

The Dimerization/Repression Domain of RFX1 is Related to a Conserved Region of its Yeast Homologues Crt1 and Sak1: A New Function for an Ancient Motif

Yael Katan-Khaykovich, Ivo Spiegel and Yosef Shaul*

Department of Molecular Genetics, The Weizmann Institute of Science, Rehovot 76100, Israel

The RFX protein family includes members from yeast to humans, which function in various biological systems, and share a DNA-binding domain and a conserved C-terminal region. In the human transcription regulator RFX1, the conserved C terminus is an independent functional domain, which mediates dimerization and transcriptional repression. This dimerization domain has a unique ability to mediate the formation of two alternative homodimeric DNA-protein complexes, the upper of which has been linked to repression. Here, we localize the complex formation capacity to several different RFX1 C-terminal subregions, each of which can function independently to generate the upper complex and repress transcription, thus correlating complex formation with repression. To gain an evolutionary perspective, we have examined whether the different properties of the RFX1 C terminus exist in the two yeast RFX proteins, which are involved in signaling pathways. Replacement of the RFX1 C terminus with those of Sak1 and Crt1, its orthologues from *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, respectively, and analysis of fusions with the Gal4 DNA-binding domain, revealed that the ability to generate the two alternative complexes is conserved in the RFX family, from *S. cerevisiae* to man. While sharing this unique biochemical property, the three C termini differed from each other in their ability to mediate dimerization and transcriptional repression. In both functions, RFX1, Sak1, and Crt1 showed high capacity, moderate capacity, and no capacity, respectively. This comparative analysis of the RFX proteins, representing different evolutionary stages, suggests a gradual development of the conserved C terminus, from the appearance of the ancestral motif (Crt1), to the later acquisition of the dimerization/repression functions (Sak1), and finally to the enhancement of these functions to generate a domain mediating highly stable protein-protein interactions and potent transcriptional repression (RFX1).

© 1999 Academic Press

Keywords: dimerization; molecular evolution; RFX family; transcriptional repression; EMSA

*Corresponding author

Introduction

Typical transcription regulators are composed of several functional domains, which mediate binding to DNA, oligomerization, and transcriptional activation or repression. Such proteins often generate homo- or heterodimers *via* their dimerization domains. Well-characterized dimerization domains are the leucine zipper motif, found, for example, in Jun, Fos, and GCN4, and the helix-loop-helix motif

Abbreviations used: DBD, DNA-binding domain; DOC, sodium deoxycholate; EDD, extended dimerization domain; GCNF, germ cell nuclear factor; HBV, hepatitis B virus; NP-40, Nonidet P-40; wt, wild-type.

E-mail address of the corresponding author: lvshaul@weizmann.weizmann.ac.il

of MyoD, E12, E47, and others (Jones, 1990; Baxevanis & Vinson, 1993). Other proteins, such as Myc and Max, contain both of these motifs. In most dimeric transcription factors, the dimerization domain is found adjacent to the DNA-binding domain (DBD), and generation of a protein dimer is essential for DNA-binding (Jones, 1990; Baxevanis & Vinson, 1993). Two protein families that are exceptions to this rule are the well-characterized nuclear receptors (Fawell *et al.*, 1990; Mangelsdorf & Evans, 1995; Mangelsdorf *et al.*, 1995; Perlmann *et al.*, 1996) and the novel RFX proteins (Reith *et al.*, 1990; Emery *et al.*, 1996a), which contain dimerization domains that are structurally and functionally independent of the DBDs.

The RFX protein family is highly conserved in evolution and includes members from yeast, *Caenorhabditis elegans*, mice, and humans (see Figure 1) (Emery *et al.*, 1996a). RFX1, RFX2, and RFX3, identified in humans and in mice, form stable homodimers or heterodimers with each other (Reith *et al.*, 1989, 1990, 1994). RFX1, the most ubiquitously expressed of these proteins, has been implicated in transcriptional regulation (Reith *et al.*, 1990; Siegrist *et al.*, 1993). The *Schizosaccharomyces pombe* RFX protein Sak1 is a positive mediator of the exit from the mitotic cell cycle (Wu & McLeod, 1995). Crt1, the recently identified RFX protein of *Saccharomyces cerevisiae*, is a transcriptional repressor, involved in the response to replication block and DNA damage (Huang *et al.*, 1998). The RFX proteins share several conserved regions, including a DNA-binding domain and the C-terminal B, C, and D regions, all of which are unique to the RFX family (Reith *et al.*, 1990, 1994; Emery *et al.*, 1996a,b). Region D of RFX1 was initially characterized as the dimerization domain (Reith *et al.*, 1990). Recently, we have demonstrated that efficient dimerization in cell extracts requires the B-C region as well, implying that RFX1 possesses a split dimerization domain (Katan-Khaykovich & Shaul, 1998). The complex structure of this domain, which we have termed Extended Dimerization Domain (EDD), suggested that it may serve for mediating different types of interactions. In accordance with this view, we have detected a novel low-mobility DNA-protein complex, the formation of which was mediated by the RFX1 EDD (Katan-Khaykovich & Shaul, 1998). RFX1 binds to several homologous DNA sites, which are either palindromic, such as the EP element found in viral enhancers (Ostapchuk *et al.*, 1986, 1989; Ben-Levy *et al.*, 1989; Dikstein *et al.*, 1990, Zhang *et al.*, 1990), or non-palindromic, containing only one EP-homologous half-site, such as the X box of MHC class II promoters (Kouskoff *et al.*, 1991; Siegrist *et al.*, 1993). All previously characterized homo- and heterodimeric complexes of RFX1 are similarly generated with both palindromic and non-palindromic DNA sites (Kouskoff *et al.*, 1991; Reith *et al.*, 1994; David *et al.*, 1995). By contrast, the novel up-shifted complex was detected only with palindromic ones. Interestingly, while a full deletion of the EDD abolished the up-

shifted complex, partial deletions increased its relative abundance. Based on our observations, we proposed the existence of two alternative complexes, which both contain RFX1 homodimers, yet differ in the nature of the intersubunit interaction, mediated by the EDD. Since other RFX family members, such as Sak1 and Crt1, contain sequences homologous to the B-C-D region of RFX1 or parts of it, dimerization and the unique ability to form alternative complexes *via* the conserved region may be common features of the RFX proteins. The properties of the conserved region in proteins other than RFX1, however, have not been tested.

RFX1 has been implicated in transcriptional regulation of viral and cellular genes. Its EP (EF-C) binding site in the hepatitis B virus (HBV) and polyomavirus enhancers (Ostapchuk *et al.*, 1989; Dikstein *et al.*, 1990; Bolwig & Hearing, 1991; Siegrist *et al.*, 1993; David *et al.*, 1995), as well as binding sites in the promoters of MHC class II (X box) (Tsang *et al.*, 1988, 1990; Kouskoff *et al.*, 1991), ribosomal protein rpL30 (Safrany & Perry, 1993, 1995), and interleukin-5 receptor $\alpha 5$ (Iwama *et al.*, 1999) genes, are positively acting regulatory elements in their natural context. However, multimers of the HBV EP site or a homologous site from the *c-myc* gene have no stimulatory effect on their own, demonstrating that the activity of EP is DNA context-dependent; moreover, these multimerized binding sites can silence the transcriptional activity of different enhancers (Weisinger *et al.*, 1988; Dikstein *et al.*, 1990; Reinhold *et al.*, 1995; Blake *et al.*, 1996). A negative effect of the EP site was observed when the p53 response element of the *mdm2* gene was inserted upstream of the HBV enhancer, resulting in the conversion of p53 from an activator to a repressor, dependent on the presence of the EP site (Ori *et al.*, 1998). An RFX1 binding site in the proliferating cell nuclear antigen promoter was recently implicated in transcriptional down-regulation (Labrie *et al.*, 1995; Lee *et al.*, 1998; Liu *et al.*, 1999). RFX1 was shown to stimulate transcription from the HBV enhancer, the HBV NRE γ site, and the MHC promoter X box, using overexpression and introduction of antisense DNA or RNA (Reith *et al.*, 1990; Siegrist *et al.*, 1993; Siegrist & Mach, 1993; Buckwold *et al.*, 1997). RFX1 can also interact with the c-Abl tyrosine kinase and stimulate the c-Abl kinase activity (Dikstein *et al.*, 1992, 1996; Agami & Shaul, 1998). Despite all these observations, the exact mechanism by which RFX1 regulates transcription remains unclear.

Previously we have demonstrated that RFX1 possesses functionally independent activation and repression domains, which can counteract each other's effect (Katan *et al.*, 1997). The fact that the repression domain overlaps the EDD suggested a possible link between transcriptional repression and formation of the novel up-shifted complex (Katan-Khaykovich & Shaul, 1998). By dissecting the RFX1 EDD, we now show a correlation between these two functions, which are both

mediated by several different subregions within the EDD. Importantly, the conserved C termini of Sak1 and Crt1 also mediated the formation of the up-shifted complex, indicating of structural similarity to the RFX1 EDD. By contrast, the C termini of RFX1, Sak1, and Crt1 differed from each other in their ability to mediate dimerization and transcriptional repression, showing high, moderate, and no capacity, respectively. This analysis suggests that an ancestral motif of unknown function in Crt1 has evolved into the complex dimerization/repression domain of Sak1 and RFX1.

Results

Dimerization properties of RFX1, Sak1, and Crt1

The homology between the C-terminal part of RFX1, which constitutes the EDD, and other RFX proteins (Figure 1) suggested that the dimerization function of this region may be conserved in the RFX family. In order to examine this, RFX1-derived chimeric proteins were generated, in which the B-C-D region of RFX1 was replaced with the conserved C terminus of its *S. pombe* homologue Sak1 (RFX1-Sak1) or its *S. cerevisiae* homologue Crt1 (RFX1-Crt1). This approach allowed us to compare the dimerization properties of the different C termini alone, in a similar context; moreover, the generation of chimeras seemed particularly suitable for studying the RFX family, since the RFX1 EDD is known to function independently of the DBD (Reith *et al.*, 1990). The dimerization status of these chimeric proteins was assayed by gel shift analysis of extracts from transfected cells, as previously described (Katan-Khaykovich & Shaul, 1998), with two homologous RFX1 binding sites: the palindromic HBV EP element and the non-palindromic X box half-site. The palindromic nature of the EP element allows the binding of two RFX1-derived

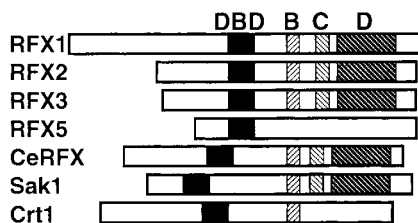


Figure 1. The RFX family. RFX proteins from man and mouse (RFX1-5), *C. elegans* (CeRFX), *S. pombe* (Sak1), and *S. cerevisiae* (Crt1) are shown, according to Emery *et al.* (1996a). The RFX1-5 (Reith *et al.*, 1989, 1990, 1994; Steimle *et al.*, 1995), sak1 (Wu & McLeod, 1995), and crt1 (Huang *et al.*, 1998) genes have been cloned. The CeRFX gene was identified by a data library search (Emery *et al.*, 1996a). The DBD and regions B, C, and D are conserved in the RFX family. The Crt1 sequence shows clear homology to the DBD and region B, and only limited homology to parts of regions C and D.

molecules to each probe, regardless of their dimerization status. With this probe, the two chimeric proteins, as well as the wild-type (wt) RFX1 and the Δ BCD RFX1 mutant, which lacks the B-C-D region, all generated dimeric complexes (Figure 2, lanes 3, 5, 7, and 9). The X box half-site, to which only one RFX1 molecule can directly bind, was used to discriminate between stably linked dimers and free monomers. As previously reported (Katan-Khaykovich & Shaul, 1998), the wt RFX1 formed only dimeric complexes with the X box (lane 4), while the dimerization-deficient Δ BCD formed only monomeric complexes (lane 10). The RFX1-Sak1 chimera generated both dimeric and monomeric complexes (lane 6), and the RFX1-Crt1 chimera formed only monomeric complexes (lane 8). Therefore, the C termini of RFX1 and Sak1 mediate the formation of highly and moderately stable dimers, respectively, while that of Crt1 shows no dimerization capacity. The dimerization property of the RFX C terminus thus shows a hierarchy in Crt1, Sak1, and RFX1, ranging from none to maximal capacity.

