ends of the forkhead domain.

Genes encoding transcription factors that contain a forkhead DNA binding domain have been found in eucaryotic organisms from the simplest unicellular fungi to mammals. It appears, however, to be in the evolution of metazoans that this gene family expanded to the complexity that we see in today’s vertebrates. Whereas, for example, the entire genome of Saccharomyces cerevisiae (1) contains 4 genes with a forkhead motif, over 50 mammalian forkhead genes have been cloned so far (reviewed in Ref. 2). Consistent with this, forkhead genes are involved in embryogenesis and pattern formation in multicellular organisms. In fact, all known examples of inactivation of forkhead genes result in more or less dramatic disturbances of embryogenesis (3–14).

How is functional specificity maintained within a gene family with so many closely related members? One level of specificity is DNA binding. Distantly related members of the forkhead family may bind to completely distinct DNA sequences (15–17). Even in cases in which the DNA binding domains are more closely related and the binding specificities are overlapping, there can—as in the case of FREAC-31 and FREAC-4 (18)—be differences in the relative affinities for a particular sequence that may have profound biological significance (18–21). There are, however, also several examples of forkhead proteins in which the DNA binding domains are so similar that it is reasonable to assume that the specificities are identical. The most extreme example is that of FREAC-4 and FREAC-9, two human forkhead proteins that are encoded by different genes and are coexpressed in certain cells. FREAC-4 and FREAC-9 have DNA binding domains with 100% identity throughout their 105 amino acids, but they have no homologies outside the forkhead domains (22). In such cases, the proteins will either be functionally redundant, as appears to be the case with the two forkhead genes in the sloppy paired locus of Drosophila (23), or specificity will reside in other functions than DNA binding. An example that illustrates that forkhead proteins expressed in the same tissue and with indistinguishable DNA binding specificities can have distinct biological functions is that of FREAC-1 and FREAC-2. Although both proteins bind to and activate the surfactant protein B promoter, the gene for another lung-specific protein, CC10, is only activated by FREAC-1 due to the presence of a cell type-specific activation domain in FREAC-1 (24).

The DNA binding function of forkhead proteins is well understood (25), but far less is known about their activities as transcriptional regulators. The majority of forkhead proteins that have been experimentally analyzed appear to be transcriptional activators, but only in a few cases have the regions of the proteins responsible for activation been identified. HNF3β activates transcription through four regions (26), and Whn, encoded by the nude locus, contains an acidic activation domain in the C terminus (27). The human forkhead protein FKHR has a potent activation domain that, when fused to the heterologous DNA binding domain of PAX3 by a translocation, is responsible for malignant transformation in alveolar rhabdomyosarcoma (28–30). Transcriptional trans-repression is known from Genesis, a forkhead protein expressed in embryonic stem cells (16), and the proto-oncogene product c-Qin or BF1 (31, 32). In the latter case, the repression domain is essential for the transforming ability of the oncogenic version of the protein, v-Qin (33, 34).

Interaction with other cellular components is necessary for transcription factors in conveying biological signals. Again, we know very little about the interactions made by forkhead proteins, but a role in the remodeling of chromatin structure as a step in transcriptional activation has been proposed (35). The Xenopus forkhead protein FAST-1 has been identified as a direct target for transforming growth factor-β/activin signaling and shown to form a complex with the transforming growth

* This work was supported by grants from The Swedish Cancer Foundation, the Fredrik and Ingrid Thuring Foundation, and the Magnus Bergvall Foundation. 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.: 46-31-7733804; Fax: 46-31-7733801; E-mail: peter.carlsson@molio.gu.se.
FREAC-2 Contains Functionally Redundant Activation Domains

factor-β signal transducer XMD2 (36). The forkhead-associated domain is a sequence motif found in some forkhead proteins, as well as in certain protein kinases, that has been proposed to participate in nuclear signaling (37).

The human forkhead gene FREAC-2 is, in the adult, expressed only in lung and placenta (19), and the encoded protein is a transcriptional activator (24). FREAC-2 binds the consensus sequence AACCTAAACAA (18, 19), and the amino acid sequence of the DNA binding domain is virtually identical to that of another human forkhead protein, FREAC-1 (24). The promoters of a number of genes that are specifically expressed in lung, including the genes for surfactant apoproteins, contain binding sites for FREAC-2 (24).

Herein, we describe mapping of functional domains in FREAC-2. We have identified the regions responsible for transcriptional activation and also shown that FREAC-2 interacts with components of the basal transcriptional machinery. Finally, we have addressed the issue of subcellular localization and identified the parts of FREAC-2 that targets the protein to the nucleus.

