Analysis of the DNA-Binding and Dimerization Activities of *Neurospora crassa* Transcription Factor NUC-1

YOAV PELEG* AND ROBERT L. METZENBERG

Department of Biomolecular Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706

Received 9 May 1994/Returned for modification 14 June 1994/Accepted 13 September 1994

NUC-1, a positive regulatory protein of *Neurospora crassa*, controls the expression of several unlinked target genes involved in phosphorus acquisition. The carboxy-terminal end of the NUC-1 protein has sequence similarity to the helix-loop-helix family of transcription factors. Bacterially expressed and in vitro-synthesized proteins, which consist of the carboxy-terminal portion of NUC-1, bind specifically to upstream sequences of two of its target genes, *pho2*⁺ and *pho-4*⁺. These upstream sequences contain the core sequence, CACGTG, a target for many helix-loop-helix proteins. A large loop region (47 amino acids) separates the helix I and helix II domains. Mutations and deletion within the loop region did not interfere with the in vitro or in vivo functions of the protein. Immediately carboxy-proximal to the helix II domain, the NUC-1 protein contains an atypical zipper domain which is essential for function. This domain consists of a heptad repeat of alanine and methionine rather than leucine residues. Analysis of mutant NUC-1 proteins suggests that the helix II and the zipper domains are essential for the protein dimerization, whereas the basic and the helix I domains are involved in DNA binding. The helix I domain, even though likely to participate in dimer formation while NUC-1 is bound to DNA, is not essential for in vitro dimerization.

Transcriptional control in eukaryotes is mediated by sequence-specific DNA-binding proteins which interact with *cis*-regulatory sequence motifs and with the transcriptional apparatus (25). In recent years, many eukaryotic transcription factors have been characterized and their DNA recognition sequences have been identified. The helix-loop-helix (HLH) family of proteins includes a large number of transcription factors involved in a variety of cellular processes, including proliferation, differentiation, and metabolism (26, 32, 48, 54). These transcription factors bind as homodimers or as heterodimers to a conserved core sequence (CANNTG) called the E box (7, 8, 33, 34, 43). The region which provides the dimerization interface in HLH proteins consists of two amphipathic helices separated by a variable length loop, whereas the contact of the proteins with the DNA is mediated by a conserved basic domain located immediately amino terminal to the HLH motif (14, 57). In myogenic HLH proteins, the basic region has a second role in directing muscle-specific transcriptional activation (9, 13, 16). Many proteins of this family contain a zipper domain which consists of a heptad repeat of leucine residues located carboxy terminal to the HLH motif. The leucine zipper domain is involved together with the HLH motif in promotion of dimerization (8, 47). HLH proteins which contain the HLH motif but lack a functional basic DNA-binding region have been identified and are thought to act as negative regulators of other HLH proteins through formation of inactive heterodimeric complexes (6, 18).

The transcription factor NUC-1 is a positive regulatory protein which directly controls the transcription of several unlinked target genes involved in phosphorus acquisition in *Neurospora crassa* (28). NUC-1 is part of a signal transduction pathway involved in supplying the cell with phosphorus under conditions of shortage. Several genes besides *nuc-1*⁺ have been

identified, on the basis of genetic and metabolic analyses, as being involved in the regulation of the phosphorus acquisition system: *nuc-2*⁺, *preg*⁺, and *pgov*⁺ (41).

Several potential functional domains have been found in the NUC-1 protein. A putative basic domain-HLH (bHLH) DNA-binding motif has been identified at the carboxy-terminal portion of NUC-1 (28). A second functional domain identified in NUC-1 is a region potentially involved in the interaction with the negative regulatory factor(s), PREG and/or PGOV (27, 29). Several glutamine- and proline-rich regions which have been proposed to act as transcriptional activation motifs are also present in the NUC-1 protein (28). The regulation of NUC-1 activity has been proposed to be posttranslational (28).

In this report, we present evidence demonstrating that the carboxy-terminal portion of NUC-1 is sufficient for specific interaction with upstream sequences of two NUC-1 target genes: *pho-4*⁺, which encodes a repressible high-affinity phosphate permease (38, 39), and *pho-2*⁺, which encodes a repressible alkaline phosphatase (22). NUC-1 binds to DNA which contains the core sequence CACGTG, as do other HLH proteins. Before the present study, NUC-1 had been considered to be a typical bHLH protein (28). Exploring functional differences between NUC-1 and known bHLH proteins led to the discovery of an additional dimerization interface in NUC-1 protein. This dimerization domain, which is located at the immediate carboxy-terminal side of helix II, is an atypical zipper domain which contains a heptad repeat of alanine and methionine rather than leucine residues. A large loop region (47 amino acids) separates the two putative helices. Mutations or deletion within the loop region did not affect function in vitro or in vivo. Analysis of mutant NUC-1 proteins for DNA binding and dimer formation suggests that the helix II and the zipper domains are essential for the protein dimerization and therefore for DNA binding. The basic region and the helix I domains are involved in the protein DNA-binding activity. The helix I domain, even though it may participate in dimer formation while NUC-1 is bound to DNA, appears not to be essential for the dimerization of the protein in vitro.

* Corresponding author. Phone: (608) 263-3238. Fax: (608) 262-5253.