Formation and stabilization of an up-shifted RFX1 complex

Since the C termini of the three RFX proteins showed sequence homology on one hand, but differential dimerization properties on the other hand, we attempted to determine whether these conserved regions share structural or biochemical properties. One unique property of the RFX1 C terminus is its ability to mediate the formation of an up-shifted DNA-protein complex. In gel shift anal-

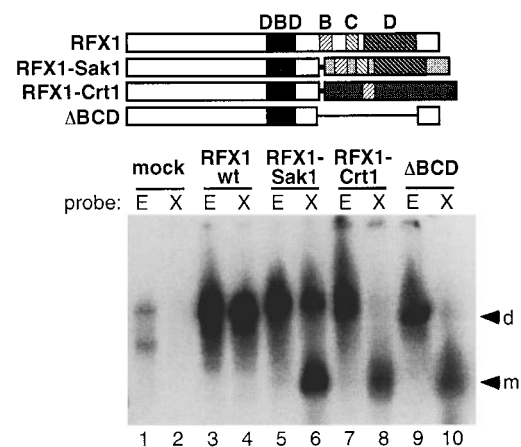


Figure 2. Dimerization properties of the RFX C terminus. 293T cells were transfected with expression plasmids of the wt RFX1, Δ BCD RFX1 mutant, and RFX1 chimeras substituted with the C termini of Sak1 and Crt1, or mock transfected (mock). Cells were extracted in the absence of detergents and assayed by gel shift with the palindromic EP (E) and non-palindromic X box (X) probes. Dimeric (d) and monomeric (m) complexes are indicated. The free probe was allowed to run out of the gel.

ysis of cell extracts, RFX1 generates two major DNA-protein complexes: the well-characterized complex a, containing RFX1 homodimers (Herrero-Sanchez *et al.*, 1992; Siegrist *et al.*, 1993; Reith *et al.*, 1994), and the recently identified up-shifted complex a* (Katan-Khaykovich & Shaul, 1998). While complex a is efficiently generated with both palindromic and non-palindromic probes, complex a* is observed only with the palindromic EP site, but not with X box half-site, and its formation is dependent on the EDD. To further investigate the EDD-mediated formation of the up-shifted complex, we wished to define conditions that may induce the preferential formation of one of these

complexes. Since the extraction buffer used in our initial identification of complex a* contained Triton X-100 (Katan-Khaykovich & Shaul, 1998), while the detergent-independent extraction protocol used for the experiment illustrated by Figure 2 did not generate the up-shifted complex, we examined the effect of Triton X-100 on complex formation (Figure 3(a) and (b)). Complex a was generated using whole-cell extracts prepared in either the absence or the presence of Triton X-100, whereas the up-shifted complex a* was seen with the EP probe only in the presence of Triton X-100 (Figure 3(a), lanes 1-4). A similar result was obtained with the overexpressed wt RFX1 (lanes 5-

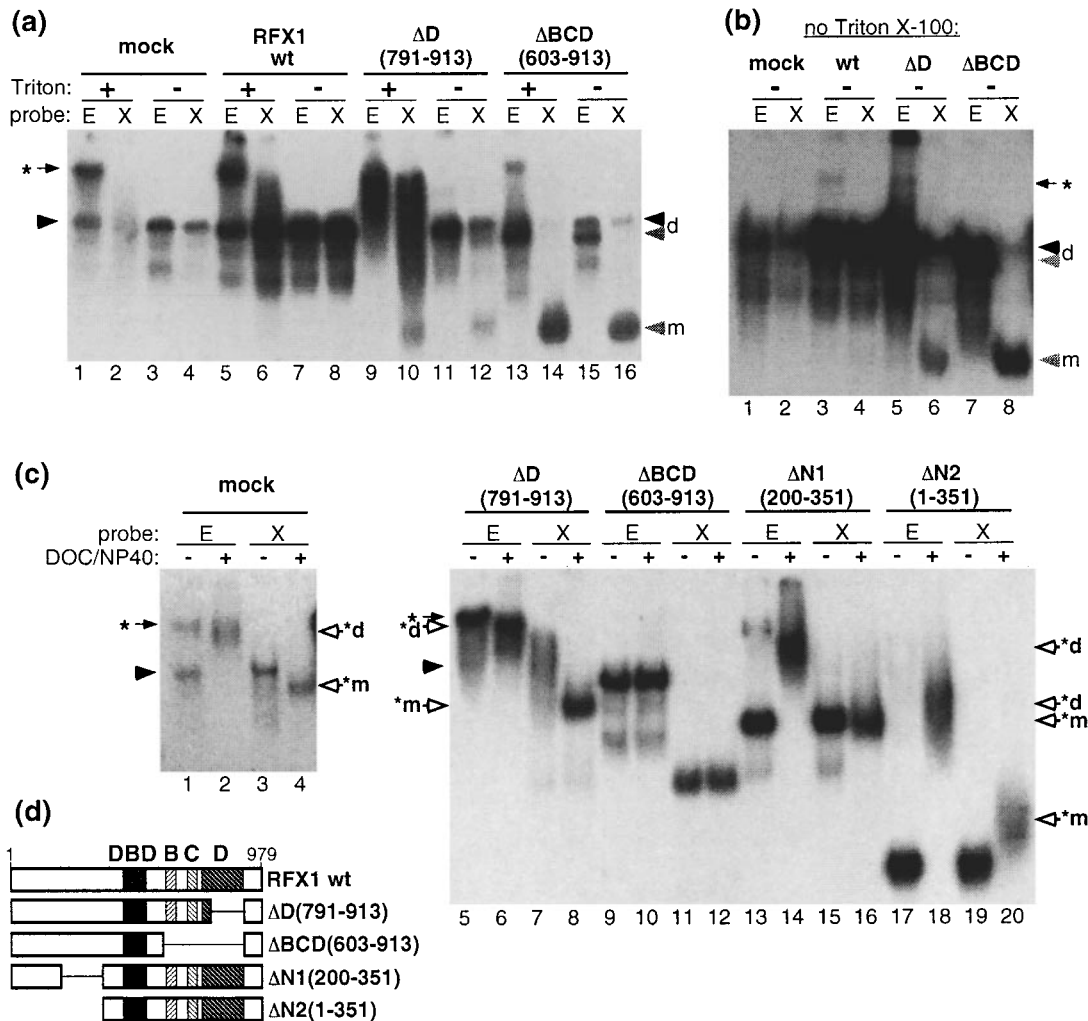


Figure 3. The effect of detergents on EP complex formation. HepSK1 cells were transfected with expression plasmids of the RFX1 derivatives shown in (d) or mock transfected (mock). Cellular extracts were analyzed by gel shift with the EP (E) and X box (X) probes. (a) and (b) Extraction was performed in the absence (-) or presence (+) of Triton X-100. The position of complex a* is indicated by an asterisk (*). Black and gray arrowheads indicate the position of complex a generated by the wt RFX1 and its deletion mutants, respectively. Complex a* seen in lane 13 is formed by the endogenous RFX1. d, dimer; m, monomer. The free probe was allowed to run out of the gel. (c) Cells were extracted in the presence of Triton X-100. A combination of DOC (0.4%) and NP-40 (1.2%) was added to the binding reaction where indicated by +. The positions of the endogenous complexes a and a* formed in the absence of DOC/NP-40 are marked as in (a). Complexes marked by open arrows appear to be dimeric (*d) and monomeric (*m) forms of complex a*, formed in the presence of DOC/NP-40. Lanes 1-4 and 5-20 represent different exposures of the same gel. (d) The structure of RFX1 and its deletion mutants lacking the indicated amino acids.

8). However, upon a long exposure of the gels, a faint a^* band could be detected with the EP probe even in the absence of detergents (Figure 3(b), lane 3).

RFX1 mutants bearing partial deletions within the EDD generate complex a^* with the EP probe and a smeary complex with the X box probe, which appears to be an unstable monomeric form of complex a^* , but do not form complex a (Katan-Khaykovich & Shaul, 1998). In the presence of Triton X-100, the ΔD mutant, deleted in region D, generated the expected complexes (Figure 3(a), lanes 9 and 10). By contrast, in the absence of detergents, ΔD formed mainly complex a , a higher dimeric form with the EP probe and a lower monomeric form with the X box (lanes 11 and 12). A highly exposed gel revealed the existence of a relatively faint a^* band even in the absence of detergent (Figure 3(b), lane 5). Deletion of the whole B-C-D region, generating the ΔBCD mutant, abolished complex a^* formation, as previously shown (Katan-Khaykovich & Shaul, 1998). The ΔBCD mutant was not affected by the extraction conditions and generated only a dimer and a monomer of complex a with the EP and X box probes, respectively, under both conditions (Figure 3(a), lanes 13-16, and (b), lanes 7 and 8). Collectively, these results demonstrate a major effect of Triton X-100 on complex a^* formation and suggest that in its absence most a^* complexes are converted into the a -form. However, Triton X-100 is not absolutely required for complex a^* formation, since in the absence of detergent a small fraction of the RFX1 complexes were of the a^* form.

To determine the role of Triton X-100, cells were extracted in the absence of detergent, and only after centrifugation and removal of the insoluble material, the supernatants were supplemented with Triton X-100. The addition of Triton X-100 to the supernatants allowed the detection of complex a^* in some experiments (data not shown), indicating that these extracts contained the components required for complex a^* formation. Therefore, the effect of Triton X-100 appears to be not in extracting an a^* -essential component but, more likely, in stabilizing complex a^* . This stabilization may be especially important during the extraction, since the later addition of Triton X-100 to the extracts was not always sufficient for complex a^* formation.

