Molecular and Cellular Biology, Aug. 1996, p. 4486–4494 Vol. 16, No. 8
0270-7306/96/$04.00+0

A Consensus Motif in the RFX DNA Binding Domain and Binding Domain Mutants with Altered Specificity

PATRICK EMERY,[1] MICHEL STRUBIN,[1] KAY HOFMANN,[2] PHILIPP BUCHER,[2] BERNARD MACH,[1] and WALTER REITH[1]*

Department of Genetics and Microbiology, University of Geneva Medical School, CH-1211 Geneva 4,[1] Switzerland, and Institut Suisse de Recherche Experimentale sur le Cancer, CH-1066 Epalinges, Lausanne,[2] Switzerland

Received 25 March 1996/Returned for modification 19 April 1996/Accepted 9 May 1996

The RFX DNA binding domain is a novel motif that has been conserved in a growing number of dimeric DNA-binding proteins, having diverse regulatory functions, in eukaryotic organisms ranging from yeasts to humans. To characterize this novel motif, we have performed a detailed dissection of the site-specific DNA binding activity of RFX1, a prototypical member of the RFX family. First, we have performed a site selection procedure to define the consensus binding site of RFX1. Second, we have developed a new mutagenesis-selection procedure to derive a precise consensus motif, and to test the accuracy of a secondary structure prediction, for the RFX domain. Third, a modification of this procedure has allowed us to isolate altered-specificity RFX1 mutants. These results should facilitate the identification both of additional candidate genes controlled by RFX1 and of new members of the RFX family. Moreover, the altered-specificity RFX1 mutants represent valuable tools that will permit the function of RFX1 to be analyzed in vivo without interference from the ubiquitously expressed endogenous protein. Finally, the simplicity, efficiency, and versatility of the selection procedure we have developed make it of general value for the determination of consensus motifs, and for the isolation of mutants exhibiting altered functional properties, for large protein domains involved in protein-DNA as well as protein-protein interactions.

The majority of DNA-binding proteins can be grouped into families of proteins sharing a characteristic DNA binding motif (15). For many of these motifs, such as the well-known helix-turn-helix, basic-helix-loop-helix, basic-leucine zipper, and zinc finger domains, a great deal has been learned about their structures, target site specificities, and modes of interaction with DNA (1, 5, 27, 33). This is currently not the case for one of the most recently identified motifs, the RFX domain. Yet it is becoming increasingly clear that this RFX domain has been conserved throughout the eukaryotic kingdom in a growing number of proteins implicated in a diverse range of biological systems (6). The RFX family currently contains eight different members: five (RFX1 to RFX5) in humans and mice (17, 18, 21, 30), one in *Caenorhabditis elegans* (6), one (sak1) in *Schizosaccharomyces pombe* (34), and one in *Saccharomyces cerevisiae* (6). Functions of RFX genes include the regulation of major histocompatibility class II gene transcription by RFX5 (30), the regulation of genes such as c-*myc* (16) and the ribosomal protein gene *rpl30* (25, 26) by RFX1, expression of the highly pathogenic human hepatitis B virus by RFX1 (9, 28), and control of the mitotic cell cycle by sak1 (34). In view of the evolutionary conservation of the RFX motif and its recruitment into proteins having crucial regulatory functions, a detailed characterization of this novel DNA binding domain is warranted.

The human RFX1 protein was the first representative of the RFX family to be isolated (17, 18). It is expressed ubiquitously in all cell types and exists in the form of a mixed population of nuclear complexes representing RFX1 homodimers and RFX1-RFX2 or RFX1-RFX3 heterodimers (21) referred to collectively as EF-C (9, 14) or MDBP (10, 35, 37). RFX1 is in many respects a prototypical member of the family; in addition to the RFX DNA binding domain, RFX1 contains several other regions, including a domain responsible for dimerization (18), that have been highly conserved not only in RFX2 and RFX3 (21) but also in the RFX genes identified in *C. elegans*, *S. pombe*, and *S. cerevisiae* (6, 34).

Taking RFX1 as a model, we have studied three aspects of the site-specific binding activity of the RFX DNA binding motif. First, we have determined the consensus sequence for the optimal binding sites of RFX1. Second, we have developed an efficient and powerful in vivo selection technique performed with yeast cells to isolate mutated DNA binding domain sequences retaining site-specific DNA binding activity. This approach relies on the screening of large numbers of highly mutated DNA binding domains for sequences that remain functional. Alignment of these mutated, yet functional, sequences permits the identification of amino acid residues that are essential for the binding activity of the RFX DNA binding domain. The mutated sequences also allow us to test the validity of a secondary structure prediction for the DNA binding domain. Finally, we have used this selection technique to generate altered-specificity mutants of RFX1, thereby identifying amino acid residues that are crucial for target site specificity. In addition to enhancing our understanding of the RFX DNA binding domain, these results should facilitate the identification of candidate target genes controlled by RFX1 and of additional proteins containing the RFX DNA binding motif. Moreover, the altered-specificity mutants represent valuable tools that will permit the development of systems in which the function of individual RFX proteins can be assayed in vivo independently of the endogenously expressed RFX factors.

MATERIALS AND METHODS

Selection of RFX1 binding sites. Two randomized oligonucleotide pools, N20 (CTAGAATTCGGCTCCAGGT-N_{20}-GACTGAGACGGATCCTGA) and N30 (CTAGAATTCGGCCTCATCTC-N_{30}-TGTCAGAGACGGATCCTGA), were

* Corresponding author. Mailing address: Department of Genetics and Microbiology, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva 4, Switzerland. Phone: (41 22) 702 56 66. Fax: (41 22) 702 57 02. Electronic mail address: reith@cmu.unige.ch.

synthesized and purified on 12% polyacrylamide denaturing gels. These oligonucleotide pools consist, respectively, of a 20- and 30-nucleotide random sequence embedded in a constant sequence. They were made double stranded by using the Klenow fragment and oligonucleotide primers complementary to their 3′ ends, purified on 6% polyacrylamide gels, and labelled with T4 polynucleotide kinase and [γ-^{32}P]ATP. Cycles of selection for RFX1 binding sites were then performed in three steps as follows. (i) The double-stranded oligonucleotides were incubated for 30 min at 20°C with in vitro-translated human RFX1 in 12 mM HEPES (*N*-2-hydroxyethylpiperazine-*N′*-2-ethanesulfonic acid) (pH 7.9)–12% glycerol–60 mM KCl–0.12 mM EDTA–0.3 mM dithiothreitol–0.3 mM phenylmethylsulfonyl fluoride–5 mM $MgCl_2$–12.5 μg of single-stranded *Escherichia coli* DNA per ml–12.5 μg of poly(dI-dC) · poly(dI-dC) per ml. During the first cycle, 15 ng of labelled oligonucleotide was used. During subsequent cycles, the amount of labelled oligonucleotide was progressively diminished to 3 ng. (ii) Binding reaction mixtures were fractionated by gel electrophoresis, and oligonucleotides bound by RFX1 homodimers were eluted and purified as described previously (12). (iii) Oligonucleotides selected in this way were then amplified by PCR with primers complementary to the constant regions flanking the randomized sequence. Three (N20) or four (N30) cycles were sufficient to select high-affinity binding sites. The selected oligonucleotides were then digested with *Eco*RI and *Bam*HI, subcloned in pBluescript (Stratagene), and sequenced. Half-lives for RFX1-DNA complexes were determined for certain binding sites by electrophoretic mobility shift assay (EMSA) experiments as described previously (19, 20), except that in vitro-translated RFX1 was used.

Yeast strains. Strains containing *his3* alleles under the control of a minimal promoter and various upstream regulatory elements were derived from KY320 (2). Py-His3 contains a high-affinity RFX1 binding site from the enhancer of polyomavirus (Py site) (GTTGCCTAGCAAC) (14) 50 bp upstream of the transcription initiation site. In the ATPyMutA, ATPyMutC, and ATPyMutT strains, the Py site is mutated to GTTGTCTAACAAC, GTTGGCTACCAAC, and GTTGACTATCAAC, respectively, and the natural dA-dT element of the *his3* gene is maintained upstream of the mutated Py sites.