MATERIALS AND METHODS

Plasmid construction and protein expression. Wild-type and mutant NUC-1 proteins were expressed in *Escherichia coli* DH5 α MCR cells (Gibco-BRL, Gaithersburg, Md.) as fusion proteins with glutathione *S*-transferase (GST), using the expression vector pGEX-2T (51). In-frame fusions were constructed by using 1-kb *Bgl*II (filled in)-*Eco*RI fragments from the wild-type *nuc-1*⁺ gene (28), from several mutant *nuc-1* alleles (27), and from one which generated by PCR (NUC1-8; see Fig. 7) cloned into the *Sma*I-*Eco*RI sites of the vector pGEX-2T. Four characterized *nuc-1* mutant alleles (27) were used: JFL411, MKG22, JFL125, and T28M1 (T28M1 was previously referred to as 01-02). A shorter version of the protein (NUC1-14; see Fig. 7) was expressed by cloning 265-bp *Pml*I-*Eco*RV fragment (the *Pml*I site was generated for expression of the NUC1-L protein; see Fig. 7) into an *Xma*I site (filled in) of pGEX-2T. The fusion proteins were expressed and affinity purified as previously described (51), with several procedural modifications (23). Overnight cultures (100 ml of each) of transformed *E. coli* cells were diluted 1:10 into LB medium containing 100 μ g of ampicillin per ml and incubated at 37°C for 2 h. Isopropylthiogalactopyranoside (IPTG) was then added to a final concentration of 0.1 mM, and the cultures were incubated for an additional 1 h. Cell harvesting, lysis, and adsorption of the fusion protein onto glutathione-agarose beads (Sigma, St. Louis, Mo.) were performed essentially as described previously (23). Proteins used in the DNA-binding assays were cleaved from the GST by thrombin while the fusion proteins were still bound to the beads (23). The proteins thus obtained consisted of the carboxy-terminal 210 amino acids of the various alleles of the NUC-1 protein. The free NUC-1 proteins were then dialyzed against buffer containing 25 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES)-KOH (pH 7.9), 0.1 M KCl, 2 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride, and 20% (vol/vol) glycerol, aliquoted, and stored at -70°C. The wild-type protein, which was used as a GST-NUC-1 fusion protein for immunization of a rabbit, was eluted from the glutathione-agarose beads by competition with reduced glutathione (51).

Preparation of rabbit polyclonal anti-GST-NUC-1 antiserum. Rabbit polyclonal anti-GST-NUC-1 antiserum was generated by subcutaneous injection of the fusion protein emulsified in complete Freund's adjuvant. Booster injections were given with incomplete adjuvant. Serum was collected 2 weeks after each immunization.

Site-directed mutagenesis. Mutations in the NUC-1 DNA-binding domain were introduced by PCR as previously described (24). Primers used in the PCRs were obtained from the DNA Core Facility, Marshall University, Huntington, W.Va. DNA sequencing analyses were performed to confirm that only the desired mutations were introduced. Sequenase version 2.0 (United States Biochemical, Cleveland, Ohio) was used in this procedure.

Plasmid constructions and in vitro protein synthesis. For in vitro expression of the various NUC-1 proteins, a 0.7-kb *Nco*I (filled in)-*Eco*RV fragment of the wild-type *nuc-1*⁺ gene (28) was cloned into the *Pvu*II-*Eco*RV sites of plasmid pSP72 (Promega, Madison, Wis.), creating plasmid pYP210. A series of plasmids encoding the different mutant *nuc-1*⁺ proteins was then constructed by replacing the 540-bp *Xho*I-*Eco*RV fragment of pYP210 with the corresponding fragments containing the various mutations from mutant *nuc-1* alleles (28) and those generated by PCR. A second plasmid containing the wild-type sequences was constructed by ligating a 970-bp *Bam*HI-*Eco*RV DNA fragment from the *nuc-1*⁺ gene into the *Bam*HI-*Pvu*II

sites of plasmid pSP73 (Promega), resulting in plasmid pYP222. The in vitro-synthesized NUC-1 proteins produced from pYP210 and pYP222 consisted of the 233 and 323 amino acids, respectively, from the carboxy-terminal end of NUC-1 (28). In vitro protein synthesis was carried out with the TNT SP6 coupled reticulocyte lysate system (Promega) in the presence of [³⁵S]methionine (DuPont, Boston, Mass.). The reactions were carried out according to the manufacturer's recommendations.

DNA-binding assays. The double-stranded DNA fragments used in the DNA-binding assays were end labeled with [α -³²P]dCTP or [α -³²P]dATP and Klenow fragment (3). For each DNA-binding reaction, labeled DNA (0.5 to 1.0 ng, 1.0×10^4 to 1.5×10^4 cpm) was incubated in 20 μ l of DNA-binding buffer containing 20 mM HEPES-KOH (pH 7.9), 50 mM KCl, 2 mM MgCl₂, 0.3 mg of bovine serum albumin per ml, 10% glycerol, 0.1 mM EDTA, 1 mM dithiothreitol, 1 μ g of poly(dI-dC) (Pharmacia, Piscataway, N.J.), and NUC-1 protein synthesized either in *E. coli* or in vitro. Unless indicated otherwise, the binding reactions were carried out with 5 ng of NUC-1 made in *E. coli* or 2 μ l of in vitro coupled transcription-translation reaction mixture. The reaction mixtures were incubated at 25°C for 30 min and loaded onto a 5% polyacrylamide gel (29:1, acrylamide/bisacrylamide). Electrophoresis was carried out at 10 V/cm, using 0.5 $\times$ Tris-borate-EDTA (50 mM Tris base, 44.5 mM boric acid, 10 mM EDTA [pH 8.0]) as the running buffer, after which the gels were dried and autoradiographed.