Having determined the stabilizing effect of Triton X-100 on complex a^* formation, we turned to examine stronger detergents, by adding a combination of sodium deoxycholate (DOC) (0.4%, w/v) and Nonidet P-40 (NP-40) (1.2%, v/v) to binding reactions performed with our standard Triton X-100 extracts (Figure 3(c)). Under these conditions, the up-shifted complex a^* was formed with the EP probe in mock-transfected cell extracts, migrating slightly faster than normally (lane 2), and a lower band was seen with the X box probe (lane 4). Complex a was not detected with either probe. The X box-bound complex formed in the presence of DOC/NP-40 is likely to be a mono-

meric form of complex a^* , since it resembles the EP-bound complex a^* , and not complex a , in its response to detergents. The ΔD mutant behaved similarly to the endogenous RFX1 and generated only complex a^* with the EP probe and a lower band with the X box, but not complex a (lanes 6 and 8). Under standard Triton X-100 conditions, the N-terminally deleted RFX1 derivatives $\Delta N1$ and $\Delta N2$ formed a strong complex a and a faint complex a^* (lanes 13, 15, 17, and 19). By contrast, in the presence of DOC/NP-40 these mutants generated complex a^* , but not complex a , similar to the endogenous RFX1 (lanes 14, 16, 18, and 20). While the addition of DOC/NP-40 resulted in the exclusive formation of the up-shifted complex a^* by all these RFX1 derivatives, ΔBCD generated only dimers and monomers of complex a with the EP and X box, respectively (lanes 10 and 12). This mutant, deficient of the entire EDD, was not affected by either mild detergents (Triton X-100; Figure 3(a), lanes 13-16) or moderate detergents (DOC/NP-40; Figure 3(c), lanes 9-12). Collectively, these results suggest that DOC/NP-40 affects complex formation by converting complex a into the a^* form. All RFX1 derivatives containing at least a portion of the EDD formed complex a^* exclusively under these conditions, while ΔBCD , which cannot generate complex a^* since it lacks the EDD, continued to generate complex a .

Formation of the up-shifted complex is an independent property of the RFX1 EDD

Deletion analysis of RFX1 demonstrated that complex a^* formation requires the EDD (Figure 3; Katan-Khaykovich & Shaul, 1998). However, since all EP-binding RFX1 derivatives must contain the RFX1 DBD, using this experimental system we could not determine whether the ability to generate the up-shifted complex is an intrinsic property of the EDD alone. Therefore, we turned to analyze complex formation in a heterologous system, by replacing the RFX1 DBD with the DBD of the yeast Gal4 activator (Figure 4(d), construct 1). The G4-RFXNC fusion was expressed in cells and assayed for complex formation in gel shift with the Gal4 binding site probe. This binding site resembles the EP element and differs from the X box in that it is palindromic, and Gal4-derived proteins bind to it as dimers. Since in the Gal4 system complex a^* could not be identified by its differential formation with palindromic *versus* non-palindromic sites, the two alternative complexes were identified on the basis of their unique response to detergents. When extraction was performed in either the absence or the presence of Triton X-100, the Gal4 DBD alone generated the same single complex (Figure 4(a), lanes 1 and 2). By contrast, two complexes were generated by G4-RFXNC, a faster-migrating band seen in both extraction conditions and a low mobility band obtained only in the presence of Triton X-100 (lanes 5 and 6). This pattern of bands was very similar to that of the endogenous complexes

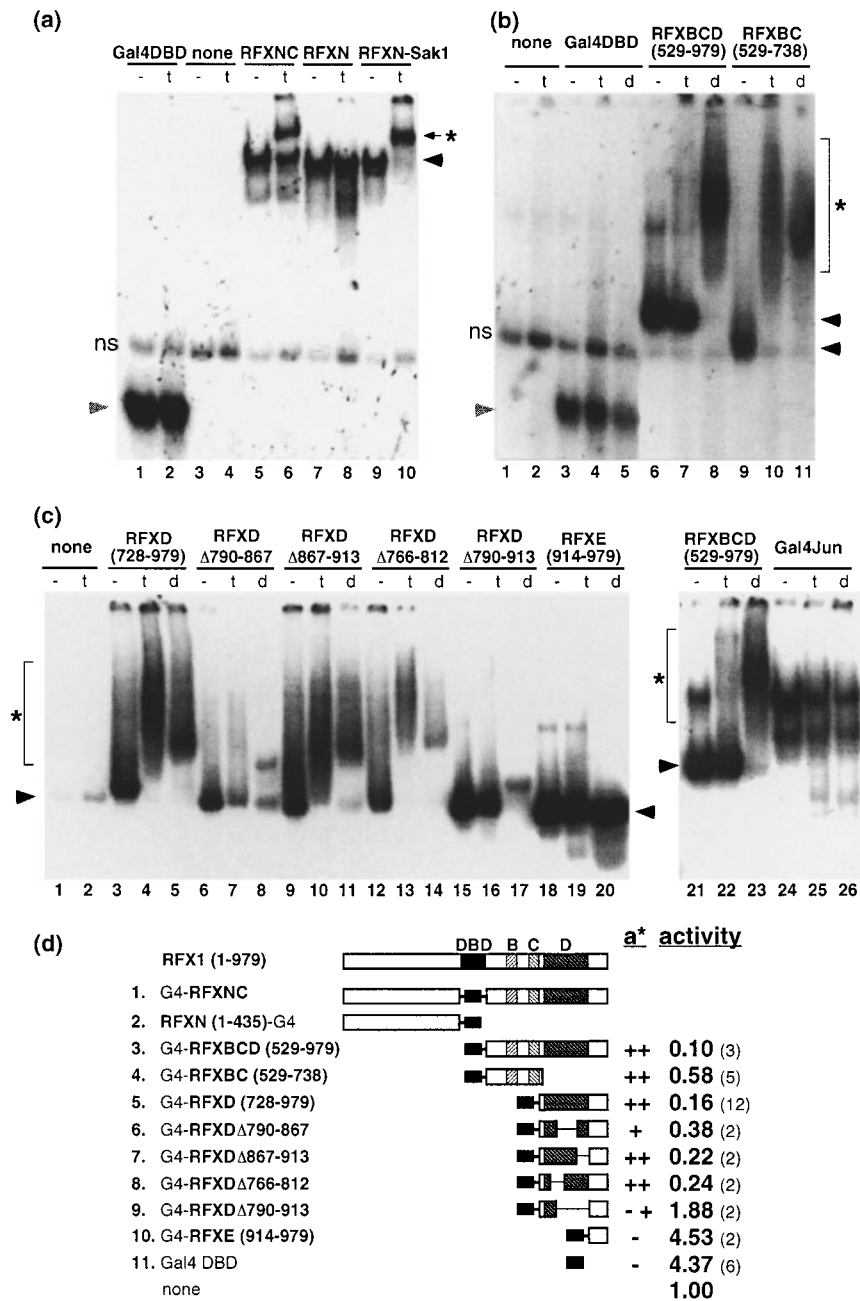


Figure 4. Formation of an up-shifted complex and transcriptional repression by Gal4-RFX1 fusions. (a) to (c) 293T cells were transfected with Gal4 derivative expression plasmids (indicated without the G4) or mock transfected (none), and cellular extracts were analyzed by gel shift with the Gal4 binding site probe. The structure of the Gal4-RFX1 derivatives is shown in (d), and RFXN-G4-Sak1 is shown in Figure 5. The amounts of transfected constructs were adjusted to give complexes with comparable intensities. The extracts were subjected to three different detergent conditions: no detergent treatment (-), extraction with Triton X-100 (t), or extraction with Triton X-100 followed by DOC/NP-40 addition during the binding reaction (d). The positions of complexes a and a* are indicated by a black arrowhead and an asterisk (*), respectively. A gray arrowhead marks the complex of the Gal4 DBD. ns, non-specific. The free probe was allowed to run out of the gel. (d) Structure and transcriptional activity of the Gal4-RFX1 fusions. The structure of RFX1 is shown at the top, with the conserved regions indicated, and the Gal4-RFX1 fusions are shown below. For constructs 2-5 and 10, the RFX1 residues contained in the fusion are indicated. Construct 1 contains the whole RFX1 sequence except for the DBD, residues 436-528. Constructs 6-9 were derived from construct 5 by creating the indicated deletions. A black box represents the Gal4 DBD. The column marked a* represents the capacity to generate the up-shifted a* complex in response to detergents, determined by the experiments shown in (b) and (c). The column marked activity represents the transcriptional activity of the Gal4 fusions in transient cotransfections, according to Katan *et al.* (1997). HepG2 cells were cotransfected with 2 μ g of the indicated Gal4 derivative expression construct, together with the E5G5-luciferase reporter plasmid, containing five Gal4 binding sites, a c-Jun expression plasmid (for activating the reporter), and the SV₂- β -galactosidase internal control plasmid. The normalized luciferase activity is shown relative to the basal activity in the absence of a Gal4 derivative. Each result represents the mean of 2-12 independent experiments, as indicated in parentheses.

formed with the EP probe, the lower band corresponding to complex a, and the higher Triton X-100-dependent band to complex a*. Therefore, these results indicate that the formation of the up-shifted complex does not depend on the RFX1 DBD or on the specific sequence of the EP DNA element.

The Gal4 system was further used for examining whether specific regions of RFX1 can generate complex a* independently, by fusing these regions to the Gal4 DBD (see Figure 4(d) for the construct structure). The N-terminal part of RFX1 could not generate complex a* (construct 2; Figure 4(a), lanes 7 and 8). The G4-RFXBCD fusion, containing the whole RFX1 C terminus, generated a strong band under both extraction conditions, in addition to higher complexes: a relatively weak band seen only in the absence of Triton X-100, and a diffuse complex obtained only in the presence of Triton X-100 (construct 3; Figure 3(b), lanes 6 and 7). This Triton X-100-dependent low-mobility complex was reminiscent of complex a*. In order to verify this, the binding reaction was supplemented with DOC/NP-40. While the Gal4 DBD was unaffected by the different detergents (Figure 4(b), lanes 3-5), G4-RFXBCD was dramatically affected (lanes 6-8); in the presence of DOC/NP-40 the lower band disappeared, and the diffuse low-mobility complex was highly intensified. The behaviour of the complexes generated by G4-RFXBCD under the different conditions was similar to that of the endogenous EP complexes, the lower band seen with and without Triton X-100 corresponding to complex a, and the up-shifted complex formed only in the presence of detergents corresponding to complex a*. Addition of DOC/NP-40 seemed to convert the lower complex into the up-shifted one, as observed with complexes a and a* of the wt RFX1 (Figure 3(c)). These findings indicate that the C-terminal part of RFX1 has the intrinsic ability to generate the up-shifted complex independently. The diffuse appearance of the up-shifted complex formed by G4-RFXBCD (Figure 4(b), lanes 7 and 8) was similar to that of complex a* formed by N-terminally deleted RFX1 derivatives with the EP probe (Figure 3(c), lanes 13, 14, 17, and 18), and was probably due to the absence of the RFX1 N-terminal sequences. The RFX1 N terminus, which cannot generate the up-shifted complex by itself (Figure 4(a), lanes 7 and 8), thus appears to assist in generating a sharp a* band, perhaps by stabilizing this complex.