RFX1 expression vectors. RFX1 expression vectors were constructed in pRS316 (29), which contains the *ura3* gene as a selectable marker. pNVRFX1 contains the entire RFX1-coding region fused at its 5′ end to sequences coding for the nuclear localization signal (NLS) of simian virus 40 large T antigen and the transcription activation domain (amino acids 413 to 490) of VP16 (32). This fusion gene is placed between the regulatory sequences of yeast TBP (3).

pNVRFX1HE, pNVRFX1MHEX1, and pNVRFX1MHEX2 were derived from pNVRFX1 by the introduction of unique restriction sites within or close to the DNA binding domain of RFX1. These restriction sites were created by the introduction of silent point mutations by using a PCR-mediated site-directed mutagenesis technique. In pNVRFX1HE, the codons for K-482–L-483 and G-500–N-501–S-502 of RFX1 are mutated, respectively, to *Hin*dIII and *Eco*RI sites. pNVRFX1MHEX1 was derived from pNVRFX1HE by adding additional *Mlu*I and *Xho*I sites at the codons for T-432–R-433 and A-513–S-514–S-515, respectively. pNVRFX1MHEX2 differs from pNVRFX1MHEX1 only by the position of the *Hin*dIII site, at the codons for K-471–L-472.

Construction of mutated RFX1 libraries. To construct libraries of RFX1 expression plasmids containing mutated DNA binding domains, double-stranded oligonucleotides containing degenerate regions were inserted between the unique restriction sites of pNVRFX1HE, pNVRFX1MHEX1, and pNVRFX1 MHEX2. Prior to insertion of these degenerate oligonucleotides, the corresponding regions were first replaced with unrelated stuffer DNA to avoid high background levels of wild-type plasmids. The degenerate oligonucleotides were designed such that the frequencies of wild-type sequences were less than $1/10^6$. To achieve this, the oligonucleotides were synthesized such that at each degenerate position the frequency of incorporation was 67% for the wild-type nucleotide and 11% for each of the three remaining nucleotides. For the RND1 and RND2 libraries, oligonucleotides degenerate between amino acids 438 and 452 and between amino acids 453 and 468 were cloned between the *Mlu*I and *Hin*dIII sites of pNVRFX1MHEX2. For the RND3 library, an oligonucleotide degenerate between amino acids 469 and 483 was cloned between the *Mlu*I and *Eco*RI sites of pNVRFX1MHEX2. For the RND4 library, an oligonucleotide degenerate between amino acids 484 and 499 was cloned between the *Eco*RI and *Hin*dIII sites of pNVRFX1HE. For the RND5 library, an oligonucleotide degenerate between amino acids 500 and 513 was cloned between the *Xho*I and *Hin*dIII sites of pNVRFX1MHEX1. For each library, a complexity of at least 2×10^6 to 3×10^6 individual clones was obtained.

Screening of mutated RFX1 libraries. Yeast cells transformed with mutated RFX1 libraries were first grown at 30°C in medium lacking uracil (URA$^-$ medium) until an optical density at 600 nm of 1.5 to 2 was attained. The cells were then diluted to an optical density at 600 nm of 0.02 in minimal medium lacking histidine (HIS$^-$ medium) containing 50 mM aminotriazole (AT) and grown until an optical density at 600 nm of 1 to 2 was attained. Plasmids were then rescued from the bulk culture and transformed into *E. coli* as described previously (22). Individual plasmids were then isolated and sequenced in the region encoding the DNA binding domain.

Isolation of altered-specificity mutants. To isolate altered-specificity mutants, the five mutated RFX1 libraries were transformed, either individually or as a pool, into yeast strains carrying the ATPyMutA, ATPyMutC, or ATPyMutT *his3* allele. For each strain, 50,000,000 transformed cells were plated on HIS$^-$ plates containing 50 mM AT. Clones able to grow were isolated only with the ATPyMutA strain. Plasmids were rescued from these clones and sequenced in the region encoding the DNA binding domain.

Transcription assays. The capacities of isolated plasmids, or pools of plasmids, to encode functional RFX1 proteins were assayed by testing their abilities to transactivate the *his3* reporter gene in yeast cells, thus permitting growth on minimal HIS$^-$ plates containing AT. Py-His3 cells transformed with pNVRFX1 and its derivatives were tested for their abilities to grow on minimal HIS$^-$ plates containing various amounts of AT (data not shown). Py-His3 cells transformed with individual mutated plasmids isolated by the selection procedure were tested on HIS$^-$ plates containing 50 mM AT (see Fig. 4). Py-His3 cells transformed with pools of 100 to 200 plasmids were replica plated on URA$^-$ plates and on HIS$^-$ plates containing 50 mM AT (see Fig. 3). Altered-specificity mutants were tested for their abilities to allow growth of Py-His3, ATPyMutA, ATPyMutC, and ATPyMutT cells on minimal plates containing 50 mM AT. In all cases, plated cells were allowed to grow for 3 to 4 days at 30°C.

EMSA. The altered-specificity mutants were transferred into the pT7-7 vector (31). Wild-type RFX1 and the altered-specificity mutants were then transcribed and translated in vitro as described previously (18). Binding activities of the in vitro translation products were then analyzed by EMSA as described previously (21) with double-stranded oligonucleotides containing the wild-type Py site and the mutated PyMutA, PyMutC, and PyMutT sites. The sequence of the pBR322 oligonucleotide used as a competitor is GATCGTCACGGCGATC (11). In the methylated version of this oligonucleotide, the underlined CG dinucleotides contain 5-methylcytosine on both strands. To ensure that equal amounts of in vitro-translated proteins were analyzed in all samples, the ^{35}S-labelled translation products were fractionated by electrophoresis on sodium dodecyl sulfate–10% polyacrylamide gels and the levels of protein synthesized were quantified by PhosphorImager analysis.

Secondary structure prediction. Secondary structure predictions from multiple sequence alignments were obtained from the PredictProtein server at the EMBL (23). In the submitted alignment of naturally occurring RFX domains, we duplicated the sequences of the nonmammalian members and RFX5 in order to decrease the weight of the highly homologous RFX1 to RFX4 subfamily, from which we included only human RFX1, RFX2, and RFX4. In order to derive a secondary structure prediction based on the mutational data, we generated an alignment of 10 artificial sequences containing at each position the wild-type residue and all substitutions observed in the selected sequences at approximately equal frequencies.

RESULTS

The consensus binding site for RFX1. In order to determine the consensus sequence recognized by RFX1, we performed a site selection procedure with oligonucleotides containing a stretch of random sequence (see Materials and Methods). Two different oligonucleotides (N20 and N30), containing, respectively, 20 and 30 randomized nucleotides, were used. Successive rounds of selection for RFX1 binding sites were performed (three rounds for N20 and four rounds for N30) until the selected population consisted essentially only of high-affinity binding sites as judged by EMSA. In EMSA, the final selected population was bound by RFX1 as efficiently as a single oligonucleotide (data not shown) containing a sequence (Py) known to be a high-affinity binding site present in the enhancer of polyomavirus (14). This selected population was subcloned, and individual clones were sequenced. An alignment of these sequences permits the establishment of the consensus RFX1 binding site (Fig. 1). Results with the N20 and N30 oligonucleotides were essentially identical. To avoid introducing a bias, only sequences in which the binding site does not infringe on the flanking constant sequences of the oligonucleotides were taken into consideration. The consensus binding site derived from the optimal alignment of the selected sequences consists of an imperfect palindrome (Fig. 1A). In all cases, at least one of the two half-sites closely matches the RGYAAC consensus motif. The second half-site is often considerably more degenerate and has a G/NGYNAC consensus motif. The spacing between the two half-sites ranges from 0 to 3 bp, with a strong preference for 1 or 2 bp (Fig. 1B). In the case of 1-bp spacers, there is a strong preference for either T or A, while in the case of 2-bp spacers, the consensus is AT (Fig. 1B).