DNase I footprinting analysis. The reaction conditions for the DNase I footprinting analysis were similar to those described for the DNA-binding assay except for the following modifications: CaCl₂ (2 mM) was added to the DNA-binding buffer, and [³²P]DNA (4×10^4 to 5×10^4 cpm) labeled by filling in a single overhanging end with the Klenow fragment was used. After addition of NUC-1 protein synthesized in *E. coli* (~200 ng), the binding reactions were allowed to proceed for 30 min at 25°C. A freshly diluted solution of DNase I (Sigma) in the DNA-binding buffer, without poly(dI-dC), was added to a final concentration of 50 ng/ml. The DNase I digestions were carried out for 30 s at 25°C. The reactions were stopped by addition of 50 μ l of solution containing 2% sodium dodecyl sulfate (SDS), 10 mM EDTA, and 0.1 mg of *E. coli* tRNA per ml, and the samples were extracted with phenol and chloroform. After ethanol precipitation, the samples were dried and redissolved in 10 μ l of sequencing loading buffer. The samples were run on an 8% sequencing gel. Sequencing reactions, which were used to generate size markers, were carried out as previously described (40).

Protein cross-linking and immunoprecipitation. In vitro-synthesized wild-type NUC-1 or mutant proteins (1 μ l for each assay) were incubated for 30 min at 25°C in 20 μ l of buffer described above for the DNA-binding assay. Cross-linking experiments were performed by the addition of dimethyl suberimidate (DMS; purchased from Pierce, Rockford, Ill.) as a cross-linking reagent, dissolved in a freshly prepared 0.3 M triethanolamine to a final concentration of 2.5 or 10 mM. The reactions were allowed to proceed for 30 min at 25°C and were quenched by the addition of glycine to 50 mM. Samples were diluted (1:10) in immunoprecipitation buffer containing 10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1% Triton X-100, 0.1% SDS, and 1 mM EDTA and cleared by the addition of 10 μ l of 50% protein A-Sepharose beads (Sigma) for 30 min at 4°C. After removal of the beads by centrifugation, 4 μ l of polyclonal anti-GST-NUC-1 antiserum was added to the supernatants. Following incubation for 1 h at 4°C, 20 μ l of 50% protein A-Sepharose beads was added, and the samples were incu-

bated for an additional 30 min at 4°C. The beads were washed three times with 1 ml of immunoprecipitation buffer and resuspended in 20 μ l of 2 $\times$ electrophoresis sample buffer. The samples were boiled and loaded onto SDS-8% polyacrylamide gels. The gels were dried, and the labeled proteins were visualized by autoradiography.

Expression of *nuc-1* alleles in *N. crassa*. Plasmids harboring the various *nuc-1* alleles were created by replacing the 537-bp *XhoI-EcoRV* fragment of the wild-type gene (28) with the corresponding fragments containing the different mutations. Spheroplasts for transformation were prepared as previously described (1) from an *N. crassa* strain with the genotype *nuc-1*^{RIP}; *pho-4*^{RIP}; *pan-2*; *a* (strain 64-36 in R. L. Metzenberg's stock collection). Spheroplasts (5×10^7) were cotransformed (56) with 1 μ g of plasmid containing *nuc-1* allele plus 1 μ g of plasmid for complementation of the auxotrophic mutation *pan-2* (obtained from J. Grotelueschen of our laboratory). Aliquots of transformation mixture were plated onto Vogel's minimal medium, which selects only for complementation of the *pan-2* mutation, and onto minimal low-phosphate, high-pH medium (41), which selects for complementation of both *nuc-1* and *pan-2* mutations. When preliminary experiments suggested that an allele gave temperature-conditional transformation, duplicate sets of plates were incubated at 30 and 39°C. Number and size of the transformant colonies were examined 36 h after plating. To avoid any confluent growth of transformants, 1/40 of the original transformation volume was plated in most cases. If no growth was observed in a preliminary experiment, the entire transformation mixture was plated. The average size of transformant colonies was determined from measurement of at least 10 random colonies.

Miscellaneous procedures. Standard DNA manipulation and protein analysis procedures were performed as described by Ausubel et al. (3). Sequence comparisons were carried out with LASERGENE software (DNASTAR, Inc. Madison, Wis.).

RESULTS

NUC-1 specifically interacts with upstream sequences of the *pho-2*⁺ and *pho-4*⁺ genes. Detailed genetic evidence implicated NUC-1 as a positive regulator which controls the expression of genes involved in phosphorus acquisition in *N. crassa*. To test whether NUC-1 controls the expression of these genes by direct interaction, we screened for NUC-1-binding sites by a gel mobility shift assay using DNA fragments from the 5'-flanking regions of two NUC-1-controlled genes, *pho-4*⁺ and *pho-2*⁺. A truncated but otherwise wild-type NUC-1 protein, made in *E. coli*, which consists of 210 amino acids from the protein carboxy-terminal end, did bind DNA (Fig. 1). This protein is hereafter described as NUC-1(c210). Full-length NUC-1 protein (which consists of 823 amino acids), synthesized either in *E. coli* or in vitro, did not bind DNA (data not shown). The DNA fragments which were retarded by NUC-1(c210) contain a common sequence, CACGTG, which is recognized by the restriction endonuclease *PmlI*. DNA fragments cleaved at the *PmlI* site did not form retarded complexes with NUC-1(c210) (Fig. 2), suggesting that this site is involved in NUC-1 interaction with the DNA. DNA-binding assays using a 181-bp *KpnI-StyI* fragment from the 5'-flanking region of the *pho-4*⁺ gene (fragment a in Fig. 2A) revealed the formation of a single DNA-protein complex which was competed for by an excess of cold specific DNA competitor but not by a nonspecific one (Fig. 1). This result indicated a sequence-specific interaction. A similar DNA-binding pattern was observed with other DNA fragments (b and c in Fig. 2A and B,