To further localize the specific region within the RFX1 C terminus responsible for complex a* formation, different Gal4-RFX1 fusions were assayed for ability to generate an up-shifted complex in response to detergents (Figure 4(b)-(d)). G4-RFXBC, containing the B-C region of RFX1, generated a high-mobility and a low-mobility complex in the absence and presence of Triton X-100, respectively (construct 4; Figure 4(b), lanes 9 and 10). In the presence of DOC/NP-40 the low-mobility complex migrated slightly faster (lane 11), as

observed with complex a* formed by the wt RFX1 (Figure 3(c)). G4-RFXD, containing region D of RFX1, generated a similar pattern of complexes (construct 5; Figure 4(c), lanes 3-5). Therefore, the RFX1 sequences contained within these fusion proteins possess the capacity to generate the up-shifted complex independently. Interestingly, in the presence of Triton X-100, G4-RFXBC and G4-RFXD generated only the up-shifted complex (Figure 4(b), lane 10, and (c), lane 4), while G4-RFXBCD generated both an upper and a lower complex (Figure 4(b), lane 7). A similar behaviour was observed in the original EP system, where the wt RFX1 formed both complexes a and a*, while the Δ D mutant generated only complex a* (Figure 3(c), lanes 1 and 5; Katan-Khaykovich & Shaul, 1998). The Gal4 system therefore seems to mimic the original EP system in the fact that C-terminally intact RFX1 derivatives form both complexes, while those containing a partial EDD form only the up-shifted one.

In order to further localize the RFX1 region mediating the formation of the up-shifted complex, several deletions were introduced into G4-RFXD. The behaviour of two deleted derivatives was generally similar to that of the original fusion, although there was a difference in the level of up-shift in response to detergents (constructs 7 and 8; Figure 4(c), lanes 9-14). A third deleted derivative showed a weaker response to detergents (construct 6; lanes 6-8); however, DOC/NP-40 still gave rise to an up-shifted complex. A large deletion within region D nearly abolished the response to Triton X-100, and only a small up-shift was seen in the presence of DOC/NP-40 (construct 9; lanes 15-17). Finally, no response to detergents was seen with a fusion containing only the extreme C terminus of RFX1 (construct 10; lanes 18-20).

Although the various Gal4-RFX1 fusions contained only portions of the RFX1 protein, the up-shifted complex generated by these fusions seems to correspond to complex a* generated by the wt RFX1, by several criteria. First, the relatively slow migration is a typical property of complex a*. Second, the upper and lower complexes formed by the Gal4-RFX1 fusions showed the same response to detergents as complexes a* and a, respectively, of the wt RFX1. Third, while a Gal4-RFX1 fusion containing the entire B-C-D region formed both the upper and the lower complex in the presence of Triton X-100, fusions containing a partial EDD formed only the upper complex; the same phenomenon was observed with EP-bound derivatives of the wt RFX1. Collectively, our results indicate that several different parts of the B-C-D region can form the up-shifted complex a* independently. The RFX1 sequences of G4-RFXBC and G4-RFXD have this capacity, and within region D this property seems to be found in at least two different parts, since none of the smaller deletions could completely abolish the up-shifted complex. To verify that this response to detergents is specific to the C-terminal part of RFX1, Gal4 fusions of several

heterologous proteins were tested, but these showed no response to detergents, as demonstrated for Gal4-Jun (lanes 24-26).

Formation of an up-shifted complex by Gal4-RFX1 fusions correlates with transcriptional repression

An analysis of deletion mutants and chimeric proteins has localized a transcriptional repression domain to the C-terminal part of RFX1 (Katan *et al.*, 1997). This repression function has the following properties: (i) it is independent, i.e. it can be transferred to a heterologous system by fusing RFX1 sequences to the Gal4 DBD; (ii) it is not localized to a single region within the RFX1 C terminus, but rather several different regions can repress transcription independently. Since the ability to generate the up-shifted complex shares these two properties, we examined a possible link between complex a* formation and transcriptional repression by comparing these functions in each of the Gal4-RFX1 fusions (Figure 4(d)). The transcriptional activity of each Gal4 fusion, in transient cotransfections with a reporter plasmid containing five Gal4 binding sites, is shown relative to the basal activity in the absence of a Gal4 derivative. In this system, the Gal4 DBD, which cannot form complex a*, induced a mild activation of transcription (Figure 4(d), construct 11). By contrast, G4-RFXBCD and each of its derivatives, G4-RFXBC and G4-RFXD, exhibited a repressive effect (constructs 3-5), as previously shown (Katan *et al.*, 1997). The variations in repression level may be due, at least partially, to the fact that the three fusions differ significantly in their intrinsic DNA-binding capacity (Katan *et al.*, 1997). In particular, the relatively weak repressive effect of G4-RFXBC may result from its low DNA-binding activity. A transcriptional inhibitory effect of the B-C region was also recently reported (Iwama *et al.*, 1999). These three fusions all generated the up-shifted complex efficiently (Figure 4(b)-(d), constructs 3-5). Three partially deleted derivatives of G4-RFXD showed a capacity to both repress transcription and generate the up-shifted complex (constructs 6-8). The ability of construct 6 to form the up-shifted complex was weaker than that of the others, as was its repressive capacity, suggesting that residues 790-867 of RFX1 may be particularly important for both functions. A severely deleted derivative of G4-RFXD, which showed only a minor capacity to form the up-shifted complex, was unable to repress transcription (construct 9). This fusion induced a very mild activation, lower than that of the Gal4 DBD. Finally, G4-RFXE, containing the extreme C terminus of RFX1, showed no capacity to form the up-shifted complex or repress transcription, and induced a mild activation, similar to that of the Gal4 DBD (construct 10). Collectively, these results show a

correlation between the two functions located within the RFX1 C terminus, the ability to generate the up-shifted complex and transcriptional repression.

Formation of an up-shifted complex by Gal4-Sak1 fusions

The RFX1 region mediating complex a* formation includes the evolutionarily conserved regions B, C, and D, and each of the a*-forming Gal4-RFX1 fusions contained at least a part of these sequences. Since the formation of the up-shifted complex appeared to be mediated by the conserved regions, it was possible that other RFX family members may share this property. This possibility was addressed by examining the ability of the *S. pombe* RFX homologue Sak1 (Wu & McLeod, 1995) to generate complex a*. A portion of Sak1 containing the B-C-D region was fused to the Gal4 DBD, and the G4-Sak1 fusion was expressed in human cells and assayed for complex formation, as described above (Figure 5). In the absence of detergents, this fusion generated a single complex (lanes 4 and 10), similar in mobility

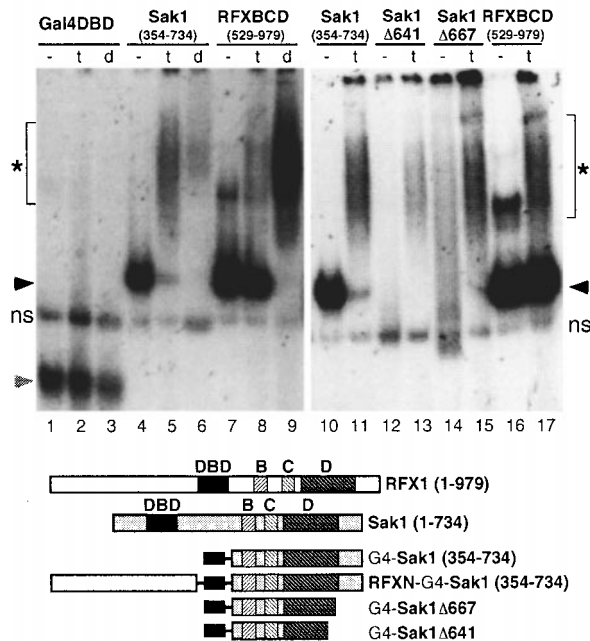


Figure 5. Structure and gel shift analysis of Gal4-Sak1 fusions. 293T cells were transfected with expression plasmids of Gal4 derivatives (indicated without the G4): 3 μ g for G4-RFXBCD (structure shown in Figure 4(d)) and 6 μ g for the rest. Cellular extracts were analyzed, and the complexes are marked, as for Figure 4. The structures of RFX1, Sak1, and the Gal4-Sak1 fusions are shown, with the conserved regions indicated. The Sak1 residues present in each construct are indicated in parentheses. RFXN-G4-Sak1 also contains residues 1-435 of RFX1. The last two constructs are truncated derivatives of G4-Sak1, terminated at the indicated residues. A black box represents the Gal4 DBD.

to complex a of G4-RFXBCD (lanes 7 and 16). In the presence of Triton X-100, G4-Sak1 generated an up-shifted complex (lanes 5 and 11), as did G4-RFXBCD (lanes 8 and 17), indicating that the Sak1 portion present in this fusion can form complex a*. Interestingly, while G4-RFXBCD formed both complexes a and a* in the presence of Triton X-100, G4-Sak1 generated mainly complex a* and only a faint band of complex a. Upon addition of DOC/NP-40, both fusions generated only complex a* (lanes 6 and 9). Sak1 was further examined by creating a Gal4 fusion containing the N terminus of RFX1 combined with the B-C-D region of Sak1 (RFXN-G4-Sak1, Figure 5). Since the RFX1 N terminus could not generate the up-shifted complex by itself (Figure 4(a), lanes 7 and 8) but only when combined with the RFX1 C terminus (lanes 5 and 6), we wished to examine whether the B-C-D region of Sak1 can replace the RFX1 C terminus in mediating complex a* formation when combined with the RFX1 N terminus. In the absence of detergent, RFXN-G4-Sak1 generated complex a, but in the presence of Triton X-100, the major complex seen was a* (Figure 4(a), lanes 9 and 10), indicating that the B-C-D region of Sak1 can replace the RFX1 C terminus. Here, too, the Sak1 fusion showed a strong preference for complex a* (lane 10), while G4-RFXNC efficiently generated both complexes a and a* (lane 6).