To confirm that the consensus motif was indeed established

A

			LEFT HALF-SITE						SPACER 0-3 N	RIGHT HALF-SITE							
A	5	10	6	3	11	13	3	4		11	0	3	30	32	0	11	11
C	5	4	4	2	10	1	25	14		3	0	16	0	0	32	9	3
G	6	8	22	7	2	17	1	7		17	32	3	1	0	0	7	5
T	8	6	0	20	9	1	3	7		1	0	10	1	0	0	4	10
TOTAL	24	28	32	32	32	32	32	32		32	32	32	32	32	32	31	29
CONSENSUS	N	N	G	T	N	R	C	C/n		R	G	Y	A	A	C	N	N

B

LENGTH OF SPACER	
0	1
1	12
2	16
3	3
TOTAL	32
CONSENSUS	1/2N

1N SPACING	
A	4
C	2
G	0
T	6
TOTAL	12
CONSENSUS	T/A

2N SPACING		
A	9	4
C	3	2
G	2	0
T	2	10
TOTAL	16	16
CONSENSUS	A	T

C

GTTGCT	A	GGCAAC	(Py)	75 min.
GTTACC	A C	AGTAAC		70 min.
GTGGTG	A T	GGCAAC		40 min.
GAAACC	A	AGGAAC		30 min.
GGTGCG	A	GGGAAC		30 min.

FIG. 1. The consensus binding site for RFX1 is an imperfect palindrome with variable spacing between the two half-sites. To obtain the consensus sequence, 32 sequences obtained by the site selection procedure were analyzed. (A) Sequences are aligned with the most strongly conserved half site (5′-RGYAAC-3′) at the right. The left half-site is more degenerate (5′-G/NGYNAC-3′ on the complementary strand). The spacing between the two half-sites ranges from 0 to 3 bp. The number of sequences containing a given nucleotide is indicated for each position of the two half-sites and for the two positions adjacent to each half-site. To avoid introducing a bias, nucleotides flanking the palindrome were ignored when they were situated within the constant regions of the oligonucleotides used for the site selection procedure. (B) Numbers of sequences containing spacer lengths of 0, 1, 2 and 3 bp and numbers of sequences containing a given nucleotide in spacers of 1 and 2 bp. (C) Half-lives for the RFX1-DNA complexes formed with five selected binding sites exhibiting a variable degree of divergence from the consensus sequence are compared with the half-life observed with the high-affinity Py binding site.

only on the basis of high-affinity binding sites, the avidity of RFX1 for several of the individual selected sequences was analyzed by means of EMSA experiments. The approximate half-lives of the RFX1-DNA complexes obtained with several representative sequences exhibiting various degrees of divergence from the consensus binding site are indicated in Fig. 1C. Divergence from the consensus indeed leads to a reduction in the stability of the RFX1-DNA complex. However, even for the most divergent sites this reduction is only marginal (2.5-fold), indicating that a significant amount of divergence from the consensus sequence is permitted.

The consensus DNA binding motif of RFX1. To identify amino acids within the DNA binding domain that are crucial for binding of RFX1, we developed an in vivo selection procedure for isolating DNA binding domain sequences that are heavily mutated yet remain functional (Fig. 2). Briefly, we constructed RFX1 expression libraries in which segments of the DNA binding domain of an NLS-VP16-RFX1 fusion protein were mutated randomly and extensively (see Materials and Methods). These libraries were transformed into a yeast strain (Py-His3) carrying a *his3* allele under the control of the high-affinity Py binding site. Yeast cells carrying plasmids encoding functional NLS-VP16-RFX1 fusion proteins were then selected for by growth in minimal medium containing AT, a competitive inhibitor of the HIS3 enzyme. Under these conditions, only cells in which expression of the *his3* gene is activated by binding of the NLS-VP16-RFX1 fusion protein can grow (data not shown). This procedure is highly efficient. Prior to selection, fewer than 5% of yeast cells transformed with the libraries can grow when plated in the presence of AT, indicating that the DNA binding domain has been extensively mutated and that the majority of plasmids encode nonfunctional NLS-VP16-RFX1 fusion proteins (Fig. 3). Following selection by growth in medium containing AT, on the other hand, greater than 95% of the plasmids rescued from the surviving yeast cells can confer growth on plates containing AT when reintroduced into the parental strain, indicating that the vast majority encode NLS-VP16-RFX1 fusion proteins retaining binding activity (Fig. 3).

Five libraries (RND1 to RND5) containing random mutations in five adjacent segments of the DNA binding domain were constructed (see Materials and Methods) and subjected to the selection procedure. Following selection by growth in the presence of AT, plasmids were rescued from the surviving cells, and individual clones were sequenced in the region encoding the DNA binding domain. For each library, over 20 independent clones were sequenced (Fig. 4). Although over 95% of the sequences obtained should encode a functional DNA binding domain (Fig. 3), we nevertheless verified this for several of the most heavily mutated sequences, particularly those containing mutations at positions that are strongly conserved in all the other sequences, by reconfirming their ability

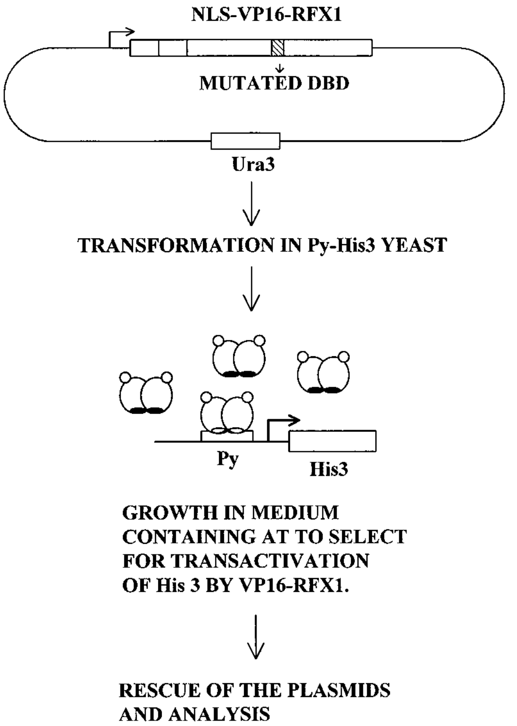

NLS-VP16-RFX1
MUTATED DBD
Ura3
TRANSFORMATION IN Py-His3 YEAST
Py
His3
GROWTH IN MEDIUM CONTAINING AT TO SELECT FOR TRANSACTIVATION OF His 3 BY VP16-RFX1.
RESCUE OF THE PLASMIDS AND ANALYSIS

FIG. 2. Selection procedure to isolate RFX1 DNA binding domains that are heavily mutated but remain functional. Expression libraries directing the synthesis of NLS-VP16-RFX1 fusion proteins containing heavily mutated DNA binding domains (DBD) are transformed into a yeast strain (Py-His3) harboring a *his3* gene placed under the control of a high-affinity RFX1 binding site (Py). If the mutated DBD remains functional, the NLS-VP16-RFX1 protein will bind to the Py site and activate expression of the *his3* gene. Yeast cells expressing the *his3* gene are selected for by growth in minimal HIS$^-$ medium containing 50 mM AT.

to confer growth on plates containing AT. Of the 32 individual sequences tested, 27 were found to encode functional DNA binding domains (Fig. 4). Only five were found to encode defective DNA binding domains, and all of these contained one or more mutations at residues that are invariant in all of the other sequences (data not shown). These nonfunctional sequences were eliminated from our analysis.