FIG. 1. DNA-binding assays with wild-type and mutant NUC-1(c210) proteins. Gel mobility shift assays were performed as described in Materials and Methods, using ³²P-labeled 181-bp *KpnI-StyI* fragment from the 5' flank of the *pho-4*⁺ gene (Fig. 2A, fragment a). The same DNA fragment, but unlabeled, was used as a specific competitor. Lanes: 1, control DNA without protein; 2, wild-type protein; 3, wild-type protein plus 100 ng of unlabeled specific competitor; 4, wild-type protein plus 130 ng of unlabeled 261-bp *StyI* fragment from the 5' flank of the *pho-4*⁺ gene (Fig. 2A, fragment d) as a nonspecific competitor; 5, mutant NUC-1(c210) protein, T28M1; 6, mutant NUC-1(c210) protein, JFL125; 7, mutant NUC-1(c210) protein, JFL411; 8, mutant NUC-1(c210) protein MKG22.

respectively) containing the consensus sequence, CACGTG (data not shown).

Several null alleles of *nuc-1* have been identified. *N. crassa* strains harboring these mutations fail to synthesize any of the phosphorus acquisition enzymes and therefore are unable to grow under conditions of low phosphate and high pH (41). Molecular characterization revealed that most of the missense mutations are localized in the in the bHLH motif of NUC-1, whereas frameshift or nonsense mutations are scattered (27). Proteins encoded by four of these mutant alleles are identical to the wild-type NUC-1(c210) except for a single amino acid substitution either in the basic region (JFL411 and MKG22) or in the putative helix I motif (JFL125 and T28M1). These were tested for binding to DNA. No retarded complexes were detected with any of these mutant proteins (Fig. 1, lanes 5 to 8), even when higher concentrations of mutant proteins were used (data not shown). Two of these mutations (JFL125 and JFL411) have been characterized as temperature sensitive (29), since *N. crassa* strains harboring these mutant alleles showed no growth at 37°C under restrictive conditions but did show slight growth at 25°C. Our results did not reveal any DNA-binding activity for proteins carrying these mutations in assays performed either at 25°C (Fig. 1) or at 37°C (data not shown).

To define more precisely the DNA recognition sequences involved in the NUC-1 interaction with DNA, we performed DNase I footprinting analysis. The results of several of the DNase I footprinting reactions are shown in Fig. 3. They demonstrate that NUC-1 binds to sequences located in the *pho-4*⁺ upstream region between bp -315 to -334 (called the -325 box) and between bp -1476 to -1492 (called the -1483 box) and to sequences in the *pho-2*⁺ upstream region located between bp -343 to -361 (called the -353 box) (Fig. 2 and 3). As predicted from the DNA-binding assays, NUC-1-binding sites contain the core sequence CACGTG (Fig. 3), which has been identified as a target DNA sequence for many other HLH proteins. The presence of two NUC-1-binding sites in the *pho-4*⁺ 5'-flanking region located more than 1.1 kb apart was not unexpected, because two NUC-1-controlled transcripts, with sizes of 3.3 and 2.4 kb, have been previously identified (38, 39). Further study will be required to determine whether

FIG. 2. Summary of NUC-1 DNA-binding reactions. ³²P-labeled fragments flanking the *pho-4*⁺ (A) and *pho-2*⁺ (B) genes on the 5' side were used in the DNA-binding assays as described in Materials and Methods. Positions of NUC-1(c210)-binding sites as revealed by gel mobility shift assays and DNase I footprinting analysis are shown as open boxes. Numbers beneath the binding sites indicate the positions of *Pml*I cleavage sites relative to the putative translation start site (+1). Solid lines represent the DNA fragments used in the binding assays. Vertical arrows point to the position of the transcription start sites of the *pho-4*⁺ gene (39) and the *pho-2*⁺ gene (22). Letters above some of the DNA fragments represent the fragments used in the DNA-binding assays and the DNase I footprinting experiments (see Fig. 1 and text for details). +, band shift was detected; -, no band shift was detected. Restriction sites: A, *Ava*I; Bg, *Bgl*II; D, *Dde*I; H, *Hinc*II; K, *Kpn*I; Na, *Nar*I; N, *Nco*I; Pm, *Pml*I; P, *Pst*I; S, *Sma*I; St, *Sty*I.

NUC-1-binding sites function as upstream activation sequences.