Transcriptional repression by Gal4-Sak1 fusions

In order to determine whether the correlation between the formation of the up-shifted complex and transcriptional repression detected in RFX1 also extends to Sak1, the transcriptional activity of G4-Sak1 was examined (Figure 6(a)). However, this fusion did not repress transcription and even induced a weak activation, similar to the Gal4 DBD. This result could be due to the absence of any transcriptionally active regions within this portion of Sak1 or to the existence of both activating and repressing regions, which neutralize each other's effects. The extreme C terminus of Sak1 contains short regions relatively rich in either acidic residues or glutamine, such as are commonly found in activation domains, and was therefore suspected to contain an activating region. In order to eliminate the effect of this putative activation domain, two truncated derivatives of G4-Sak1 were generated (Figure 5). In contrast to the original fusion, both G4-Sak1 Δ 667 and G4-Sak1 Δ 641 repressed transcription, in a dose-dependent manner (Figure 6(b) and (c)). These results indicate that in Sak1, as in RFX1, the B-C-D region can repress transcription. As suspected, the extreme C terminus of Sak1 may contain an activation domain, which neutralized the repressive effect of the B-C-D region in G4-Sak1. The repressive effect of the Sak1 fusions was substantially weaker than that of G4-RFXBCD (Figure 6(b)). Gel shift and Western analyses indicated that the Gal4-

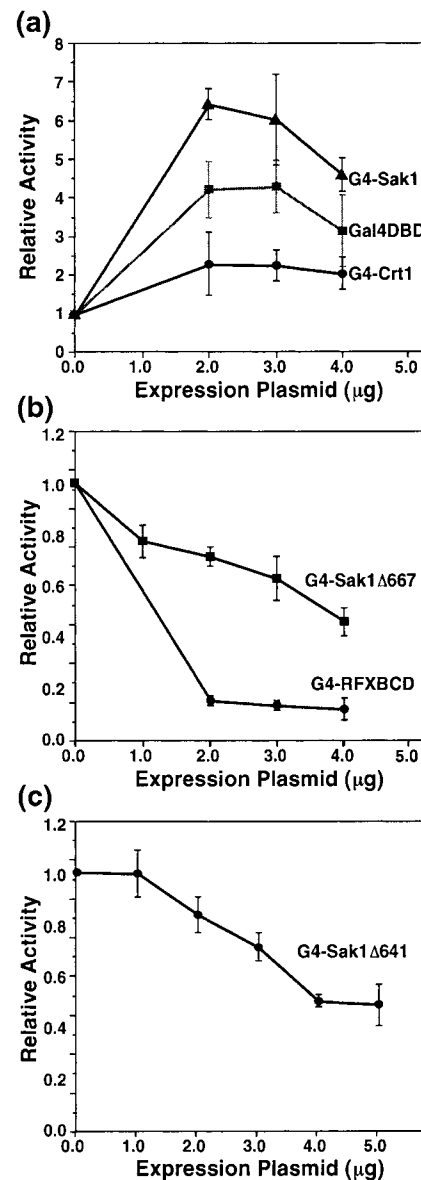


Figure 6. Transcriptional activity of Gal4 fusions containing the C termini of RFX proteins. HepG2 cells were cotransfected with increasing amounts of the indicated Gal4 derivative expression constructs (structure shown in Figures 4(d), 5, and 7), together with the E5G5-luciferase reporter plasmid, containing five Gal4 binding sites, a c-Jun expression plasmid (for activating the reporter), and the SV₂- β -galactosidase internal control plasmid. The normalized luciferase activity is shown relative to the basal activity in the absence of a Gal4 derivative. Each result represents the mean and SD of 2-3 independent experiments.

Sak1 fusions were expressed less well than G4-RFXBCD (Figure 5 and data not shown; notice that in Figure 5 the amount of transfected G4-RFXBCD DNA was half of that of the Sak1 fusions). Therefore, the weaker repressive capacity of the Gal4-Sak1 fusions may be partially due to their ineffi-

cient expression. Collectively, our results demonstrate a correlation between complex a* formation and transcriptional repression, in both the human protein RFX1 and its *S. pombe* homologue Sak1.

Complex formation and transcriptional activity of Crt1 fusions

The *S. cerevisiae* RFX protein Crt1, which regulates the response to replication block and DNA damage (Huang *et al.*, 1998), contains the conserved region B but seems to bear no obvious homology to regions C and D. To examine whether Crt1 shares with its homologues the ability to generate alternative complexes, the C-terminal part of Crt1 was fused to the Gal4 DBD, and this G4-Crt1 fusion was expressed in human cells and assayed by gel shift, as above (Figure 7(a)). In the absence of detergents, G4-Crt1 formed a single lower band of complex a, but in the presence of Triton X-100 the lower and upper complexes a and a*, respectively, were generated (lanes 4 and 5). The addition of DOC/NP-40 resulted in the exclusive formation of complex a* (lane 6). The behaviour of G4-Crt1 under the different conditions resembled that of G4-Sak1 (Figure 7(a), lanes 7-9, and Figure 5, lanes 4-6) and G4-RFXBCD (Figure 5, lanes 7-9), indicating that the ability to generate the up-shifted complex is conserved in the RFX family from *S. cerevisiae* to man. However, the ratio between the two complexes under standard conditions of mild detergent (Triton X-100) differed between the three fusions, as G4-Sak1 showed a clear preference for complex a*, while G4-RFXBCD and G4-Crt1 formed complex a at least as efficiently as complex a*.

To further examine the conservation of the complex formation property in Crt1, the RFX1-Crt1 chimera, in which the RFX1 C terminus was replaced with the corresponding region of Crt1, was tested in gel shift with the EP and X box probes under standard Triton X-100 conditions (Figure 7(b)). While the C-terminally deleted Δ BCD RFX1 mutant generated only complex a with the EP probe (lane 5), the RFX1-Crt1 chimera formed both complexes a and a* (lane 7), similar to the wt RFX1 (lane 3). Therefore, the Crt1 C terminus can functionally replace the B-C-D region of RFX1 and mediate the formation of alternative complexes. However, the complex formation pattern of RFX1-Crt1 with the non-palindromic X box probe differed from that of the wt RFX1. While RFX1 generated a dimeric complex a (lane 4), the RFX1-Crt1 chimera generated a lower distinct band and a smeary complex above it (lane 8). These lower complexes were reminiscent of those formed by RFX1 mutants with the X box, the lower band migrating similarly to that of Δ BCD (lane 6), and the smeary complex appearing similar to that of Δ D (lane 10). Since the complexes formed by Δ BCD and Δ D with the X box appear to be monomeric forms of complexes a

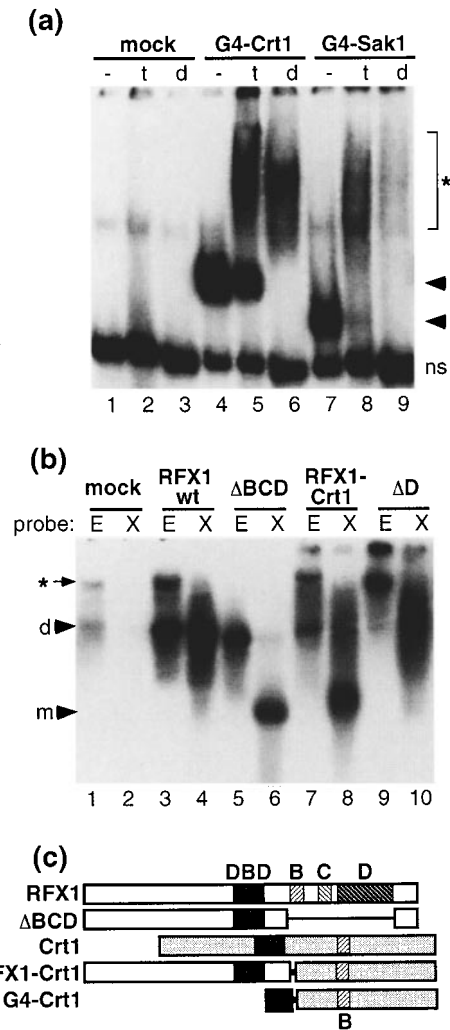


Figure 7. Formation of an up-shifted complex by Crt1 fusions. (a) Gel shift analysis of Gal4 fusions (structure shown below and in Figure 5) overexpressed in 293T cells was performed as for Figures 4 and 5. (b) RFX1 derivatives (structure shown below) were overexpressed in 293T cells. Cells were extracted in the presence of Triton X-100 and analyzed by gel shift with the EP (E) and X box (X) probes. The positions of complexes a and a* are indicated by an arrowhead and an asterisk (*), respectively. d, dimer; m, monomer. (c) Structure of RFX1, Crt1, and their derivatives. A black box in G4-Crt1 represents the Gal4 DBD.

and a*, respectively, it is likely that the RFX1-Crt1 chimera generated with the X box a mixture of a-type and a*-type monomeric complexes, consistent with the lack of dimerization capacity in this protein (Figure 2).

To examine whether the ability to repress transcription is conserved in the Crt1 C terminus, the transcriptional activity of G4-Crt1 was analyzed. G4-Crt1 did not repress transcription, but rather induced a very mild activation, somewhat weaker than that of the Gal4 DBD (Figure 6(a)). Thus, despite its ability to mediate

the formation of the up-shifted complex, the Crt1 C terminus shows no repression capacity in this assay. Altogether, this comparative analysis of the RFX family reveals on one hand a conservation of the ability to generate an up-shifted complex *via* the C terminus, and on the other hand a gradual alteration in the dimerization and transcriptional repression properties of this region, ranging from none to moderate and high capacity in Crt1, Sak1, and RFX1, respectively.

Discussion

Members of the novel RFX family of DNA-binding proteins function in various biological systems, such as transcriptional regulation of viral and cellular genes (RFX1), cell fate regulation in *S. pombe* (Sak1) (Wu & McLeod, 1995), and the DNA damage and replication block checkpoint in *S. cerevisiae* (Crt1) (Huang *et al.*, 1998). Since the RFX proteins share, in addition to a DNA-binding domain, a conserved C-terminal region, characterization of this region should be important for understanding how the functions of these proteins are exerted. Region D was initially characterized as the RFX1 dimerization domain (Reith *et al.*, 1990), yet full dimerization of RFX1 requires both regions D and B-C, which together constitute the Extended Dimerization Domain (EDD) (Katan-Khaykovich & Shaul, 1998). Sequence analysis of the RFX family revealed that the *S. pombe* protein Sak1 contains the B-C-D region, similar to other family members from *C. elegans*, mice, and humans, whereas the *S. cerevisiae* protein Crt1 shows obvious homology only to region B (Figure 1; Emery *et al.*, 1996a). While such homology is suggestive of common structural or functional characteristics between RFX1 and its yeast orthologues, only experimental analysis could determine if, and to what extent, this is actually the case. Our analysis reveals the unique nature of the EDD, as a complex domain that has evolved from an ancient motif of unknown function, to mediate an extremely stable protein-protein interaction, as well as transcriptional repression.

One unusual property of the EDD of RFX1 is its ability to generate alternative homodimeric DNA-protein complexes (Katan-Khaykovich & Shaul, 1998). Here we show that this property is evolutionarily conserved, as the C-terminal parts of Sak1 and Crt1 can generate the up-shifted complex a* (Figures 5 and 7), indicating structural similarity between the C termini of the three proteins. However, analysis of their dimerization properties revealed functional differences, in that the C terminus of RFX1 mediated very stable dimerization, that of Sak1 mediated moderately stable dimerization, and that of Crt1 showed no dimerization capacity (Figure 2). Therefore, while the C-terminal motif is present in the most ancient known RFX protein, Crt1, the dimerization function was probably acquired later in evolution. In view of the con-

comitant appearance of the C-D region and the dimerization function in Sak1, it is likely that the C-D region is responsible for the major interaction formed by the EDD and may constitute the core dimerization motif, consistent with the *in vitro* dimerization capacity of region D (Reith *et al.*, 1990). The fact that the distance between regions C and D is conserved, while the distance between regions B and C is more variable (Figure 1), further suggests that regions C and D function together in mediating dimerization, while the role of region B may be different or accessory.