Alignment of the sequences obtained allows us to identify a consensus motif composed of amino acids that are essential for the binding activity of the DNA binding domain of RFX1 (Fig. 4 and 5). The consensus appears to consist of two conserved motifs (amino acids 438 to 467 and 476 to 510) separated by an 8-amino-acid segment exhibiting strong degeneracy (amino acids 468 to 475). The C-terminal motif is considerably more conserved (89%) than the N-terminal one (57%). Prominent among the critical amino acids are hydrophobic residues and aromatic residues distributed throughout the domain and a cluster of basic residues restricted to the C-terminal motif. Additional characteristic amino acids include an invariant cysteine residue and five invariant or highly conserved glycine residues.

All 48 critical amino acids defined as described above for the DNA binding domain of RFX1 are conserved in several other members of the RFX family (Fig. 5). In fact, 30 of these positions are either invariant or replaced by conservative changes in all currently known RFX genes, even though several of these (RFX4, RFX5, and ScRFX) are known to have target site specificities that differ from that of RFX1 (reference 4 and unpublished data). On the other hand, amino acids that are dispensable for binding of RFX1 have diverged enormously in the other family members. Consequently, the residues determined to be crucial for the DNA binding domain of RFX1 define a consensus motif that is to a large extent valid for the entire RFX family.

Protein secondary structures can often be predicted with a high degree of accuracy when they are based on adequate multiple sequence alignments (24). We therefore derived secondary structure predictions from an alignment of all of the known naturally occurring RFX DNA binding domain sequences (see Materials and Methods). The N-terminal moiety of the RFX domain is predicted to contain two α helices (I and II), while the C-terminal moiety is predicted to contain a third α helix (III) followed by a β strand (Fig. 5). This secondary structure prediction was tested for its compatibility with the entire panel of mutated sequences isolated by the selection procedure (see Materials and Methods). This analysis provides independent support for helices I and III and the β strand but does not confirm the presence of helix II.

Isolation of altered-specificity mutants. The efficiency of the selection technique used to determine the DNA binding domain motif suggested that it should be well suited to isolate altered-specificity mutants of RFX1. We therefore generated yeast strains in which the *his3* gene was placed under the control of Py sites in which the second residue of each RGC AAC half site was mutated to either A (RACAAC), T (RTC AAC), or C (RCCAAC). These strains are unable to grow on

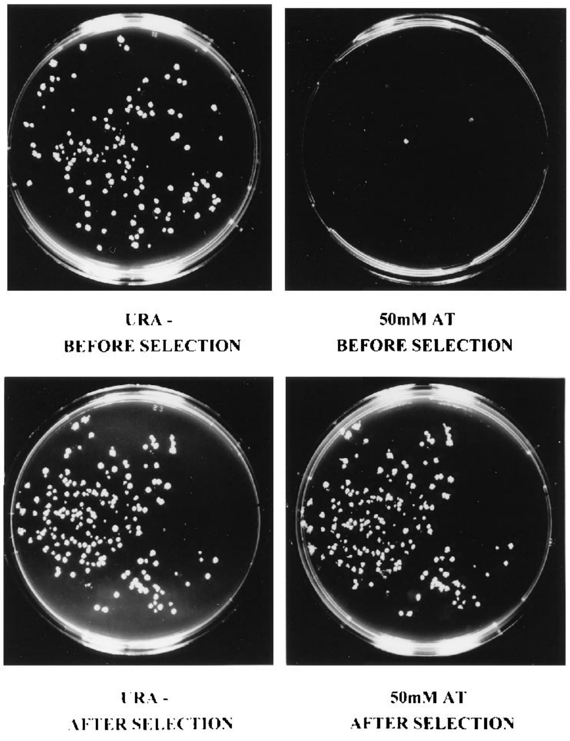
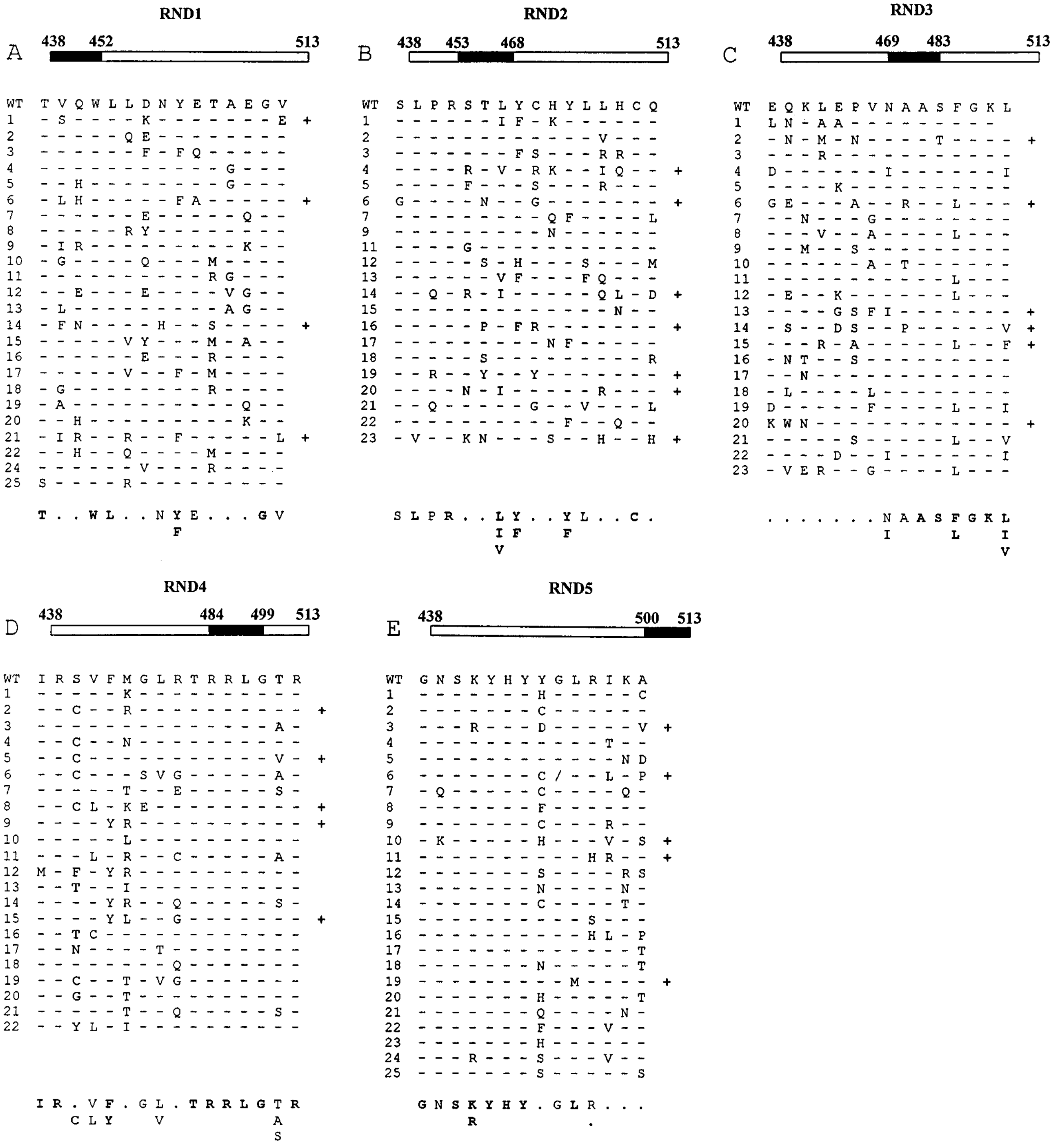