EXPERIMENTAL PROCEDURES

Plasmid Constructs and Mutagenesis—The FREAC-2 expression plasmid has been described elsewhere (24). Constructs expressing truncated FREAC-2 proteins were generated by creating deletions in the FREAC-2 plasmid with restriction enzymes or Bal31 exonuclease.

Triple alanine substitution mutations in the FREAC-2 coding sequence were generated with oligonucleotide directed mutagenesis, essentially as described previously (38), using the following mutagenic primers: GCCATCAAGTGGCCGGCCGCGCAAGCTCGCCGGC, (m295–297), AACAGCGCGCCTGCGCCGCGCGCAAGTTCTCTCAGCT (m319–321), TCCGACAGTCTCCTCCGGCCGCGCAAGGAGCAGCTT (m338–340), CTGGGACCTCCTCCTGCCTGCGCGCCTCACTACTCTACAGT (m360–362), and AGTCAGAGACCTCGCCTGCGCGCCTCACTACTCTACAGT (m385–387). The m173 mutant was created by insertion of a double-stranded oligonucleotide adapter (GCCGCCGCACGCCG and TCCGCGCCGCCG) into the unique DraIII site of FREAC-2. Internal, in-frame deletions in FREAC-2 were created by joining fragments from the triple alanine substitution constructs at the NotI sites introduced by mutagenesis.

A vector for expression of Gal4 fusions (pEG4) was created by inserting an in-frame XbaI fragment (CCGGAATTGTCGACTGCGGCCGCAAGCTTTCT-AGATAGCTAG) downstream of codon 147 of Gal4 in pCMV-Gal4 (39). Constructs expressing amino-terminally truncated FREAC-2 proteins fused to the Gal4 DNA binding domain were then made by cloning NotI-XbaI fragments from the triple alanine substitution constructs between the corresponding sites in pNG4. Fragments encoding internal regions of FREAC-2-activation domain were generated by polymerase chain reaction with primers tagged with SacI (forward) and XbaI (reverse) sites. These polymerase chain reaction products were inserted between the SacI and XbaI sites of pNPG and verified by sequencing.

The reporter 4×FREAC-luc (24) contains four FREAC-2 binding sites upstream of a minimal apoB promoter that drives the expression of luciferase. To create a reporter construct for Gal4 fusion proteins (Gal4-luc), six copies of a double-stranded oligonucleotide containing a Gal4 binding site (GATCCGGACTGTCCTCCGAGATC) were cloned into the BglII site of apoB (24).

FREAC-2 expression constructs were modified for in vitro transcription-translation by deleting the SacI fragment containing the CMV promoter in the pEYRFO vector (39), thereby moving the T7 promoter of the vector immediately upstream of the FREAC-2 cDNA sequence.

To express TFIIIB in mammalian cells, the human TFIIIB cDNA (40) was cloned between the BamHI and XbaI sites of pEV38 (39).

Transfections, Luciferase Assays, and Gel Shift Assays—Cell lines COS-7 and HepG2 were grown on collagen-coated plastic in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal calf serum, penicillin and streptomycin. Liposome-mediated transfections were performed as described previously (24) in 24-well tissue culture plates using Lipofectin or LipofectAMINE (Life Technologies, Inc.). Two days posttransfection, the medium was removed, and the cells were rinsed once with 0.5× Tris-buffered saline (1× Tris-buffered saline is 50 mM Tris-Cl, pH 7.8, 130 mM NaCl, 10 mM KCl, 5 mM MgCl2). Cell lysis buffer (25 mM Tris-phosphate, pH 7.8, 2 mM dithiothreitol, 2 mM 1,2-diamino-cyclohexane-N,N,N’,N’-tetraacetic acid, 10% glycerol, 1% Triton X-100) was added (50 μl/well), and the plates were incubated on ice for 15 min. When extracts for gel shift assays should be made from the nuclei after the luciferase assay, protease inhibitors (20 μg/ml each of leupeptin, chymostatin and aprotinin; 5 μg/ml each of pepstatin A and antipain; 1 mM phenylmethylsulfonyl fluoride) were included in the cell lysis buffer. The plates were then centrifuged, and the cytoplasmic extract was carefully removed, without disturbing the nuclei attached to the bottom of the well, and assayed for luciferase activity. The plates were put back on ice, and 30 μl of nuclear extraction buffer (100 mM Hepes, pH 7.9, 500 mM KCl, 25 mM MgCl2, 35% glycerol, 10 mM dithiothreitol, protease inhibitors) was added to each well. The washes of the wells were transferred to 1.5-ml tubes and cleared by centrifugation. Supernatants were diluted 5-fold with H2O and assayed for FREAC-2 or Gal4 DNA binding activities in gel shifts (24) with the probes GATCCAACTGTAACATCGGAGATC (FREAC-2) and GATCCGAGACTGTCCTCCGGAGATC (Gal4). Each construct was transfected in duplicate in at least three independent transfections.