The molecular basis for the different mobilities of complexes a and a* has not been clearly established. Although the lower mobility of complex a* may result from the presence of additional proteins besides the RFX1 dimer in this complex, several observations argue against this possibility (Katan-Khaykovich & Shaul, 1998). The fact that complex a* is efficiently formed even by highly overexpressed RFX1 derivatives indicates that this complex does not contain an additional component of limiting concentration. In addition, in immunoprecipitation experiments we have not been able to detect any protein that stably interacts with the overexpressed RFX1; nor could we detect a stable interaction between a*-forming RFX1 mutants that are unable to dimerize (such as ΔD), which would suggest a higher-order oligomerization of RFX1 itself (data not shown). While the possibilities of additional proteins or RFX1 oligomerization cannot be entirely ruled out, alternative RFX1 conformations are more likely to account for the different mobilities of the two complexes. The ability to convert one complex into the other under certain detergent conditions, with complex a* formation being favoured by detergents (Figure 3), further supports this possibility.

Our previous study suggested that the formation of the up-shifted complex a*, mediated by the RFX1 EDD, is linked to dimerization. For example, efficient dimerization and complex a* formation are both properties of the cellular RFX1, but not the recombinant bacterially expressed protein (Katan-Khaykovich & Shaul, 1998). These two functions of the EDD, however, differ in their sequence requirements. While different parts of the RFX1 EDD can form the up-shifted complex independently (Figure 4), stable dimerization requires an intact EDD (Reith *et al.*, 1990; Katan-Khaykovich & Shaul, 1998). It therefore appears that the EDD is composed of several subdomains, each of which can acquire the a* form, and their combined properties result in a stable interaction. Thus, acquisition of the a* form by the different RFX1 subregions appears to be a process that precedes dimerization. That acquisition of the a* form precedes dimerization is also the case from an evolutionary point of view, as the C termini of Crt1, Sak1, and RFX1 can all generate the up-shifted complex, but only those of Sak1 and RFX1 can mediate dimerization.

Interestingly, mutational analysis of RFX1 demonstrated an inverse correlation between complex a* formation and dimerization. While the wild-type RFX1 and a Gal4-RFX1 fusion containing the entire EDD formed both complexes a and a*, partial EDD deletions, which abolish dimerization, resulted in the exclusive formation of the up-shifted complex (Figures 3, 4, and 7(b); Katan-Khaykovich & Shaul, 1998). Based on these observations, we have suggested that the a* form represents a specific conformation of the EDD, in which the strength dimerization is reduced, as compared to complex a. Thus, the two complexes may contain RFX1 homodimers that differ in the nature of the intersubunit interaction, mediated by the EDD (Figure 8(a), top row). Each RFX1 subunit seems to have the intrinsic ability to acquire the a* conformation; however, the strong interaction between the subunits can constrain the EDD in the a conformation, resulting in an equilibrium between a-type and a*-type dimers. It should be noted that, although it is suggested here that the difference between complexes a and a* lies in the conformation of the EDD, this model is not incompatible with the existence of additional proteins in complex a*, a possibility that seems less likely but cannot be ruled out. The effect of detergents on complex formation is consistent with the above model. Thus, comparable levels of the two complexes are observed under conditions of mild detergent (Triton X-100), which appear to stabilize the a* conformation. The absence of detergents or conditions of moderate detergents (DOC/NP-40) shift the equilibrium towards the a or a* form, respectively (Figure 3). The effect of DOC/NP-40 may be due to both stabilization of the a* conformation and disruption of the intersubunit interaction, leading to the exclusive formation of complex a*.

While the three RFX proteins share the ability to acquire the a* form, they also seem to differ with respect to this property. Most notably, whereas dimerization-deficient RFX1 derivatives (such as Δ D) generate complex a* exclusively, the Crt1 fusions, which are also unable to dimerize *via* the conserved region, generate both complexes a and a* to comparable levels (Figure 7). Thus, the intrinsic ability of Crt1 to acquire the a* form appears to be weaker than that of RFX1. Based on their specific dimerization and complex formation properties, the different RFX proteins can be described in terms of the above model (Figure 8(a)). The *S. cerevisiae* protein Crt1 has a limited capacity to acquire the a* conformation and no dimerization ability, and thus exists in the form of a-type and a*-type monomers. In the *S. pombe* protein Sak1, the capacity to acquire the a* conformation is stronger; consequently, this protein is found mainly in the a* conformation, as observed with the Gal4-Sak1 fusions. Importantly, the B-C-D region of Sak1 can form a moderately stable interaction, so that some of the Sak1 proteins, although

not all, generate dimers. An enhanced intersubunit interaction in the human protein RFX1 causes this protein to exist exclusively in the form of stable dimers. Moreover, in addition to the stable interaction in complex a*, RFX1 can form a new, stronger type of interaction, which constrains the EDD in the a conformation. Thus, the ancestral conserved motif of Crt1 may have developed into the complex EDD of RFX1, which can generate different types of interactions, corresponding to the two alternative complexes.

A comparison between the three RFX proteins reveals that they differ in their dimerization properties and in their DNA binding sites. The known binding sites of Crt1, present in the promoters of its target genes, are non-palindromic half-sites resembling the X box (Huang *et al.*, 1998). The binding sites of Sak1 are presently unknown. RFX1 binds to half-sites, such as the X box, yet it shows more stable binding to palindromic EP sites, such as those found in viral enhancers (David *et al.*, 1995). This stabilization is apparently due to the fact that the EDD, although not essential for DNA-binding, stabilizes the binding of RFX1 to the palindromic EP site (Katan-Khaykovich & Shaul, 1998). Thus, the alteration in dimerization status, from monomers in Crt1, to moderately stable dimers in Sak1, to very stable dimers in RFX1, is paralleled by the alteration in the DNA binding site, from half-sites bound by Crt1 to palindromic sites bound by RFX1 (Figure 8(b)). In view of these data, it appears that the ancestral RFX protein was bound as a monomer to a DNA half-site. It already contained the conserved C-terminal motif, yet this motif could not mediate dimerization. The later acquisition and enhancement of the dimerization function, accompanied by the duplication of the original DNA half-site to form an inverted repeat, resulted in a protein dimer stably bound to a palindromic DNA site. Thus, the EDD of RFX1 appears to have evolved from the original conserved motif present in Crt1 in order to mediate stable binding to palindromic EP sites.

The role of the conserved C terminus in Crt1 remains unclear. It is possible that it serves a function entirely different from that of the RFX1 EDD, and the ability to mediate protein-protein interactions was acquired later in evolution. Alternatively, the Crt1 C terminus may participate in a protein-protein interaction too weak to be detected by our assay. Since the Crt1 binding sites are half-sites, there seems to be no need for an interaction that stabilizes the binding to palindromic DNA sites. However, the promoters of Crt1 target genes contain several Crt1 binding sites (Huang *et al.*, 1998), so that an interaction between Crt1 molecules bound to different sites within a promoter could result in cooperative binding or the formation of a higher-order structure. Whether such an interaction indeed exists, and its role in the function of Crt1, remain to be determined.

The significance of the unique biochemical properties of the RFX proteins to their function

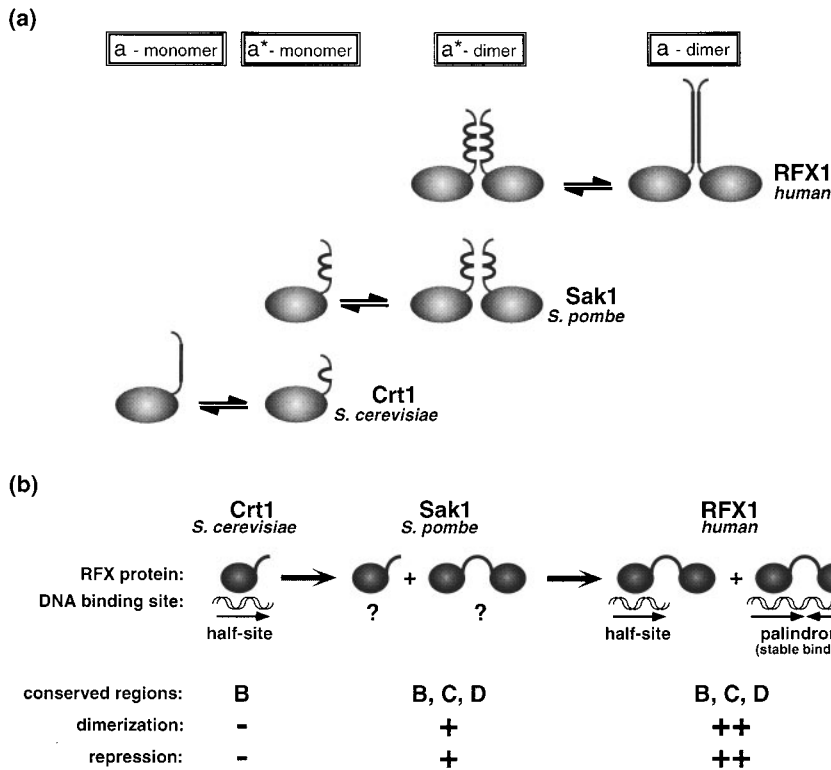


Figure 8. (a) A model comparing between Crt1, Sak1, and RFX1, based on their dimerization and complex formation properties. Monomeric and dimeric a and a* forms of the different RFX proteins are shown schematically. A thick black line represents the conserved C-terminal region in its a conformation (straight line) or a* conformation (curved line). The increase in the number of curves from Crt1 to RFX1 indicates an enhanced capacity to acquire the a* conformation. The Crt1 C terminus has a limited capacity to acquire the a* conformation, and cannot mediate stable dimerization; therefore Crt1 exists as monomers of the a and a* forms. The C terminus of Sak1 has an increased capacity to acquire the a* conformation, and can mediate moderately stable dimerization; therefore Sak1 exists as dimers and monomers, which are mainly of the a* form. The RFX1 C terminus (EDD) has a high dimerization capacity, so that all RFX1 proteins exist as stable dimers. The a*-formation capacity of RFX1 is also

high, and every molecule has an intrinsic ability to acquire the a* conformation. However, the existence of a very strong interaction between the dimer subunits constrains some of the molecules in the a conformation; a somewhat weaker interaction allows other dimers to acquire the a* conformation, resulting in an equilibrium between a type and a* type dimers. Mutations of RFX1 that disrupt the intersubunit interaction drive the equilibrium to the exclusive formation of complex a* (b) Alterations in the properties of RFX proteins and their DNA binding sites. Top: Evolutionary progression from protein monomers to stably linked dimers, in parallel with the duplication of a half-site DNA to generate an inverted repeat. The RFX proteins are shown schematically as dimers or monomers, bound to an EP-homologous DNA half-site or palindrome. Based on the dimerization properties of their conserved C termini, Crt1, Sak1, and RFX1 appear to exist as monomers, moderately stable dimers, and very stable dimers, respectively. The known DNA binding sites of Crt1 are half-sites (Huang *et al.*, 1998), while RFX1 binds to both half-sites and palindromes. RFX1 binds more stably to palindromic DNA sites than to half-sites (David *et al.*, 1995), palindromic binding being stabilized by the EDD (Katan-Khaykovich & Shaul, 1998). The binding sites of Sak1 are not known. Bottom: The conserved regions, dimerization capacity, and transcriptional repression capacity of the C termini of RFX proteins are shown. The dimerization and repression capacities are marked as none (-), moderate (+), and high (++), and show a parallel enhancement with evolutionary progression.