FIG. 3. Efficiency of the selection procedure. A pool of 100 to 200 plasmids from the RND1 library prior to selection and a pool of 100 to 200 plasmids rescued after the selection procedure were transformed into Py-His3 yeast cells. Transformed cells were replica plated on URA$^-$ medium to select for maintenance of the plasmids and on medium containing 50 mM AT to select for plasmids encoding functional NLS-VP16-RFX1 fusion proteins. Prior to selection, fewer than 5% of the plasmids in the library permit growth in the presence of AT. After the selection procedure, over 95% of the rescued plasmids permit growth in the presence of AT. Similar results were obtained with the four other libraries (data not shown).



```
WT  T V Q W L L D N Y E T A E G V
1   - S - - - - K - - - - - - - E   +
2   - - - - - Q E - - - - - - - -
3   - - - - - - F - F Q - - - - -
4   - - - - - - - - - - - G - - -
5   - - H - - - - - - - - G - - -
6   - L H - - - - - F A - - - - -   +
7   - - - - - - E - - - - - Q - -
8   - - - - - R Y - - - - - - - -
9   - I R - - - - - - - - - K - -
10  - G - - - - Q - - - M - - - -
11  - - - - - - - - - - R G - - -
12  - - E - - - E - - - - V G - -
13  - L - - - - - - - - - A G - -
14  - F N - - - - H - - S - - - -   +
15  - - - - - V Y - - - M - A - -
16  - - - - - - E - - - R - - - -
17  - - - - - V - - F - M - - - -
18  - G - - - - - - - - R - - - -
19  - A - - - - - - - - - - Q - -
20  - - H - - - - - - - - - K - -
21  - I R - - R - - F - - - - - L   +
22  - - H - - Q - - - - M - - - -
24  - - - - - - V - - - R - - - -
25  S - - - - R - - - - - - - - -

    T . . W L . . N Y E . . . G V
                    F
```

```
WT  S L P R S T L Y C H Y L L H C Q
1   - - - - - - I F - K - - - - - -
2   - - - - - - - - - - - - V - - -
3   - - - - - - - F S - - - R R - -
4   - - - - R - V - R K - - I Q - -   +
5   - - - - F - - - S - - - R - - -
6   G - - - - N - - G - - - - - - -   +
7   - - - - - - - - - Q F - - - - L
9   - - - - - - - - - N - - - - - -
11  - - - - G - - - - - - - - - - -
12  - - - - - S - H - - - S - - - M
13  - - - - - - V F - - - F Q - - -
14  - - Q - R - I - - - - - Q L - D   +
15  - - - - - - - - - - - - - N - -
16  - - - - - P - F R - - - - - - -   +
17  - - - - - - - - - N F - - - - -
18  - - - - - S - - - - - - - - - R
19  - - R - - Y - - Y - - - - - - -   +
20  - - - - N - I - - - - - R - - -   +
21  - - Q - - - - - G - - V - - - L
22  - - - - - - - - - - F - - Q - -
23  - V - - K N - - - S - - H - - H   +

    S L P R . . L Y . . Y L . . C .
                I F     F
                V
```

```
WT  E Q K L E P V N A A S F G K L
1   L N - A A - - - - - - - - -
2   - N - M - N - - - - T - - - -   +
3   - - - R - - - - - - - - - - -
4   D - - - - - - I - - - - - - I
5   - - - - K - - - - - - - - - -
6   G E - - - A - - R - - L - - -   +
7   - - N - - - G - - - - - - - -
8   - - - V - - A - - - - L - - -
9   - - M - - S - - - - - - - - -
10  - - - - - - A - T - - - - - -
11  - - - - - - - - - - - L - - -
12  - E - - K - - - - - - L - - -
13  - - - - G S F I - - - - - - -   +
14  - S - - D S - - P - - - - - V   +
15  - - - R - A - - - - - L - - F   +
16  - N T - - S - - - - - - - - -
17  - - N - - - - - - - - - - - -
18  - L - - - - L - - - - - - - -
19  D - - - - - F - - - - L - - I
20  K W N - - - - - - - - - - - -   +
21  - - - - - S - - - - - L - - V
22  - - - - D - - I - - - - - - I
23  - V E R - - G - - - - L - - -

    . . . . . . . N A A S F G K L
                  I       L     I
                                V
```

```
WT  I R S V F M G L R T R R L G T R
1   - - - - - K - - - - - - - - - -
2   - - C - - R - - - - - - - - - -   +
3   - - - - - - - - - - - - - - A -
4   - - C - - N - - - - - - - - - -
5   - - C - - - - - - - - - - - V -   +
6   - - C - - - S V G - - - - - A -
7   - - - - - T - - E - - - - - S -
8   - - C L - K E - - - - - - - - -   +
9   - - - - Y R - - - - - - - - - -   +
10  - - - - - L - - - - - - - - - -
11  - - - L - R - - C - - - - - A -
12  M - F - Y R - - - - - - - - - -
13  - - T - - I - - - - - - - - - -
14  - - - - Y R - - Q - - - - - S -
15  - - - - Y L - - G - - - - - - -   +
16  - - T C - - - - - - - - - - - -
17  - - N - - - - T - - - - - - - -
18  - - - - - - - - Q - - - - - - -
19  - - C - - T - V G - - - - - - -
20  - - G - - T - - - - - - - - - -
21  - - - - - T - - Q - - - - - S -
22  - - Y L - I - - - - - - - - - -

    I R . V F . G L . T R R L G T R
        C L Y     V             A
                                S
```

```
WT  G N S K Y H Y Y G L R I K A
1   - - - - - - - H - - - - - C
2   - - - - - - - C - - - - - -
3   - - - R - - - D - - - - - V   +
4   - - - - - - - - - - - T - -
5   - - - - - - - - - - - - N D
6   - - - - - - - C / - - L - P   +
7   - Q - - - - - C - - - - Q -
8   - - - - - - - F - - - - - -
9   - - - - - - - C - - - R - -
10  - K - - - - - H - - - V - S   +
11  - - - - - - - - - - H R - -   +
12  - - - - - - - S - - - - R S
13  - - - - - - - N - - - - N -
14  - - - - - - - C - - - - T -
15  - - - - - - - - - - S - - -
16  - - - - - - - - - - H L - P
17  - - - - - - - - - - - - - T
18  - - - - - - - N - - - - - T
19  - - - - - - - - - M - - - -   +
20  - - - - - - - H - - - - - T
21  - - - - - - - Q - - - - N -
22  - - - - - - - F - - - V - -
23  - - - - - - - H - - - - - -
24  - - - R - - - S - - - V - -
25  - - - - - - - S - - - - - S

    G N S K Y H Y . G L R . . .
          R             .
```

FIG. 4. Establishment of a consensus sequence for the DNA binding domain of RFX1. Yeast cells were transformed with the mutated RFX1 libraries and subjected to the selection procedure. Plasmids were then rescued from the surviving yeast cells, and their DNA binding domain regions were sequenced. The sequences shown are derived from the RND1 (A), RND2 (B), RND3 (C), RND4 (D), and RND5 (E) libraries. Only amino acids differing from those of the wild-type (WT) RFX1 sequence are indicated. A deleted amino acid in one sequence (RND5, sequence 6) is indicated (/). The coordinates of the mutated segments (solid boxes) within the DNA binding domain of RFX1 (open boxes) are indicated above the sequences; the coordinates refer to the complete amino acid sequence of RFX1 (18). The consensus sequence is given below the sequences; residues that are invariant or replaced by similar amino acids in all sequences (boldface) and residues that are found preferentially in the majority of sequences (lightface) are indicated. Dots indicate highly divergent positions. +, sequences for which binding activity has been reconfirmed (see text).

minimal plates containing low concentrations of AT, even in the presence of wild-type RFX1 (data not shown), confirming the fact that RFX1 does not bind to the mutated Py sites (Fig. 6). The five mutated RFX1 libraries were transformed, either singly or as a pool, into each strain, and the transformed cells were plated on minimal plates containing AT. This led to the isolation of four altered-specificity mutants of RFX1 (AltSp1 to AltSp4) conferring AT-resistant growth to the strain carry-