Bacterial Protein Expression—Amino acids 146–408 of FREAC-2 was expressed as a histidine-tagged protein in E. coli M15[pREP4] using the vector pQE30 (Qiagen). Induced bacteria were lysed in 6 M guanidine-HCl, 0.1 M sodium phosphate, 10 mM Tris-Cl, 1 mM β-mercaptoethanol, pH 8.0, and mixed with nickel-NTA (Qiagen) resin. The resin was packed in a column and washed with 8 M urea, 0.1 M sodium phosphate, 10 mM Tris-Cl, 1 mM β-mercaptoethanol, pH 8.0 (Buffer B), and the bound protein was renatured with a linear gradient of Buffer B to 100 mM NaCl, 20 mM Tris-Cl, pH 7.4. 20% glycerol, 2 mM β-mercaptoethanol, 0.02% NaN3 using an Amersham Pharmacia Biotech FPLC.

Epitope (HA)-tagged human TBP (41) was expressed in E. coli BL21(DE3)[pLysS], and the soluble protein was immobilized on protein A-Sepharose beads (Amersham Pharmacia Biotech) that had first been saturated with the anti-HA monoclonal antibody 12CA5 (42).

Protein-Protein Interaction Assays—Coupled in vitro transcription-translation of general transcription factors and parts of FREAC-2 was performed with TNT-T7 (Promega) and [35S]methionine (Amersham Pharmacia Biotech). In vitro-translated general transcription factors were mixed with FREAC-2/nickel-NTA beads, and in vitro-translated FREAC-2 peptides were mixed with TBP-Sepharose beads in 50 μl of 20 mM Tris-Cl, pH 7.4, 100 mM NaCl, 5 mM MgCl2, 10% glycerol, 0.1% Nonidet P-40, 2 mM β-mercaptoethanol, 0.1 mM mg/ml bovine serum albumin. When nickel-NTA beads were used, 10 μM imidazole was included in the binding and washing buffers to eliminate nonspecific binding. To verify that retention of the labeled protein was due to true protein-protein interactions, rather than being mediated by protein-DNA interaction, the DNA binding domain of FREAC-2 peptides was mixed with TBP-Sepharose beads in 50 μl of 20 mM Tris-Cl, pH 7.4, 100 mM NaCl, 5 mM MgCl2, 10% glycerol, 0.1% Nonidet P-40, 2 mM β-mercaptoethanol, 0.1 mM mg/ml bovine serum albumin. When nickel-NTA beads were used, 10 μM imidazole was included in the binding and washing buffers to eliminate nonspecific binding. To verify that retention of the labeled protein was due to true protein-protein interactions, rather than being mediated by protein-DNA interaction, the DNA binding domain of FREAC-2 peptides was mixed with TBP-Sepharose beads in 50 μl of 20 mM Tris-Cl, pH 7.4, 100 mM NaCl, 5 mM MgCl2, 10% glycerol, 0.1% Nonidet P-40, 2 mM β-mercaptoethanol, 0.1 mM mg/ml bovine serum albumin. When nickel-NTA beads were used, 10 μM imidazole was included in the binding and washing buffers to eliminate nonspecific binding.

Control binding reactions were performed with nickel-NTA beads or protein A-Sepharose beads saturated with 12CA5 monoclonal antibody. The binding reactions were rocked gently for 1 h at 4°C, and the beads were collected by low speed centrifugation and washed four times with binding buffer. Bound proteins were detected by SDS-polyacrylamide gel electrophoresis and fluorography.

Given Fluorescent Proteins (GFP) Fusions—The FREAC-2 cDNA and parts thereof generated by polymerase chain reaction were cloned between the BamHI and XhoI sites in the GFP fusion vector pGFP-C1 (CLONTECH). Transfection of these constructs into COS-7 cells generated fusion proteins in which parts of FREAC-2 are fused to the C terminus of GFP. Cells were seeded on a coverslip in a tissue culture dish and transfected with liposomes as described above. Twenty-four hours posttransfection, the cells were rinsed in phosphate-buffered saline and mounted in a drop of phosphate-buffered saline on a microscope slide. The subcellular localization of the fusion proteins was visualized in vivo by fluorescence microscopy and photographed through a fluorescein isothiocyanate filter set.