Mutations in the basic and the helix I domains do not interfere with NUC-1 dimer formation. Previously described proteins which belong to the HLH family are characterized by having several surfaces involved in protein oligomerization. The HLH motif has been shown to be essential for the dimerization properties of bHLH proteins (14, 57), whereas in bHLH-zipper domain (bHLH/Z) proteins, dimerization is mediated both by the HLH domain and the leucine repeats (8, 47). Before the present study, NUC-1 protein had been considered a bHLH protein (28), with no zipper extension. Therefore we expected that its dimerization properties would not be different from those of other bHLH proteins. To investigate whether NUC-1 has oligomerization properties similar to those of other HLH proteins, we tested different mutant NUC-1(c210) proteins which failed to bind DNA for the ability to interfere with the wild-type NUC-1(c210) DNA-binding activity. This approach has been previously taken by

Voronova and Baltimore (57) to demonstrate that E47 proteins with mutations in the basic domain are able to interfere with binding by wild-type protein whereas mutant proteins with changes in the helices or loop domains are unable to do so. DNA-binding assays were performed by incubating a constant amount of wild-type NUC-1(c210) protein with increasing concentrations of mutant NUC-1(c210) proteins. The results demonstrate that proteins with a mutation in the basic domain, JFL411 (Fig. 4A, lanes 6 to 8) and MKG22 (data not shown), were able to reduce the amount of the retarded wild-type NUC-1(c210)-DNA complex. Surprisingly, the same pattern of inhibition of wild-type NUC-1(c210) DNA-binding activity was obtained by using proteins with mutations in the helix I domain, T28M1 (Fig. 4A, lanes 3 to 5) and JFL125 (data not shown). In contrast, bacterially produced protein with a single mutation in the helix II domain (Fig. 4B) did not bind DNA and did not affect DNA binding by the wild-type protein.

The order of addition to the incubation mix appears to be unimportant. No differences were observed in the inhibition of

FIG. 3. DNase I footprinting analysis. Reactions were carried out as described in Materials and Methods, using fragments from the *pho-4*⁺ (A) and the *pho-2*⁺ (B) 5'-flanking regions, with (+) or without (−) NUC-1(c210) protein made in *E. coli*. (A) A 364-bp *KpnI-XhoI* DNA fragment from the *pho-4*⁺ promoter was cloned into the *KpnI-SalI* sites in the multiple cloning sites of the vector pGEM3Zf(+) (Promega). The fragment used in the DNase I footprinting was then removed as a 197-bp *EcoRI-StyI* fragment. The upper strand was labeled with [³²P]dCTP at the *StyI* site, and the lower strand was labeled with [³²P]dATP at the *EcoRI* site. This binding site is called the −1483 box (see Fig. 2A). (B) A 223-bp *AvaI-DdeI* fragment was removed from the *pho-2*⁺ 5'-flanking region, and the lower strand was labeled with [³²P]dCTP in the *AvaI* site. This binding site is called the −353 box (see Fig. 2B). The protected DNA sequences are shown to the right of the lanes for DNase I footprinting reactions. The CACGTG core sequence is boxed in each lane. Products of the A+G and C+T Maxam and Gilbert sequencing reactions were used as size markers.

the wild-type NUC-1 binding to DNA when the proteins were preincubated prior to the addition of DNA (Fig. 4A) or when the wild-type protein was allowed to associate with DNA prior to the addition of the mutant proteins (data not shown). Since in our experiments the target DNA is added in a molar excess over NUC-1(c210) protein, these results imply that the wild-type protein is able to dissociate from its target DNA to monomers which then reassociate to form inactive complexes with subsequently added mutant protein.

Our results clearly indicate that some functional differences distinguish NUC-1 from E47. The most reasonable explanation is that the ability of NUC-1 helix I mutant proteins to interfere with the wild-type binding to DNA (Fig. 4A) resulted from the formation of inactive complexes which were unable to bind DNA. This result clearly indicates that the helix I mutations tested did not abolish the proteins dimerization.

To further study the oligomerization properties of wild-type and mutant NUC-1 proteins, we used chemical cross-linking. Experiments were performed with in vitro-synthesized ³⁵S-labeled NUC-1 fragments which consisted of 233 and 323 amino acid residues from the protein carboxy terminus [NUC-1(c233) and NUC-1(c323), respectively]. Cross-linked wild-type NUC-1(c233) protein appears to consist of a homodimer, as judged by the molecular weight of the slowly migrating

complex (Fig. 5A). NUC-1(c233) forms homodimers even in the absence of its target DNA sequences (Fig. 5A, lanes 2 and 3). Homodimer formation was also obtained when NUC-1(c233) proteins containing mutations in the basic domain or in the helix I domain were tested (Fig. 5A, lanes 6 to 9). Furthermore, the mutant proteins formed heterodimers with a larger version of the wild-type protein, as judged by the appearance of a new, more slowly migrating protein complex when cross-linking experiments were performed with in vitro-synthesized unlabeled wild-type NUC-1(c323) protein with various labeled NUC-1(c233) proteins (Fig. 5B). Formation of homodimers or heterodimers by the wild-type protein in the absence of its target DNA (Fig. 5A and B) is similar to that observed previously (42, 57).

NUC-1 has a high degree of similarity in its DNA-binding domain with other HLH proteins. The functional differences between NUC-1 and E47 presumably stem from their differences in structure. Comparison of the amino acid sequence of the NUC-1 DNA-binding domain with sequences of other transcription factors containing the putative HLH motif revealed similarity in the basic region and the putative helix I and II motifs (Fig. 6). There is a substantial resemblance between the NUC-1 DNA-binding domain and that of the PHO4 protein of the yeast *Saccharomyces cerevisiae*, which is func-