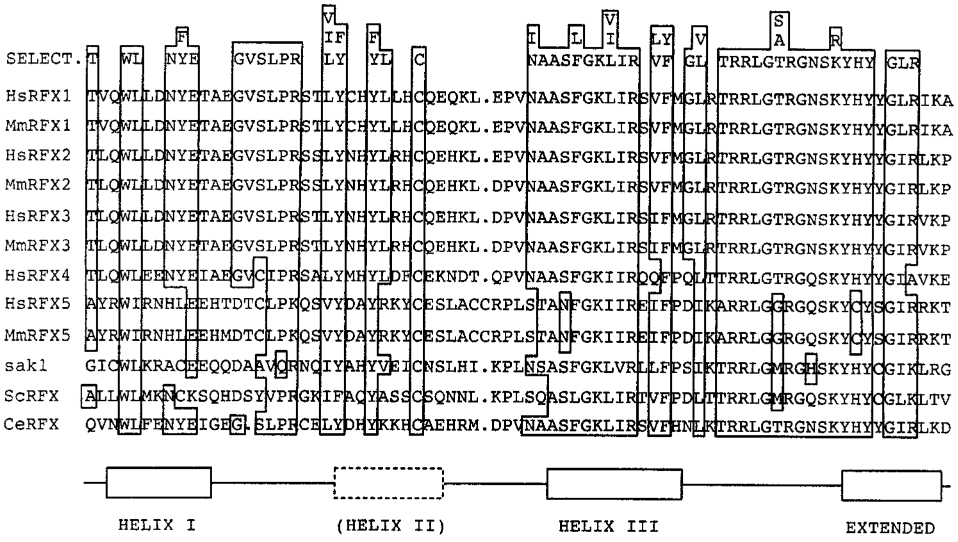

F
V IF
F
I L V I
LY
V
S A
R
SELECT. T WL NYE GVSLPR LY YL C NAASFGKLIR VF GL TRRLGTRGNSKYHY GLR
HsRFX1 TVQWLLDNYETAEGVSLPRSTLYCHYLLHCQEQKL.EPVNAASFGKLIRSVFMGLRTRRLGTRGNSKYHYYGLRIKA
MmRFX1 TVQWLLDNYETAEGVSLPRSTLYCHYLLHCQEQKL.EPVNAASFGKLIRSVFMGLRTRRLGTRGNSKYHYYGLRIKA
HsRFX2 TLQWLLDNYETAEGVSLPRSSLYNHYLRHCQEHKL.EPVNAASFGKLIRSVFMGLRTRRLGTRGNSKYHYYGIRLKP
MmRFX2 TLQWLLDNYETAEGVSLPRSSLYNHYLRHCQEHKL.DPVNAASFGKLIRSVFMGLRTRRLGTRGNSKYHYYGIRLKP
HsRFX3 TLQWLLDNYETAEGVSLPRSTLYNHYLRHCQEHKL.DPVNAASFGKLIRSIFMGLRTRRLGTRGNSKYHYYGIRVKP
MmRFX3 TLQWLLDNYETAEGVSLPRSTLYNHYLRHCQEHKL.DPVNAASFGKLIRSIFMGLRTRRLGTRGNSKYHYYGIRVKP
HsRFX4 TLQWLEENYEIAEGVCIPRSALYMHYLDFCEKNDT.QPVNAASFGKIIRQQFPQLTTRRLGTRGQSKYHYYGIAVKE
HsRFX5 AYRWIRNHLEEHTDTCLPKQSVYDAYRKYCESLACCRPLSTANFGKIIREIFPDIKARRLGGRGQSKYCYSGIRRKT
MmRFX5 AYRWIRNHLEEHMDTCLPKQSVYDAYRKYCESLACCRPLSTANFGKIIREIFPDIKARRLGGRGQSKYCYSGIRRKT
sak1 GICWLKRACEEQQDAAVQRNQIYAHYVEICNSLHI.KPLNSASFGKLVRLLFPSIKTRRLGMRGHSKYHYCGIKLRG
ScRFX ALLWLMKNCKSQHDSYVPRGKIFAQYASSCSQNNL.KPLSQASLGKLIRTVFPDLTTRRLGMRGQSKYHYCGLKLTV
CeRFX QVNWLFENYEIGEG.SLPRCELYDHYKKHCAEHRM.DPVNAASFGKLIRSVFHNLKTRRLGTRGNSKYHYYGIRLKD
HELIX I
(HELIX II)
HELIX III
EXTENDED

FIG. 5. Alignment between the consensus sequence obtained by the selection procedure and the DNA binding domain sequences of all known RFX proteins. For RFX1 to RFX5, the human (Hs) and mouse (Mm) sequences are included. CeRFX, sak1, and ScRFX are from *C. elegans*, *S. pombe*, and *S. cerevisiae*, respectively. Amino acids that are identical or similar to the selected sequence (SELECT.) are boxed. The secondary structure prediction is indicated below the sequences. Positions of the predicted α helices and β strand (EXTENDED) are indicated by boxes; helix II is shown as a dashed box because the mutational analysis did not provide support for its presence.

ing the PyMutA allele of *his3* (Fig. 6). All four altered-specificity mutants were derived from the RND5 library. The ability of the altered-specificity mutants to bind to the PyMutA site was verified in vivo by retransformation into the strain carrying the PyMutA *his3* allele and in vitro by EMSA (Fig. 6). In contrast to wild-type RFX1, the altered-specificity mutants bind efficiently to the PyMutA site. On the other hand, they cannot bind to the PyMutC and PyMutT sites and have thus acquired specificity for an A at the second position of each half-site.

The altered-specificity mutants have retained the ability to recognize the wild-type Py site (Fig. 6). Another feature char-

		Binding to			
		PyWt	PyMutC	PymutT	PyMutA
wild type	E Q K L E P V N A A S F G K L	++++	-	-	-
AltSp1	- - - V D - F I R - Y - - - -	+++	-	-	+++
AltSp2	- E - - - - L - R - C - - R -	+++	-	-	++
AltSp3	G - - - - - A - R P - - - - -	+++	-	-	+++
AltSp4	D K R Q - - - - R P C - - - -	++	-	-	++
AltSpR	- - - - - - - - R - - - - - -	++	-	-	-
AltSpRP	- - - - - - - - R P - - - - -	++	-	-	++

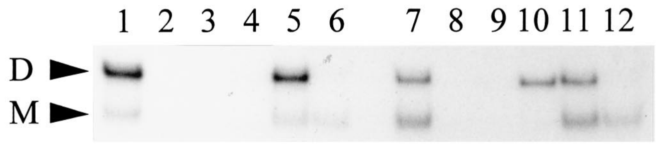


FIG. 6. Isolation of RFX1 altered-specificity mutants. (Top) Sequences and binding-site specificities of wild-type RFX1 and the altered-specificity mutants. Only amino acids that differ from those in the sequence of wild-type RFX1 are indicated. AltSp1 to AltSp4 were isolated from the RND5 library by the selection procedure. AltSpR and AltSpRP were constructed such that they contain only the R-477 and R-477–P-478 mutations, respectively. Affinities for the wild-type (PyWt) and mutated (PyMutC, PyMutT, and PyMutA) binding sites were determined by EMSA experiments, and their approximate levels are shown at the right. (Bottom) EMSA performed with wild-type RFX1 (lanes 1 to 6) and with the altered-specificity mutant AltSp1 (lanes 7 to 12). Similar results were obtained with the other altered-specificity mutants (data not shown). Oligonucleotides used for the binding experiments contained the PyWt (lanes 1, 5 to 7, 11, and 12), PyMutC (lanes 2 and 8), PyMutT (lanes 3 and 9), and PyMutA (lanes 4 and 10) binding sites. No competitor was added in lanes 1 to 4 and 7 to 10. Unlabelled competitor oligonucleotides included during the binding reactions contained the nonmethylated (lanes 5 and 11) or methylated (lanes 6 and 12) versions of a sequence from pBR322 (see Materials and Methods). Positions of protein-DNA complexes due to binding of RFX1 monomers (M) and dimers (D) are indicated (see Discussion).

acteristic of RFX1 has also been retained, namely, the ability to distinguish between the methylated and nonmethylated versions of certain target sequences containing CpG dinucleotides (11). Both wild-type RFX1 and the altered-specificity mutants can, for example, bind to a sequence from pBR322 (11) only when its CpG dinucleotides are methylated (Fig. 6).