RESULTS

Mapping of trans-Activation Domains in FREAC-2—To define which parts of the FREAC-2 protein are responsible for trans-activation of transcription, we generated a series of 5' and 3' deletions in the FREAC-2 cDNA (Fig. 1A). Expression constructs for these truncated FREAC-2 proteins were transfected into mammalian cell lines together with a luciferase reporter plasmid containing FREAC-2 binding sites upstream of a minimal promoter. Several cell lines derived from different tissues, including the lung cell lines NIH-H441 and WI38, were tested, and the activation properties of the FREAC-2 derivatives were found to be similar in all of the cell lines. We
therefore chose to use the human hepatoma cell line HepG2 and the monkey kidney cell line COS-7, because high and reproducible transfection efficiencies were obtained with these cells.

Deletion of the amino acids on the N-terminal side of the forkhead domain did not influence activation by FREAC-2, whereas deletions from the C-terminal end reduced its activation of the reporter (Fig. 1A). To establish whether the drop in activity reflects removal of activation domains or simply instability of the truncated proteins, we made extracts from the nuclei of transfected cells and used gel shift to assay for proteins with FREAC-2 DNA binding specificity. As shown in Fig. 1C, two of the truncated proteins (1–387 and 1–293) were partially degraded, and the luciferase activities produced in the corresponding transfections (Fig. 1A) are therefore unreliable. However, the rest of the proteins appear to be stable and to bind DNA normally. FREAC-2 proteins truncated at amino acid 241 or 189 inhibit luciferase expression from the reporter to less than 5% of the value obtained with cotransfection of the empty cloning vector. The mechanism behind this inhibition is not known, but the simplest explanation is that endogenous forkhead proteins with overlapping DNA binding specificities are capable of activating the reporter and are displaced by FREAC-2. Consistent with this explanation, the 4×FREAC-luc
reporter produces an activity severalfold higher than the corresponding construct without FREAC-2 sites, apoB-luc, and the truncated FREAC-2 proteins fail to inhibit the luciferase expression from apoB-luc (data not shown).

The 1–328 truncated protein activates with approximately one-half the potency of the full-length protein, and further deletions, moving the C-terminus of the protein from 328 to 241, resulted in a 100-fold drop in activity. We took this to indicate the importance of amino acid sequences between 241 and 328 for transcriptional activation, with a possible contribution from regions C-terminal of 328. To analyze this part of the protein further, we used in vitro mutagenesis to introduce six triple alanine substitutions in the part of the FREAC-2 protein C-terminal of the forkhead DNA binding domain. The restriction enzyme sites introduced by mutagenesis were also used to create internal deletions in the FREAC-2 cDNA (Fig. 1B). Surprisingly, deletions in the region between 241 and 328 did not lead to significant reductions of activity. In fact, deletion of everything C-terminal of the forkhead domain, except for the last 23 amino acids (Δ173–385), did not impair transcriptional activation by FREAC-2 in this assay (Fig. 1B). This result suggested that the most C-terminal 23 amino acids of FREAC-2 can act as an independent activation domain. To test whether this is the case or the context of the DNA binding forkhead domain is a prerequisite for activation, we fused segments of different length from the C-terminal half of FREAC-2 with the DNA binding domain of Gal4. When the ability of these fusion proteins to activate a reporter containing multiple Gal4 binding sites were analyzed, the most C-terminal 23 amino acids of FREAC-2 were again found to act as a potent and independent activation domain (Fig. 2A). This part of the FREAC-2 protein will be referred to as activation domain 1 (AD1).