FIG. 4. Wild-type NUC-1 protein DNA-binding activity with increasing amounts of mutant NUC-1 proteins. Gel mobility shift assays were performed as described in Materials and Methods, using the same DNA fragment used in the DNA-binding experiments shown in Fig. 1. The wild-type and mutant NUC-1(c210) proteins made in *E. coli* were added to the reaction mixture prior to the addition of the DNA and incubated for 15 min at 25°C before supplementation of the reaction mixture with DNA. (A) Incubation of the wild-type protein NUC-1(c210) with mutant proteins in the basic and helix I domains. The wild-type protein was added at 5 ng per reaction (lanes 2 to 8). Each mutant protein was added, as indicated, at 2-, 10-, and 20-fold molar excesses over the wild-type protein (lanes 3 to 8). Lanes: 1, control DNA without protein; 2, wild-type NUC-1(c210) protein; 3 to 5, wild-type NUC-1(c210) protein with increasing concentrations of T28M1 mutant NUC-1(c210) protein (helix I region mutation); 6 to 8, wild-type NUC-1(c210) protein with increasing concentrations of JFL411 mutant NUC-1(c210) protein (basic region mutation). (B) Incubation of wild-type protein with helix II mutant protein. Increasing concentrations of bacterially expressed protein, helix II mutant (Fig. 7, NUC1-8), were incubated with constant amounts of wild-type protein. The wild-type protein was added at 5 ng per reaction (lanes 2 to 5), and the mutant protein was added, as indicated, at 2-, 10-, and 20-fold molar excess over the wild-type protein. Lanes: 1, helix II mutant protein (NUC1-8) at 5 ng; 2, wild-type NUC-1(c210) protein; 3 to 5, wild-type NUC-1(c210) protein with increasing concentrations of NUC1-8 protein.

tionally analogous to NUC-1 (35, 44). These two proteins have an overall 32% identity and 51% similarity (28), which suggests that they are evolutionarily related. The various bHLH and bHLH/Z proteins contain a heptad repeat of hydrophobic amino acids in the putative helix II in the bHLH and bHLH/Z proteins, a motif which extends to the proposed leucine zipper in bHLH/Z proteins (boxed residues in Fig. 6). Interestingly, NUC-1 contains a heptad repeat of the amino acids alanine and methionine at the immediate carboxy-terminal end of the helix II motif in the same position as the leucine repeats in the bHLH/Z proteins (Fig. 6). This potential zipper domain bears otherwise little homology to other zipper domains (Fig. 6).

Analysis of the sequence also indicates the presence of significant structural difference, in the DNA-binding domains, between NUC-1 and other HLH proteins. NUC-1 has an extraordinarily large loop separating the two putative helices (Fig. 6). This loop, which consists of 47 amino acids residues, is rich in charged amino acids (19 residues) and glycine (14 residues). The NUC-1 loop region is typical, however, in being rich in amino acids (glycine, aspartic acid, and serine) which are abundant in other loops and in being low in hydrophobic residues which are strongly disfavored in loops (36).

DNA-binding and dimerization properties of various *nuc-1* alleles. To test further the function of the different domains in the NUC-1 protein, we introduced a series of mutations into the various domains, including the potential zipper domain and the loop region. We then analyzed their effects on DNA binding and dimerization.

Proteins which were altered by a single mutation or double mutations in the basic domain or in the helix I domain had the DNA-binding and dimerization properties expected from behavior of the previously described mutants (i.e., they formed dimers but failed to bind DNA) (Fig. 7).

We introduced mutations into the loop region which altered some of the charged amino acids. We also deleted most of the loop region, including many of the charged residues, so that only nine amino acid residues remained (Fig. 7, allele L). Neither the amino acid substitutions nor the deletion abolished

FIG. 5. Chemical cross-linking of wild-type and mutant NUC-1 proteins. Cross-linking experiments were performed by using DMS as a cross-linking reagent with ^{35}S -labeled in vitro-synthesized proteins followed by immunoprecipitation as described in Materials and Methods. (A) Cross-linking of wild-type NUC-1(c233) protein (lanes 1 to 5) and NUC-1(c233) mutant proteins (lanes 6 to 9). DMS was not included (lane 1) or was added at a final concentration of 2.5 mM (lanes 2, 4, and 6 to 9) or 10 mM (lanes 3 and 5). Unlabeled DNA was used in some of the reactions (lanes 4 and 5) at 1 ng per reaction. The same DNA was used in the mobility shift assays (Fig. 1). (B) Cross-linking of labeled wild-type and mutant NUC-1(c233) proteins with unlabeled wild-type NUC-1(c233) protein. DMS was added at a final concentration of 2.5 mM. (C) Cross-linking of NUC-1(c233) mutant proteins (Fig. 7, NUC1-6 to NUC1-9). DMS was added at a final concentration of 2.5 mM. Gibco-BRL prestained protein molecular weight standards were used as size markers. Positions of the monomer (M), homodimer (HD), and heterodimer (HTD) are indicated.

FIG. 6. Comparison of the DNA-binding domains of NUC-1 and selected HLH proteins. Amino acids are represented by a single-letter code. Boundaries of the basic, helix I, loop, helix II, and zipper domains are drawn as previously described (20). Residues shared by NUC-1 and other HLH proteins are shaded. Boxed residues denote the heptad repeat of hydrophobic amino acids which comprise the putative helix II and zipper domains. Sources for sequences: NUC-1 (28); Max (8); c-Myc (5); Myn (47); Mad (4); PHO4 (35, 59); E47 (42); myogenin (58); MyoD (15); CBF1 (10).

the dimerization or the DNA-binding activities of the proteins (Fig. 5C and 7). These findings, therefore, strongly suggest that most of the loop region is not essential for the DNA-binding or dimerization activities of the protein.