All four altered-specificity mutants harbor an A-to-R change at position 477. This is the only amino acid substitution shared by all four mutants, indicating that it plays a key role in the altered specificity. It is not sufficient, however, because an RFX1 mutant (AltSpR) engineered to contain only the R-477 mutation cannot bind to the PyMutA site (Fig. 6). Consequently, a combination of R-477 and one or more additional changes must be required. Since no substitutions other than R-477 are shared by all four mutants, several different combinations must exist. One of the combinations that is sufficient to obtain the altered specificity is R-477 together with a change from A to P at position 478, a mutation that is shared by AltSp3 and AltSp4. Indeed, an RFX1 mutant (AltSpRP) engineered to contain only the R-477 and P-478 mutations can bind efficiently to the PyMutA site (Fig. 6).

DISCUSSION

The RFX DNA binding domain represents a novel motif that has been strongly conserved throughout the evolution of eukaryotic organisms, ranging from yeasts to humans, and has been recruited into regulatory proteins having diverse functions (6). A detailed understanding of the nature and site-specific DNA binding activity of the RFX domain is therefore of general importance in the context of a wide range of different regulatory systems functioning in a large variety of different organisms. Here we present a detailed characterization of the DNA binding domain of RFX1, a prototypical member of the RFX family.

A precise definition of the target sequence recognized by RFX1 is a prerequisite for the future identification of candidates for genes that are likely to be regulated by this transcription factor. By using a site selection procedure, the consensus binding site for an RFX1 homodimer was determined to be an imperfect palindromic sequence (GTNRCC/N-N$_{0–3}$-RGYAAC) consisting of two 6-bp half-sites separated by a spacer region that is variable in length (0 to 3 bp) but is preferentially either 1 or 2 bp long. There is a strict dependence on at least one near-perfect half-site matching the RGYAAC consensus sequence. The second half-site may be considerably more degenerate (G/NGYNAC on the complementary strand). The fact that only a single high-affinity half-site is required is consistent with previous observations that binding of RFX1 is not strictly dependent on dimerization and that binding of monomers to a single half-site can be observed (Fig. 6) (28). The consensus binding site determined here for RFX1 refines and extends previous definitions of a consensus site for RFX1 to RFX3 (8, 28, 35–37). In those earlier studies the consensus site was extrapolated from a limited number of naturally occurring sites. Moreover, rather than using recombinant RFX1 as done here, those previous studies analyzed a population of native complexes, referred to collectively as EF-C (8) or MDBP (35, 36), which comprise a mixture of different RFX1 to RFX3 homodimers and heterodimers (21).

To identify residues that are crucial for activity of the DNA binding domain of RFX1, we developed a highly efficient procedure to select functional DNA binding domains from libraries containing large numbers of heavily mutated sequences. The sequences selected by this procedure permit the establishment of a consensus motif that discriminates between amino acids that are essential for binding of RFX1 and those that are dispensable. This consensus motif does not describe exclusively amino acids that are critical for RFX1, the protein used for the analysis. It also highlights amino acids that are likely to be important for all members of the RFX family. Indeed, the selection procedure appears to have mimicked quite closely the evolutionary pressures that have led to conservation of the RFX motif in different proteins and species. All of the 48 critical positions in the consensus motif have been conserved in several other members of the RFX family, and 30 of these have been conserved in all members currently known. On the other hand, residues dispensable for binding of RFX1 have diverged extensively in different RFX proteins. The 30 positions conserved in all known RFX domains probably reflect to a large extent their common structural elements and protein-DNA interactions. The 18 positions conserved in only a subset of the RFX domains may, at least in part, play a role in target site specificity. For instance, all 18 of these positions have been conserved in RFX2 and RFX3, which are known to have a target site specificity very similar to that of RFX1 (21), while only some of them have been conserved in RFX4, RFX5, and ScRFX, which have target site specificities different from that of RFX1 (references 4 and 30 and unpublished data). The consensus sequence derived from RFX1 therefore provides both a highly conserved core motif that will facilitate the identification of additional members of the RFX family and positions that may underlie differential target site specificities.

The accuracy of predictions of protein secondary structure is greatly increased when they are based on multiple sequence alignments of divergent sequences (24). We have therefore derived a secondary structure prediction from all of the known naturally occurring RFX domains. This prediction suggests the presence of three α helices upstream of a β strand. Analysis of the panel of mutated sequences generated by the selection procedure provides independent support, and thus strengthens the prediction, for helices I and III and the β strand. On the other hand, it does not support the presence of helix II. The selection procedure described here can thus be quite useful for obtaining a multiple sequence alignment that can be used to confirm, and perhaps even generate, a secondary structure prediction. Within the predicted helix I, residues that are conserved in the entire RFX family, as well as those that are important for binding of RFX1, are clustered on the same face of the helix. Helix III would consist almost entirely of residues that are conserved and important for binding and would be amphipathic, with a hydrophobic face on one side and a hydrophilic face having a net positive charge on the opposite side. Interestingly, mutations conferring an altered specificity to RFX1 lie at the N-terminal end of helix III. It is therefore tempting to speculate that helix III is a recognition helix contacting the DNA.

The method we have developed to define the consensus motif of RFX1 is simple yet extremely powerful and versatile. This is emphasized by the fact that it not only has allowed us to scan a relatively large, 75-amino-acid domain for critical residues but also has permitted us to generate altered-specificity mutants requiring a combination of at least two independent mutations. Moreover, the latter was achieved despite a complete lack of available data on the region of the DNA binding domain that contacts the DNA. The strength of our system resides in the use of *his3* as a selectable reporter gene. Use of the *his3* gene permits selective growth of functional clones in liquid medium, which allows an essentially unlimited number of clones to be screened rapidly and simultaneously. Moreover, after the selection procedure, functional clones can be rescued and analyzed either individually or as a pool. In fact, preliminary results indicate that it is possible to generate a consensus sequence directly by sequencing the entire pool rather than individual clones. These two features represent clear advantages over another recently described system based on a *lacZ* reporter gene (13). In the latter system, functional clones can be selected only individually, and a practical limit to the number of clones that can be screened simultaneously is imposed by the fact that visual selection of functional clones requires growth on plates. Finally, it should be mentioned that the system we describe here is not restricted to the analysis of DNA binding domains. It could, for example, be incorporated into a two-hybrid system (7) in order to characterize the essential amino acids, and to select altered functional mutants, of any domain involved in a protein-protein interaction. In the context of the RFX family, the conserved dimerization motif (6, 21) represents an ideal candidate domain. This dimerization domain is novel and exhibits no homology to other motifs known to be involved in protein-protein interactions.

The R-477 mutation shared by all of the isolated altered-specificity mutants is essential but not sufficient for the altered specificity. Additional changes are required. One possible combination is R-477 plus P-478. However, this combination is neither the optimal one nor the only one possible. For instance, a comparison between the binding efficiencies of the AltSp3 and AltSpRP mutants (Fig. 6) indicates that changes in addition to R-477 and P-478 contribute to the greater binding affinity of AltSp3. Moreover, P-478 is found in only two of the four selected mutants, indicating that alternative combinations can confer the same altered-specificity phenotype. Inspection of the sequences in Fig. 6 suggests that other likely combinations appear to be R-477 plus a change at V-475 and/or a change at S-479. A mutation of S-479 to C is a good candidate, since it was selected, like P-478, in two different altered-specificity mutants (Fig. 6).