The results of transfections with deletion mutants and Gal4 fusions suggested that FREAC-2 contains at least two independent and functionally redundant activation domains. More specifically, the 100-fold difference between 1–241 and 1–328 (Fig. 1A) indicated the presence of activating sequences in the central part of the protein (from now on referred to as activation domain 2 (AD2)), whereas the activities of Δ173–385 and G4/386–408 (Figs. 1B and 2A) define AD1 in the C terminus. As the results of the internal deletions showed (Fig. 1B), it was impossible to investigate AD2 with AD1 present; even deletion of more than half the FREAC-2 protein (Δ173–385) did not reduce activity as long as AD1 was left intact (Fig. 1B). Analysis of AD2 was further complicated by the fact that a precise deletion of AD1 from FREAC-2 (1–387) yielded an unstable protein (Fig. 1C). To identify the activating regions that constitute AD2, we therefore fused sequences from the internal part of FREAC-2 to the DNA binding domain of Gal4. The 212 amino acids from the C-terminal end of the forkhead domain to the N-terminal end of AD1 (amino acids 174–388) were divided into four subregions of 51–56 amino acids each (Fig. 2B). When analyzed by cotransfection with the Gal4-luciferase reporter, the first (174–225), second (225–278), and fourth (332–388) subregions contributed to activation (Fig. 2B). Individually, none of the subregions produced any substantial activation, but together (G4/174–388), they acted in a synergistic manner and activated the Gal4 reporter 150-fold (Fig. 2B). Judged from gel shifts with extracts from transfected cells, all the FREAC-2/Gal4 fusion proteins, except G4/174–408, appear to be stable and to bind DNA normally (Fig. 2C). In the case of G4/174–408 the major part of the Gal4 DNA binding activity consists of dimers of just the Gal4 DNA binding domain, which arise from proteolyis, and very little of the full-length dimer is seen. Because the truncated Gal4 dimers bind DNA and compete with the full-length protein for the binding sites but do not activate transcription, this may account for the lower activity of G4/174–408 compared with G4/174–388.

**FREAC-2 Interacts with General Transcription Factors**—The parts of FREAC-2 responsible for activation of transcription thus appear to be located in the region C-terminal of the forkhead domain. To investigate whether this part of the protein interacts through direct protein-protein binding with components of the transcriptional preinitiation complex, we used a pull-down assay. Amino acids 146–408 of FREAC-2 were expressed in bacteria as a His-tagged protein and immobilized on a nickel-agarose affinity matrix. Human general transcription factors, synthesized and 35S-labeled by in vitro translation, were assayed for specific binding to the FREAC-2 coated beads (Fig. 3). TFIIA (the 55-kDa precursor of the α- and β-subunits), TFIIIB, TBP, TFIIEα, TFIIEβ, TAFii32, and TAFii55 were tested. Two proteins, TBP and TFIIB, were found to interact specifically with FREAC-2, whereas none bound to the control beads consisting of nickel-agarose (Fig. 3).

The domains of the general transcription factors responsible for the interaction with FREAC-2 were mapped by repeating the FREAC-2 binding assay with truncated forms of TBP and TFIIB (Fig. 4). The conserved TBP core domain in the C-terminal half of the protein contains two imperfect (38% identity) 61 amino acid repeats (41) and is folded in a pseudosymmetrical saddle shape (43–46). Proteins from which part (TBP1–267) or all (TBP1–236) of the C-terminal repeat had been deleted bound FREAC-2 with undiminished affinity, but when the deletion was extended to remove part of the N-terminal repeat (TBP1–180), the interaction was lost. Thus, the second half of the first repeat of core TBP (amino acids 181–236) appears to be the target for FREAC-2 binding. Although not similar in either amino acid sequence or threedimensional structure, the organization of the TFIIB protein resembles that of TBP in having a unique N-terminal part followed by a C-terminal core domain made up of two imperfect (20% identity) 84-amino acid repeats (40). The two repeats in the core of TFIIB folds into two separate domains with similar three-dimensional structure (47). Deletion of more than half of the last of the two repeats in TFIIB (TFIIB1–265 and TFIIB1–251) did not impede binding to FREAC-2, whereas deletion into the first repeat did (TFIIB1–173). As with TBP, the part of TFIIB that interacts with FREAC-2 is within the core of the protein, or more precisely, between amino acids 174 and 251; i.e. in the C-terminal end of the N-terminal repeat or the N-terminal end of the C-terminal repeat.

Given the localization of activation domains in the FREAC-2 protein, we wanted to determine whether the amino acid sequences interacting with the general transcription factors coincided with any of the regions responsible for activation. To do this, we expressed epitope-tagged human TBP in bacteria and bound to protein A-Sepharose with a monoclonal antibody. Various parts of FREAC-2 were synthesized by in vitro translation and tested for binding to the immobilized TBP protein (Fig. 5). None of the regions that were identified as activators of transcription interacted with TBP. Instead, the TBP binding region of FREAC-2 was localized to amino acids 146–166, i.e. to the most C-terminal end of the forkhead domain. Because this stretch of amino acids is part of the DNA binding domain of FREAC-2—although unable to bind DNA by itself—we repeated the binding assay in the presence of a high concentration of DNase I to ensure that the retention of FREAC-2 resulted from a bona fide protein-protein interaction and was not mediated by DNA-protein interactions. Incubation with nuclease did not influence the binding between FREAC-2 and TBP,
which confirms that DNA does not mediate this association (Fig. 5).