The putative helix II domain is positioned immediately carboxy terminal to the loop region. Single and double mutations which alter some of the amino acid residues in this region abolished the dimerization and DNA-binding activities (Fig.

5C and 7). Furthermore, a previously identified mutant NUC-1 protein (from allele JFL427, containing a stop codon in the carboxy-terminal side of helix II domain) was also unable to form dimers or to bind to DNA (Fig. 7).

NUC-1 contains a region which consists of a heptad repeat of the amino acids methionine and alanine in a position similar to that of the leucine repeat in the zipper domain in bHLH/Z proteins (Fig. 6). Several mutations were introduced into this

	BASIC	HELIX I	LOOP	BINDING ACTIVITY	DIMERIZATION	COMPLEMENTATION
NUC1-WT	SKRTSHKLABQGRNRINSALQEIATLLPKAPAKEGGDGDGHSSSGGGGSGGADREDKREKDKDKAGGGIPNS			+	+	+
T28M1		P-----		-	+	- (0)
JFL125*		D-----		-	+	- (0)
JFL411*	-E-----			-	+	- (0)
MKG22	-K-----			-	+	- (0)
NUC1-1	-EG-----			-	+	- (0)
NUC1-2*		K-----		-	+	- (0)
NUC1-3			N-P-----	-	+	- (0)
NUC1-4			P-T-----	-	+	- (0)
NUC1-5				+	+	+
NUC1-6				+	+	+
NUC1-L			A-----	+	+	+
NUC1-14				G-----	-	N

	HELIX II	HYDROPHOBIC REPEAT	BINDING ACTIVITY	DIMERIZATION	COMPLEMENTATION
NUC1-WT	KASTVEMAI EYIKQLQKEVADANKRAEFAERKLVEKMGQGAA		+	+	+
JFL427	-----		-	-	- (0)
NUC1-7	-----NN-----		-	-	- (0)
NUC1-8	-----P-----		-	-	- (0)
NUC1-9	-----P-----		-	-	-/+ (1370)
NUC1-10	-----P-----		-	-	- (0)
NUC1-11	-----P-----R-----		+	+	+
NUC1-12	-----G-----P-----		-/+	-/+	+
NUC1-13	-----P-----		-	-	-/+ (1595)

FIG. 7. Summary of DNA binding, dimerization, and in vivo complementation of various NUC-1 proteins. Allele numbers are shown on the left, and the results are shown on the right side. +, the NUC-1 protein is functionally active; -, no activity; -/+, reduced activity (see text for details). The positions and amino acid changes for each allele are indicated. Boundaries of the different domains were as designated in Fig. 6. Heptad hydrophobic residues are underlined. Numbers in parentheses in the complementation column represent numbers of transformants per 1 μ g of DNA obtained under conditions selective for *nuc-1* and for *pan-2* (see text for details). Transformation assays were carried out as described in Materials and Methods. Numbers represent averages of two independent transformation experiments. A phosphate permease gene, *pho-4*, of the host strain is also inactivated. However, this does not interfere with growth because an alternative phosphate permease, encoded by *pho-5*⁺, is still present and allows growth under low-phosphate, high-pH conditions if an active NUC-1 protein is present (55). An asterisk indicates a temperature-sensitive allele. Numbers for the temperature-sensitive alleles are for the restrictive temperature (39°C). Under this condition, over 1,000 transformant colonies per μ g of plasmid DNA were obtained with the wild-type *nuc-1*⁺ allele. NUC-1-WT, wild-type NUC-1. N, not tested.

FIG. 8. Interference with the DNA binding of the wild-type NUC-1 protein by mutant proteins. Gel mobility shift assays were performed as described in Materials and Methods. The reaction conditions were the same as described for Fig. 4 except that a 50-fold molar excess over the wild-type protein concentration was used (lanes 2 to 4). Lanes: 1, wild-type NUC1(c210) protein; 2, wild-type NUC1(c210) protein with excess of mutant NUC1-14 protein; 3, wild-type NUC1(c210) with excess of mutant T28M1 protein; 4, wild-type NUC1(c210) protein with excess of mutant NUC1-8 protein.

potential zipper domain to assess the importance of this region for NUC-1 activity. The properties of these mutant NUC-1 proteins suggest that the region which is essential for the protein dimerization extends beyond the putative helix II and includes at least part of the potential zipper domain. Several mutations in this region abolished protein dimerization and DNA binding (Fig. 5C and 7, alleles 9, 10, and 13). Another of the mutant alleles (Fig. 7, allele 12) showed reduced binding and dimerization activities, as judged by the low intensity of the slowly migrating bands, whereas allele 11 could not be distinguished from wild type in these respects (Fig. 7). These results indicated that mutations in the helix II and the zipper domain abolish DNA-binding activity as well as dimerization of the protein.

The findings that proteins with mutations in the helix I domain, including the insertion of the helix-breaker proline, interfere with the wild-type protein binding to DNA (Fig. 4A) and form homo- and heterodimers (Fig. 5A and B) indicate that the helix I domain may not be essential for protein dimerization in vitro. To test this assumption, we analyzed the functions of a truncated, *E. coli*-produced protein which lacks the entire basic and helix I domains as well as most of the loop region (Fig. 7, NUC1-14). This protein, which contains intact helix II and zipper domains, was able to reduce the amount of the wild-type protein binding to DNA when added in excess (Fig. 8), whereas protein with a mutation in the helix II domain was unable to do so (Fig. 8, NUC1-8; Fig. 4B). Judging from the intensity of the band given by the DNA-protein complexes, the extent of interference with the wild-type NUC-1(c210) DNA binding by the truncated NUC1-14 protein was somewhat reduced (about ninefold) compared with the interference by the same molar excess of a protein containing a mutation in the helix I domain (Fig. 8). Furthermore, chemical cross-linking followed by SDS-polyacrylamide gel electrophoresis and Western blot (immunoblot) analysis of *E. coli*-produced wild-type NUC-1(c210) and NUC1-14 proteins revealed the formation of homodimers, whereas no dimer was formed by a helix II-mutant protein (NUC1-8) produced in *E. coli* (results not shown).