RFX1 has been implicated in the regulation of a number of cellular and viral genes, including c-*myc* (16), *rpl30* (25, 26), and genes controlled by the EnhI enhancer of hepatitis B virus (9, 28). However, elucidation of its precise role in these systems has been hampered by two features of RFX1. First, RFX1 is expressed constitutively in all cell types examined (21), and analysis of its function in vivo by transfection experiments consequently cannot be performed in the absence of the endogenous protein. Second, in nuclear extracts, RFX1 exists both as homodimers and as heterodimers with other family members (21), and the precise nature of the RFX1-containing complex that is implicated therefore remains unclear. In view of these problems, the altered-specificity mutants that we have generated for RFX1 are extremely valuable tools. They could, for instance, be used in transfection experiments to evaluate the function of RFX1 at suspected binding sites in vivo, without interference from the endogenous protein. Moreover, A-477 and A-478 are conserved in RFX1, RFX2, and RFX3, suggesting that the R-477 and P-478 mutations required to generate an RFX1 altered-specificity mutant should also be sufficient to confer altered specificity on RFX2 and RFX3. The combined use of RFX1, RFX2, and RFX3 altered-specificity mutants would permit an evaluation of the respective roles of RFX1 homodimers and heterodimers at a given target site in vivo.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation and the Louis Jeantet Foundation.

REFERENCES

1. **Burley, S. K.** 1994. DNA-binding motifs from eukaryotic transcription factors. Curr. Opin. Struct. Biol. **4:**3–11.
2. **Chen, W., and K. Struhl.** 1988. Saturation mutagenesis of a yeast his3 "TATA element": genetic evidence for a specific TATA-binding protein. Proc. Natl. Acad. Sci. USA **85:**2691–2695.
3. **Cormack, B. P., M. Strubin, A. S. Ponticelli, and K. Struhl.** 1991. Functional differences between yeast and human TFIID are localized to the highly conserved region. Cell **65:**341–348.
4. **Durand, B., M. Kobr, W. Reith, and B. Mach.** 1994. Functional complementation of MHC class II regulatory mutants by the purified X box binding protein RFX. Mol. Cell. Biol. **14:**6839–6847.
5. **Ellenberger, T.** 1994. Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. Curr. Opin. Struct. Biol. **4:**12–21.
6. **Emery, P., B. Durand, B. Mach, and W. Reith.** 1996. RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. Nucleic Acids Res. **24:**803–807.
7. **Fields, S., and O. Song.** 1989. A novel genetic system to detect protein-protein interactions. Nature (London) **340:**245–246.
8. **Garcia, A. D., P. Ostapchuk, and P. Hearing.** 1991. Methylation-dependent and -independent DNA binding of nuclear factor EF-C. Virology **182:**857–860.
9. **Garcia, A. D., P. Ostapchuk, and P. Hearing.** 1993. Functional interaction of nuclear factors EF-C, HNF-4, and RXR alpha with hepatitis B virus enhancer I. J. Virol. **67:**3940–3950.
10. **Huang, L. H., R. Wang, M. A. Gama-Sosa, S. Shenoy, and M. Ehrlich.** 1984. A protein from human placental nuclei binds preferentially to 5-methylcytosine-rich DNA. Nature (London) **308:**293–295.
11. **Khan, R., X. Y. Zhang, P. C. Supakar, K. C. Ehrlich, and M. Ehrlich.** 1988. Human methylated DNA-binding protein, determinants of a pBR322 recognition site. J. Biol. Chem. **263:**14374–14383.
12. **Kobr, M., W. Reith, C. Herrero Sanchez, and B. Mach.** 1990. Two DNA-binding proteins discriminate between the promoters of different members of the major histocompatibility complex class II multigene family. Mol. Cell. Biol. **10:**965–971.
13. **Nurrish, S. J., and R. Treisman.** 1995. DNA binding specificity determinants in MADS-box transcription factors. Mol. Cell. Biol. **15:**4076–4085.
14. **Ostapchuk, P., J. F. X. Diffley, J. T. Bruder, B. Stillman, A. J. Levine, and P. Hearing.** 1986. Interaction of a nuclear factor with the polyomavirus enhancer region. Proc. Natl. Acad. Sci. USA **83:**8550–8554.
15. **Pabo, C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61:**1053–1095.
16. **Reinhold, W., L. Emens, A. Itkes, M. Blake, I. Ichinose, and M. Zajac-Kaye.** 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.
17. **Reith, W., E. Barras, S. Satola, M. Kobr, D. Reinhart, C. Herrero Sanchez, and B. Mach.** 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.
18. **Reith, W., C. Herrero Sanchez, M. Kobr, P. Silacci, C. Berte, E. Barras, S. Fey, and B. Mach.** 1990. MHC class II regulatory factor RFX has a novel DNA-binding domain and a functionally independent dimerization domain. Genes Dev. **4:**1528–1540.
19. **Reith, W., M. Kobr, P. Emery, B. Durand, C. A. Siegrist, and B. Mach.** 1994. Cooperative binding between factors RFX and X2bp to the X and X2 boxes of MHC class II promoters. J. Biol. Chem. **269:**20020–20025.
20. **Reith, W., C. A. Siegrist, B. Durand, E. Barras, and B. Mach.** 1994. Function of major histocompatibility complex class II promoters requires cooperative binding between factors RFX and NF-Y. Proc. Natl. Acad. Sci. USA **91:**554–558.
21. **Reith, W., C. Ucla, E. Barras, A. Gaud, B. Durand, C. Herrero Sanchez, M. Kobr, and B. Mach.** 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.
22. **Robzyk, K., and Y. Kassir.** 1992. A simple and highly efficient procedure for rescuing autonomous plasmids from yeast. Nucleic Acids Res. **20:**3790.
23. **Rost, B., and C. Sander.** 1993. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. **232:**584–599.
24. **Russell, R. B., and M. J. Sternberg.** 1995. Structure prediction. How good are we? Curr. Biol. **5:**488–490.
25. **Safrany, G., and R. P. Perry.** 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.
26. **Safrany, G., and R. P. Perry.** 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.
27. **Schmiedeskamp, M., and R. E. Klevit.** 1994. Zinc finger diversity. Curr. Opin. Struct. Biol. **4:**28–35.
28. **Siegrist, C. A., B. Durand, P. Emery, E. David, P. Hearing, B. Mach, and W. Reith.** 1993. RFX1 is identical to EF-C and functions as a transactivator of

the hepatitis B virus enhancer. Mol. Cell. Biol. **13:**6375–6384.

29. **Sikorski, R. S., and P. Hieter.** 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics **122:**19–27.
30. **Steimle, V., B. Durand, E. Barras, M. Zufferey, M. R. Hadam, B. Mach, and W. Reith.** 1995. A novel DNA binding regulatory factor is mutated in primary MHC class II deficiency (bare lymphocyte syndrome). Genes Dev. **9:**1021–1032.
31. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82:**1074–1078.
32. **Triezenberg, S. J., R. C. Kingsbury, and S. L. McKnight.** 1988. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. Genes Dev. **2:**718–729.
33. **Wright, P. E.** 1994. POU domains and homeodomains. Curr. Opin. Struct. Biol. **4:**22–27.
34. **Wu, S. Y. and M. McLeod.** 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.
35. **Zhang, X. Y., C. K. Asiedu, P. C. Supakar, R. Khan, K. C. Ehrlich, and M. Ehrlich.** 1990. Binding sites in mammalian genes and viral gene regulatory regions recognized by methylated DNA-binding protein. Nucleic Acids Res. **18:**6253–6260.
36. **Zhang, X. Y., N. M. Inamdar, P. C. Supakar, K. Wu, K. C. Ehrlich, and M. Ehrlich.** 1991. Three MDBP sites in the immediate-early enhancer-promoter region of human cytomegalovirus. Virology **182:**865–869.
37. **Zhang, X.-Y., N. Jabrane-Ferrat, C. K. Asiedu, S. Samac, B. M. Peterlin, and M. Ehrlich.** 1993. The major histocompatibility complex class II promoter-binding protein RFX (NF-X) is a methylated DNA-binding protein. Mol. Cell. Biol. **13:**6810–6818.