In cotransfection experiments, overexpression of TFIIB potentiates the activation by FREAC-2 approximately 10-fold (Fig. 6). This TFIIB effect depends on the integrity of the activation domains and was lost when FREAC-2 was truncated at amino acid 189 (Fig. 6).

The Forkhead Domain of FREAC-2 Contains a Bipartite Nuclear Localization Signal—As a DNA-binding protein and transcription factor, FREAC-2 must enter the nucleus to perform its activity. In order to identify amino acid sequences in FREAC-2 that may function as a nuclear localization signal (NLS), we fused the FREAC-2 cDNA, and fragments thereof, with the coding sequence of GFP (Fig. 7B) and analyzed the subcellular localization of GFP fusion proteins in transfected COS-7 cells. While GFP alone was distributed throughout the cell, FREAC-2 fused to GFP was found exclusively in the nucleus (Fig. 7B); a result that confirms that FREAC-2 is indeed a nuclear protein. The minimal part of FREAC-2 that is entirely localized to the nucleus comprises amino acids 64–166, i.e. exactly the forkhead domain. When deletions were made in the forkhead domain from either the C or N terminus, the resulting fusion proteins were still preferentially nuclear, but with a significant amount of the proteins present in the cytoplasm. Sixteen amino acids from the C terminus of the forkhead domain (amino acids 151–166) gave a predominantly nuclear localization to the fusion protein, but a similar result was obtained when the forkhead domain with these same 16 amino
acids deleted was fused to GFP (amino acids 58–152). In contrast, 38 amino acids (115–152) from the central part of the forkhead domain did not contribute to nuclear localization; the distribution of this fusion protein was indistinguishable from that of GFP alone. The forkhead domain thus appears to contain sequences in its N terminus, as well as its C terminus, that aid in translocation of the protein from the cytoplasm to the nucleus. Both regions are necessary for a complete nuclear localization.

**DISCUSSION**

FREAC-2 activates transcription through two independent activation domains, AD1 and AD2. AD1 is well defined as the most C-terminal 23 amino acids of the protein. Also, when fused to the heterologous DNA binding domain of Gal4, these 23 amino acids activate expression from a reporter construct, which confirms their function as a *bona fide* activation domain.

The amino acid sequence of AD1 is homologous to the C terminus of FREAC-1 (Fig. 7B) (24). This is the only region, apart from the forkhead domain, where FREAC-1 and FREAC-2 have any higher degree of similarity. As might be expected, the C terminus of FREAC-1 has activating properties similar to AD1 of FREAC-2. A transcriptional activation domain from the extreme C terminus of another forkhead protein, HNF3β, has been described (49). However, there is no similarity in amino acid sequence between the activation domains of HNF3β and AD1 in FREAC-2.

AD2 is spread out over approximately 200 amino acids and thus spans half of the FREAC-2 protein. In contrast to AD1, it consists of noncontiguous ranges of amino acids. Activation by AD2 may depend on the tertiary structure of the protein to bring amino acid motifs from distant locations together into a functional structure. The reasons for treating it as a single activation domain are the following observations: when the 215 amino acids containing the AD2 activity were divided into four subregions, none of these subregions was able to activate on its own. But when combined, three subregions (numbers 1, 2, and 4) enhanced each other’s activation in a synergistic rather than additive manner. If the mechanism by which AD2 activates transcription involves binding of a target protein, these results are compatible with interaction with a single target and contribution to the binding by all the three activating subregions. The amino acid sequence of AD2 shows no obvious homology to other known proteins; it is rich in serines, glycines, histidines, and prolines; it contains a stretch of poly-histidine; and it has comparatively few charged amino acids.

No synergy appears to arise from the presence of both AD1 and AD2 in FREAC-2. When analyzed by deletions in the FREAC-2 protein as well as by fusion to Gal4, no significant changes in reporter gene activity were observed as a result of removal of either AD1 or AD2, as long as the other was left.

---

2 Mahlapuu, M., Pelto-Huikko, M., Aitola, M., Enerbäck, S., and Carlsson, P. (1998) Dev. Biol., in press.
intact. With the assay and reporter used, the two activation domains thus appear to be functionally redundant, which of course does not exclude the possibility that they may serve distinct purposes in other contexts.