is suggested by observations, which link between the formation of the up-shifted complex *in vitro* and transcriptional repression *in vivo* (Figure 4). The repressive effect of RFX1 (Katan *et al.*, 1997) may underlie the transcriptional inhibitory activity of its DNA binding sites, observed in several different systems (Weisinger *et al.*, 1988; Reinhold *et al.*, 1995; Blake *et al.*, 1996; Ori *et al.*, 1998; Liu *et al.*, 1999). Formation of the up-shifted complex correlates with transcriptional repression in several ways. First, both functions are mediated by the conserved B-C-D region of RFX1, or the EDD. Second, both are independent functions of the EDD that can be transferred to the heterologous Gal4 system. Third, in both cases the activity was found in several different parts of the B-C-D region. Thus, the B-C-D region of RFX1 appears to be composed of sev-

eral partially redundant subregions, each of which can function independently to generate complex a* and repress transcription. The combined activity of these subregions results in the efficient formation of the up-shifted complex and potent repression. Fourth, the correlation between these properties extends to the B-C-D region of Sak1 (Figures 5 and 6). Collectively, these data confirm and extend the results of our previous study (Katan-Khaykovich & Shaul, 1998), suggesting that the two properties of the B-C-D region are functionally related, so that the repressive effect of RFX1 (and its homologue Sak1) may be exerted *via* the formation of the up-shifted complex.

The C termini of RFX1 and Sak1 differ from that of Crt1 in their ability to mediate dimerization and transcriptional repression. The B-C-D region of

RFX1 exhibits stronger repression activity than that of Sak1 (Figure 6). While the weaker repression of the Gal4-Sak1 fusions may be partially due to their lower expression or to their C-terminal truncations, the large difference between the repression activity of the different fusions (twofold relative to the basal level by G4-Sak1 Δ 667 and G4-Sak1 Δ 641, as compared to eightfold by G4-RFXBCD) suggests that the intrinsic repression capacity of Sak1 is lower than that of RFX1. Thus, the repression and dimerization functions of the RFX C terminus seem to exhibit the same hierarchy in the three proteins: high capacity in RFX1, moderate capacity in Sak1, and no capacity in Crt1 (Figure 8(b)). The ability to generate the a* form, which is common to these three proteins, is therefore not sufficient for transcriptional repression, as is the case for dimerization. While the conserved motif mediating complex a* formation seems to exist in Crt1, both its repression and dimerization properties can be detected only in Sak1 (moderately) and RFX1 (strongly). As discussed above, even the intrinsic ability of Crt1 to acquire the a* form appears to be weaker than that of RFX1. It is likely that structural alterations that the C-terminal motif has undergone in the course of evolution account for such differences; the nature of these alterations awaits further analysis.

The lack of intrinsic repression activity in the Crt1 C terminus is interesting in view of the fact that Crt1 is a transcriptional repressor, which negatively regulates the expression of its target genes by interacting with the Tup1-Ssn6 corepressor complex (Huang *et al.*, 1998). Consequently, the three RFX proteins appear to repress transcription by different mechanisms: by the intrinsic activity of the dimerization/repression domain in RFX1 and Sak1, and *via* the Tup1-Ssn6 complex in Crt1. The RFX family thus shows a conservation of the repression function, although the mechanisms of repression differ. While our results do not rule out an interaction between RFX1 or Sak1 and the corresponding Tup1 homologues, they raise the possibility that in RFX1 and Sak1 the dimerization/repression domain fulfils a function analogous to that of the Crt1-Tup1-Ssn6 interaction.

In comparison with some well-characterized dimerization motifs, such as the leucine zipper and the helix-loop-helix, the EDD of RFX proteins shows unique characteristics, in being independent of the DBD, in its complex structure and ability to generate alternative homodimeric complexes, and in the apparent enhancement of its dimerization capacity during evolution. Finally, we note that in several respects the EDD resembles the well-characterized ligand-binding domain (LBD) of the nuclear receptors. The LBD is an independent functional domain, which mediates dimerization and can exist in alternative conformations (Fawell *et al.*, 1990; Mangelsdorf & Evans, 1995; Perlmann *et al.*, 1996; Moras & Gronemeyer, 1998), as appears to

be the case with the EDD. It has been proposed that the alteration in the LBD conformation involves the disruption of an intramolecular protein-protein interaction, which arrests the LBD in the apo conformation (Vivat *et al.*, 1997; White *et al.*, 1997; Moras & Gronemeyer, 1998). This conformational transition occurs upon ligand binding and can be mimicked by point mutations. A similar situation seems to exist in RFX1, where mutations and detergents may alter the conformation of the EDD, apparently by disrupting an interaction that arrests it in the a conformation; however, the RFX1 mutations disrupt an intermolecular interaction between the dimer subunits. The nuclear receptors were proposed to exist in an equilibrium between alternative LBD conformations, and thus ligands may function by stabilizing or destabilizing a specific conformation (Schulman *et al.*, 1996). Interestingly, it was recently demonstrated that the LBD of the germ cell nuclear factor (GCNF), an orphan nuclear receptor, has the potential to adopt two different conformations with distinct dimerization properties (Greschik *et al.*, 1999). Such a situation, which may allow the regulation of GCNF dimerization, is similar to that proposed for the RFX1 EDD (Figure 8(a)). Yet ligand-binding transcription factors are not unique to the nuclear receptor family, as the aryl hydrocarbon receptor, a basic helix-loop-helix protein, is ligand-inducible (Burbach *et al.*, 1992). While no ligand of RFX proteins is presently known, it is possible that such molecules exist, which may act to induce conformational transitions in the conserved C terminus. Whatever are the mechanisms controlling these conformational transitions, they are likely to play an important role in the various biological processes regulated by the RFX proteins.

Materials and Methods

Cell culture and transfections

Cells were cultured in Dulbecco's modified Eagle minimal essential medium (GIBCO Laboratories) containing 100 units/ml penicillin and 100 μ g/ml streptomycin, supplemented with 8% (v/v) fetal bovine serum. Transfection, by the calcium phosphate precipitation method, and analysis of luciferase activity were performed as described (Katan *et al.*, 1997). For luciferase assays, 6 cm plates of HepG2 cells were cotransfected with 0-4 μ g of a Gal4 derivative expression plasmid, 1.5 μ g of the E5G5-luciferase reporter plasmid, 1 μ g of the SV2- β -galactosidase internal control plasmid, and 0.3 μ g of a c-Jun expression plasmid. The amount of SV2 elements and the total amount of DNA were kept constant in each experiment. The normalized luciferase activity of each plate was calculated by dividing the results of the luciferase assay by those of the β -galactosidase assay.

Plasmid constructions

Plasmids expressing the wt RFX1, RFX1 deletion mutants, RFX1-Sak1, and RFX1-Crt1 are based on pSG5RFX1, which expresses the RFX1 cDNA under the control of SV2 (Siegrist *et al.*, 1993), and the expressed proteins are tagged at their N termini with the HA epitope. Expression plasmids containing the Gal4 DBD (residues 1-147) are based on the pECE vector, and their expression is directed by SV2. The E5G5-luciferase reporter plasmid contains five tandem copies of the Gal4 binding site and five tandem copies of the HBV enhancer E element. In order to generate fusion constructs, Sak1 (residues 354-734) and Crt1 (residues 392-811) sequences were amplified by PCR from genomic DNA of *S. pombe* and *S. cerevisiae*, respectively. These sequences were cloned in-frame downstream of the Gal4 DBD, to generate G4-Sak1 and G4-Crt1, respectively, or downstream of an RFX1 sequence truncated at residue 603, to generate RFX1-Sak1 and RFX1-Crt1, respectively. RFXN-G4-Sak1 was generated by cloning the Sak1 sequence downstream of the Gal4 DBD in RFXN-G4. G4-Sak1 Δ 667 was generated by PCR cloning of Sak1 residues 354-667 downstream of the Gal4 DBD. A *SpeI-XbaI* deletion in G4-Sak1 generated G4-Sak1 Δ 641, truncated after residue 641. The structure of the other plasmids has been described (Katan *et al.*, 1997).

Gel shift analysis

Whole-cell extracts were prepared by lysing 6 cm plates with 50-100 μ l of buffer A (20 mM Hepes-KOH (pH 7.9), 250 mM NaCl, 1% Triton X-100, 5 mM EDTA, 1 mM DTT, protease inhibitors, and phosphatase inhibitors), or by using a similar buffer lacking Triton X-100, followed by three rounds of freeze/thawing. The lysates were cleared by centrifugation. Gel shift analysis was conducted as described (Dikstein *et al.*, 1990), with several modifications. The binding reaction was performed for 45 minutes on ice with 2×10^4 cpm of labeled probe and 6-8 μ l of whole-cell extract, in a total volume of 15 μ l, and the samples were run on a 5% polyacrylamide gel. The free probe was allowed to run out of the gel, in order to achieve a good separation of the DNA-protein complexes. The double-stranded oligonucleotides (sequences shown by Katan *et al.* (1997) for the Gal4 binding site, and by Katan-Khaykovich & Shaul (1998) for the EP and X box sites) were end-labeled by a fill-in reaction. Where indicated, DOC (0.4%) and NP-40 (1.2%) were added to the binding reaction together with the cell extracts.

Acknowledgments

We thank S. Budilovsky for her excellent technical assistance. This work was supported by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities and by the Ebner Family Biomedical Research Foundation at the Weizmann Institute of Science in Memory of Alfred and Dolfi Ebner.

References

Agami, R. & Shaul, Y. (1998). The kinase activity of c-Abl but not v-Abl is potentiated by direct interaction with RFX1, a protein that binds the enhan-

cers of several viruses and cell-cycle regulated genes. *Oncogene*, **16**, 1779-1788.