Several lines of evidence support the idea that transcriptional activation involves recruitment of the transcriptional preinitiation machinery to the promoter DNA, and activation domains of transcription factors have been shown to interact specifically with components of the general transcription complex (reviewed in Ref. 50). We were therefore interested to see whether the FREAC-2 activation domains interacted with any of the general transcription factors involved in PolII transcription. A panel of general transcription factors were tested for

**FIG. 5.** The part of FREAC-2 that interacts with TBP is the most C-terminal 21 amino acids of the forkhead domain (146–166). Parts of FREAC-2 were synthesized and 35S labeled by in vitro translation and tested for binding to immobilized TBP. Loaded in the input lane is 5% of the protein used in the binding reaction. The lane labeled TBP beads shows epitope-tagged human TBP expressed in E. coli and immobilized on protein A-Sepharose with a monoclonal antibody, and the lane labeled control beads shows protein A-Sepharose with just the monoclonal antibody. +DNase I indicates that DNase I was included in the binding reaction to eliminate possible contributions from protein-DNA interactions to the retention of the labeled protein.

**FIG. 6.** FREAC-2 dependent activation of transcription by TFIIB. COS-7 cells were transfected with the 4×FREAC-luc reporter and the indicated FREAC-2 constructs. Shaded bars represent the ratio between luciferase activities produced by cotransfections with and without a TFIIB expression plasmid. The averages of four transfection experiments are shown, and error bars represent the S.E.

**FIG. 7.** The forkhead domain of FREAC-2 contains a bipartite nuclear localization signal. A, fluorescence micrographs of COS-7 cells transfected with plasmids expressing GFP and fusions between GFP and parts of FREAC-2. Numbers refer to amino acids of FREAC-2 fused to GFP. Green fluorescence reveals the localization of the GFP moiety, and areas of intense fluorescence appear white or yellowish due to overexposure of the film. B, schematic representation of the parts of FREAC-2 that were fused to GFP. The hatched box corresponds to the forkhead domain.
interaction with the C-terminal part of FREAC-2, containing the activation domains AD1 and AD2, in a protein-protein binding assay. Two factors, TBP and TFIIB, were found to bind specifically to FREAC-2. These two components of the general transcription complex have been found to interact with the activation domains of several other transcription factors, e.g. NFκB, C/EBPα, and VP16 (51–54).

Transcriptional activation by FREAC-2 is potentiated by overexpression of TFIIB, and this effect depends on the C-terminal part of FREAC-2 containing the activation domains. One component in the transcription regulating activity of FREAC-2 thus appears to be activation through recruitment of TFIIB by the C-terminal part of the protein.

However, when we mapped the TBP interaction on the FREAC-2 protein (Fig. 8), the amino acids responsible for binding to TBP were found to be nonoverlapping with the activation domains. Instead, TBP was found to interact with 21 amino acids in the C terminus of the forkhead domain. The function of this interaction remains obscure, but it does not appear to be required for activation by AD1 or AD2 because both retain their activities when fused to the heterologous DNA binding domain of Gal4. The region of FREAC-2 that binds TBP is well conserved within the forkhead family (19), and it is therefore possible that the ability to interact with TBP is a general characteristic of forkhead proteins.

The target sites for FREAC-2 on both TBP and TFIIB are located in the conserved core domains of the proteins. The core of TBP embraces the TATA box DNA like a saddle, where the sequence repeats correspond to the two sides of the saddle (43–46). The two repeats of the TFIIB core are folded into two separate globular domains of similar structure, connected by a short linker sequence (47). The cleft between the two domains of this dumbbell-shaped protein binds to the stirrup of the C-terminal half of the TBP saddle, and the two TFIIB domains are arranged more or less along the axis of the DNA (47).

Because FREAC-2 contacts the N-terminal repeat of the TBP core, it will bind to the half of the saddle that faces away from TFIIB. On TFIIB, the site of FREAC-2 interaction is close to the cleft where the TBP stirrup binds and could be on either the C-terminal or the N-terminal domain. Consequently, the target sites for FREAC-2 on TBP and TFIIB are not juxtaposed when these proteins are bound to the TATA box, and if the interactions are to take place simultaneously, the parts of FREAC-2 involved in binding have to be well separated.