- Baxeavanis, A. D. & Vinson, C. R. (1993). Interactions of coiled coils in transcription factors: where is the specificity? *Curr. Opin. Genet. Dev.* **3**, 278-285.
- Ben-Levy, R., Faktor, O., Berger, I. & Shaul, Y. (1989). Cellular factors that interact with the hepatitis B virus enhancer. *Mol. Cell Biol.* **9**, 1804-1809.
- Blake, M., Niklinski, J. & Zajac-Kaye, M. (1996). Interactions of the transcription factors MIBP1 and RFX1 with the EP element of the hepatitis B virus enhancer. *J. Virol.* **70**, 6060-6066.
- Bolwig, G. M. & Hearing, P. (1991). Interaction of nuclear factor EF-1A with the polyomavirus enhancer region. *J. Virol.* **65**, 1884-1892.
- Buckwold, V. E., Chen, M. & Ou, J.-H. (1997). Interaction of transcription factors RFX1 and MIBP1 with the gamma motif of the negative regulatory element of the hepatitis B virus core promoter. *Virology*, **227**, 515-518.
- Burbach, K. M., Poland, A. & Bradfield, C. A. (1992). Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *Proc. Natl Acad. Sci. USA*, **89**, 8185-8189.
- David, E., Garcia, A. D. & Hearing, P. (1995). Interaction of EF-C/RFX-1 with the inverted repeat of viral enhancer regions is required for transactivation. *J. Biol. Chem.* **270**, 8353-8360.
- Dikstein, R., Faktor, O., Ben-Levy, R. & Shaul, Y. (1990). Functional organization of the hepatitis B virus enhancer. *Mol. Cell Biol.* **10**, 3682-3689.
- Dikstein, R., Heffetz, D., Ben-Neriah, Y. & Shaul, Y. (1992). c-abl has a sequence-specific enhancer binding activity. *Cell*, **69**, 751-757.
- Dikstein, R., Agami, R., Heffetz, D. & Shaul, Y. (1996). p140/c-Abl that binds DNA is preferentially phosphorylated at tyrosine residues. *Proc. Natl Acad. Sci. USA*, **93**, 2387-2391.
- Emery, P., Durand, B., Mach, B. & Reith, W. (1996a). RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucl. Acids Res.* **24**, 803-807.
- Emery, P., Strubin, M., Hofmann, K., Bucher, P., Mach, B. & Reith, W. (1996b). A consensus motif in the RFX DNA binding domain and binding domain mutants with altered specificity. *Mol. Cell Biol.* **16**, 4486-4494.
- Fawell, S. E., Lees, J. A., White, R. & Parker, M. G. (1990). Characterization and colocalization of steroid binding and dimerization activities in the mouse estrogen receptor. *Cell*, **60**, 953-962.
- Greschik, H., Wurtz, J.-M., Hublitz, P., Kohler, F., Moras, D. & Schule, R. (1999). Characterization of the DNA-binding and dimerization properties of the nuclear orphan receptor germ cell nuclear factor. *Mol. Cell Biol.* **19**, 690-703.
- Herrero-Sanchez, C., Reith, W., Silacci, P. & Mach, B. (1992). The DNA-binding defect observed in major histocompatibility complex class II regulatory mutants concerns only one member of a family of complexes binding to the X boxes of class II promoters. *Mol. Cell Biol.* **12**, 4076-4083.
- Huang, M., Zhou, Z. & Elledge, S. J. (1998). The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. *Cell*, **94**, 595-605.
- Iwama, A., Pan, J., Zhang, P., Reith, W., Mach, B., Tenen, D. G. & Sun, Z. (1999). Dimeric RFX proteins contribute to the activity and lineage speci-

- ficity of the interleukin-5 receptor alpha promoter through activation and repression domains. *Mol. Cell Biol.* **19**, 3940-3950.
- Jones, N. (1990). Transcriptional regulation by dimerization: two sides to an incestuous relationship. *Cell*, **61**, 9-11.
- Katan, Y., Agami, R. & Shaul, Y. (1997). The transcriptional activation and repression domains of RFX1, a context-dependent regulator, can mutually neutralize their activities. *Nucl. Acids Res.* **25**, 3621-3628.
- Katan-Khaykovich, Y. & Shaul, Y. (1998). RFX1, a single DNA-binding protein with a split dimerization domain, generates alternative complexes. *J. Biol. Chem.* **273**, 24504-24512.
- Kouskoff, V., Mantovani, R. M., Candeias, S. M., Dorn, A., Staub, A., Lisowska-Grospierre, B., Griscelli, C., Benoist, C. O. & Mathis, D. J. (1991). NF-X, a transcription factor implicated in MHC class II gene regulation. *J. Immunol.* **146**, 3197-3204.
- Labrie, C., Lee, B. H. & Mathews, M. B. (1995). Transcription factors RFX1/EF-C and ATF-1 associate with the adenovirus E1A-responsive element of the human proliferating cell nuclear antigen promoter. *Nucl. Acids Res.* **23**, 3732-3741.
- Lee, B. H., Liu, M. & Mathews, M. B. (1998). Regulation of the human proliferating cell nuclear antigen promoter by the adenovirus E1A-associated protein p107. *J. Virol.* **72**, 1138-1145.
- Liu, M., Lee, B. H. & Mathews, M. B. (1999). Involvement of RFX1 protein in the regulation of the human proliferating cell nuclear antigen promoter. *J. Biol. Chem.* **274**, 15433-15439.
- Mangelsdorf, D. J. & Evans, R. M. (1995). The RXR heterodimers and orphan receptors. *Cell*, **83**, 841-850.
- Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M. & Chambon, P., *et al.* (1995). The nuclear receptor superfamily: the second decade. *Cell*, **83**, 835-839.
- Moras, D. & Gronemeyer, H. (1998). The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* **10**, 384-391.
- Ori, A., Zauberman, A., Doitsh, G., Paran, N., Oren, M. & Shaul, Y. (1998). p53 binds and represses the HBV enhancer: an adjacent enhancer element can reverse the transcription effect of p53. *EMBO J.* **17**, 544-553.
- Ostapchuk, P., Diffley, J. F., Bruder, J. T., Stillman, B., Levine, A. J. & Hearing, P. (1986). Interaction of a nuclear factor with the polyomavirus enhancer region. *Proc. Natl Acad. Sci. USA*, **83**, 8550-8554.
- Ostapchuk, P., Scheirle, G. & Hearing, P. (1989). Binding of nuclear factor EF-C to a functional domain of the hepatitis B virus enhancer region. *Mol. Cell Biol.* **9**, 2787-2797.
- Perlmann, T., Umesono, K., Rangarajan, P. N., Forman, B. M. & Evans, R. M. (1996). Two distinct dimerization interfaces differentially modulate target gene specificity of nuclear hormone receptors. *Mol. Endocrinol.* **10**, 958-966.
- Reinhold, W., Emens, L., Itkes, A., Blake, M., Ichinose, I. & Zajac-Kaye, M. (1995). The *myc* intron-binding polypeptide associates with RFX1 *in vivo* and binds to the major histocompatibility complex class II promoter region, to the hepatitis B virus enhancer, and to regulatory regions of several distinct viral genes. *Mol. Cell Biol.* **15**, 3041-3048.
- Reith, W., Barras, E., Satola, S., Kobr, M., Reinhart, D., Herrero-Sanchez, C. & Mach, B. (1989). Cloning of the major histocompatibility complex class II promoter binding protein affected in a hereditary defect in class II gene regulation. *Proc. Natl Acad. Sci. USA*, **86**, 4200-4204.
- Reith, W., Herrero-Sanchez, C., Kobr, M., Silacci, P., Berte, C., Barras, E., Fey, S. & Mach, B. (1990). MHC class II regulatory factor RFX has a novel DNA-binding domain and a functionally independent dimerization domain. *Genes Dev.* **4**, 1528-1540.
- Reith, W., Ucla, C., Barras, E., Gaud, A., Durand, B., Herrero-Sanchez, C., Kobr, M. & Mach, B. (1994). RFX1, a transactivator of hepatitis B virus enhancer I, belongs to a novel family of homodimeric and heterodimeric DNA-binding proteins. *Mol. Cell Biol.* **14**, 1230-1244.
- Safrany, G. & Perry, R. P. (1993). Transcription factor RFX1 helps control the promoter of the mouse ribosomal protein-encoding gene rpl30 by binding to its alpha element. *Gene*, **132**, 279-283.
- Safrany, G. & Perry, R. P. (1995). The relative contributions of various transcription factors to the overall promoter strength of the mouse ribosomal protein L30 gene. *Eur. J. Biochem.* **230**, 1066-1072.
- Schulman, I. G., Juguilon, H. & Evans, R. M. (1996). Activation and repression by nuclear hormone receptors: hormone modulates an equilibrium between active and repressive states. *Mol. Cell Biol.* **16**, 3807-3813.
- Siegrist, C. A. & Mach, B. (1993). Antisense oligonucleotides specific for regulatory factor RFX-1 inhibit inducible but not constitutive expression of all major histocompatibility complex class II genes. *Eur. J. Immunol.* **23**, 2903-2908.
- Siegrist, C. A., Durand, B., Emery, P., David, E., Hearing, P., Mach, B. & Reith, W. (1993). RFX1 is identical to enhancer factor C and functions as a transactivator of the hepatitis B virus enhancer. *Mol. Cell Biol.* **13**, 6375-6384.
- Steimle, V., Durand, B., Barras, E., Zufferey, M., Hadam, M. R., Mach, B. & Reith, W. (1995). A novel DNA-binding regulatory factor is mutated in primary MHC class II deficiency (bare lymphocyte syndrome). *Genes Dev.* **9**, 1021-1032.
- Tsang, S. Y., Nakanishi, M. & Peterlin, B. M. (1988). B-cell-specific and interferon-gamma-inducible regulation of the HLA-DR alpha gene. *Proc. Natl Acad. Sci. USA*, **85**, 8598-8602.
- Tsang, S. Y., Nakanishi, M. & Peterlin, B. M. (1990). Mutational analysis of the DRA promoter: *cis*-acting sequences and *trans*-acting factors. *Mol. Cell Biol.* **10**, 711-709.
- Vivat, V., Zechel, C., Wurtz, J. M., Bourguet, W., Kagechika, H., Umemiya, H., Shudo, K., Moras, D., Gronemeyer, H. & Chambon, P. (1997). A mutation mimicking ligand-induced conformational change yields a constitutive RXR that senses allosteric effects in heterodimers. *EMBO J.* **16**, 5697-5709.
- Weisinger, G., Remmers, E. F., Hearing, P. & Marcu, K. B. (1988). Multiple negative elements upstream of the murine *c-myc* gene share nuclear factor binding sites with SV40 and polyoma enhancers. *Oncogene*, **3**, 635-646.
- White, R., Sjöberg, M., Kalkhoven, E. & Parker, M. G. (1997). Ligand-independent activation of the oestrogen receptor by mutation of a conserved tyrosine. *EMBO J.* **16**, 1427-1435.

Wu, S. Y. & McLeod, M. (1995). The sak1 + gene of *Schizosaccharomyces pombe* encodes an RFX family DNA-binding protein that positively regulates cyclic AMP-dependent protein kinase-mediated exit from the mitotic cell cycle. *Mol. Cell Biol.* **15**, 1479-1488.

Zhang, X. Y., Asiedu, C. K., Supakar, P. C., Khan, R., Ehrlich, K. C. & Ehrlich, M. (1990). Binding sites in mammalian genes and viral gene regulatory regions recognized by methylated DNA-binding protein. *Nucl. Acids Res.* **18**, 6253-6260.

Edited by A. Klug

(Received 8 July 1999; received in revised form 27 September 1999; accepted 1 October 1999)