NLSs are amino acid motifs that will cause a protein to be translocated to the nucleus through docking with NLS receptors (55). Usually, an NLS is a short stretch rich in basic amino acids with a bias toward lysines over arginines (56, 57). However, many deviations from this general rule have been described, and a classification with different types of NLS motifs has been suggested (58). In DNA-binding proteins, the NLS is almost invariably found within, or in immediate proximity to, the DNA binding domain (59). From an evolutionary point of view, this arrangement makes sense. Nuclear localization is a prerequisite for DNA binding, and making the NLS an intrinsic property of the DNA binding domain will ensure that this function is always at hand when the DNA binding motif is used in new contexts, e.g. through exon shuffling. Hence, it is not surprising that the NLS of FREAC-2 is found within the forkhead domain. From inspection of the primary structure, the stretch of basic amino acids in the C-terminal end of the forkhead domain appears to be the best candidate, and when these 16 amino acids, GSFRRRRPFRRKCCQA (151–166), were fused to GFP, the localization of the fusion protein was indeed preferentially nuclear. However, a significant amount of the protein was still cytoplasmic, in contrast to what we saw with the entire forkhead domain, in which all of the GFP fusion protein was nuclear. Furthermore, when the basic stretch of 16 amino acids was removed from the forkhead domain, the resulting GFP fusion protein was still preferentially nuclear, although a considerable cytoplasmic leakage confirmed the importance of the basic stretch for complete nuclear localization. The fact that 38 amino acids from the middle of the forkhead domain have no NLS activity shows that the NLS is bipartite and requires sequences from both ends of the forkhead domain. Because both the N-terminal and the C-terminal sequences have substantial NLS activity when isolated, they are perhaps better described as two independent NLSs, neither of which is sufficient for complete nuclear localization. Given the high degree of sequence conservation within the forkhead domain, it is not unlikely that the NLS is organized in a similar manner throughout this protein family, although examples from the steroid receptor family show that the location of the NLS can differ considerably among proteins with the same general structure of their DNA binding domains (48). That a requirement for sequences from both ends of the forkhead domain for nuclear localization is not unique for FREAC-2 is shown by studies of HNF3β, in which an intact forkhead domain has been reported to be necessary for nuclear localization (26).

Acknowledgments—We thank Drs. R. D. Roeder, Z. F. Burton, D. Reinberg, and R. Tjian for clones of general transcription factors. Thanks are also due to Kerstin Dahlenberg and Lennart Asp for technical assistance.

REFERENCES

1. Goffeau, A. et al. (1997) Nature 387, (suppl.) 1–105
2. Kaufmann, E., and Knochel, W. (1996) Mech. Dev. 57, 3–20
3. Ang, S. L., and Rossant, J. (1994) Cell 78, 561–574
4. Weinstein, D. C., Buiz Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M., and Darnell, J., Jr. (1994) Cell 78, 557–588
5. Xuan, S., Baptista, C. A., Balas, G., Tao, W., Soares, V. C., and Lai, E. (1995) Neuron 14, 1141–1152
6. Hatini, V., Huh, S. O., Herzlinger, D., Soares, V. C., and Lai, E. (1996) Genes Dev. 10, 1467–1478
7. Nehls, M., Pfeifer, D., Schorpp, M., Hedrich, H., and Boehm, T. (1994) Nature 367, 105–107
8. Weigel, D., Jurgens, G., Kuttenn, F., Seifert, E., and Jackle, H. (1989) Cell 57, 645–658
9. Haake, U., Kaufmann, K., Hartmann, C., Jurgens, G., Knochel, W., and Jackle, H. (1995) EMBO J. 14, 5306–5317
10. Grossniklaus, U., Pearson, R. K., and Gehring, W. J. (1992) Genes Dev. 6, 1090–1051
11. Miller, L. M., Gallegos, M. E., Morisseau, B. A., and Kim, S. K. (1993) Genes Dev. 7, 933–947
12. Winnier, G. E., Hargett, L., and Hogan, B. L. (1997) Genes Dev. 11, 928–940
13. Labosky, P. A., Winnier, G. E., Jelton, T. L., Hargett, L., Ryan, A. K., Rosenfeld, M. G., Parlow, A. F., and Hogan, B. L. (1997) Development 124, 1263–1274
14. Dou, X., Ye, X., Stewart, C., Lai, E., and Li, S. C. (1997) Neuron 18, 539–551
15. Li, C., Lai, C. F., Sigman, D. S., and Gaylor, R. B. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 7739–7743
16. Sutton, J., Costa, R., Klug, M., Field, L., Xu, D., Largaespada, D. A., Fletcher, C. F., Jenkins, N. A., Copeland, N. G., Riemer, M., and Hromas, R. (1996) J. Biol. Chem. 271, 23126–23133
17. Schlake, T., Schorpp, M., Nehls, M., and Boehm, T. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 3842–3847
FREAC-2 Contains Functionally Redundant Activation Domains