Expression of the various *nuc-1* alleles in *N. crassa*. We examined whether the different mutant NUC-1 proteins can function in *N. crassa* to activate the phosphorus-regulated genes, thus restoring the ability of a *nuc-1* host to grow under low-phosphate, high-pH conditions. As a host in the transfor-

mation, we used a pantothenic acid-requiring strain in which the wild-type *nuc-1*⁺ gene had been inactivated by using repeat-induced point mutation, a process which detects duplications premeiotically and irreversibly inactivates both copies of the duplicated gene (50). This host was cotransformed with a plasmid for complementation of the auxotrophic mutation *pan-2* and with any of a variety of plasmids carrying altered versions of *nuc-1*. The majority of the mutant proteins which were unable to dimerize and/or to bind to DNA in vitro were completely unable to activate the phosphorus-regulated genes to allow growth under low-phosphate, high-pH conditions (Fig. 7). Two previously described temperature-sensitive alleles (Fig. 7, alleles JFL125 and JFL411) and one of the PCR-generated mutants (Fig. 7, allele 2) allowed weak growth of transformants at 30°C but none at 39°C. Two of the mutant alleles gave somewhat unexpected results (Fig. 7, alleles 9 and 13). The proteins made by these mutant alleles were unable to dimerize or to bind to DNA, as judged by the in vitro assays, but they were able to complement the *nuc-1* mutation in vivo. However, they did so only weakly, the growth of strains bearing these two alleles being slower than that of strains carrying the wild-type allele. An average size of about 1 mm was observed for colonies of transformants harboring these two alleles, compared with over twice that size for the wild-type allele, and this difference became more pronounced when the plates were incubated beyond 36 h (data not shown). Other mutant alleles which complemented the *nuc-1* mutation formed colonies indistinguishable from wild-type colonies in vigor and morphology (data not shown). The results of the transformation assays were highly significant, since over 1,000 transformants were counted in each assay recorded as positive, whereas no transformants were obtained when several of the mutant *nuc-1* alleles were used (Fig. 7). All of the transformation assays, including those that gave no colonies under conditions selective for both *nuc-1* and *pan-2* (Fig. 7), gave rise to over 1,000 transformants under conditions selective for only the *pan-2* mutation (data not shown). This result indicates that the failure to complement the *nuc-1* mutation (Fig. 8) was due to absence of functional NUC-1 protein. Under conditions selective for *pan-2* alone, the size of the colonies was in the range of 3 to 4 mm in all of the different transformation assays. It was notable that a somewhat larger number of transformants was obtained when the wild-type *nuc-1* allele was used than when other complementing alleles were used (Fig. 7). Possibly this difference reflects a reduced likelihood on the part of these *nuc-1* alleles of getting through some early, critical stage of expression.

DISCUSSION

In this paper, we report that the *N. crassa* transcription factor NUC-1 is a DNA-binding protein with several unique properties. Bacterially expressed NUC-1(c210) interacts specifically with upstream sequences of two NUC-1 target genes, *pho-2*⁺ and *pho-4*⁺. NUC-1(c210) binds to DNA fragments containing the core sequence CACGTG, which has been identified as a target sequence for class B, bHLH and bHLH/Z transcription factors (11). The NUC-1 carboxy-terminal end has a similar organization and a high degree of sequence similarity to other transcription factors belonging to the HLH family of proteins (Fig. 6). However, despite this resemblance, NUC-1 shows several unique features compared with other HLH proteins.

Initially, NUC-1 protein had been considered as a member of the bHLH family of proteins without a zipper domain at the carboxy-terminal side of the putative helix II domain. Therefore, we expected that NUC-1 functions would be similar to

those of other well-studied bHLH proteins. One of the better-characterized bHLH proteins is E47 (57). This protein has been subjected to a detailed mutational analysis (57), which determined that the DNA-binding and dimerization domains in bHLH proteins are distinct. Voronova and Baltimore (57) showed that the E47 proteins altered in the basic region interfere with the DNA binding of the wild-type protein in a dominant fashion, whereas proteins altered in the HLH domain were unable to do so. When we tested NUC-1 mutant proteins (Fig. 4A) by using a similar approach, we found that increasing concentrations of proteins altered in the basic region interfered with the DNA binding of the wild-type protein (Fig. 4A), whereas protein with a mutation in the helix II did not (Fig. 4B). Surprisingly, helix I mutant proteins behaved similarly to the basic region mutant proteins; namely, they interfered with the DNA binding of the wild-type protein (Fig. 4A). Cross-linking experiments (Fig. 5A and B) also indicated that these mutant proteins can form homo- and heterodimers in a way similar to that observed for the wild-type protein. These functional differences between E47 and NUC-1 could be explained by structural dissimilarities. One of the possibilities which may provide an explanation for the behavior of proteins altered in helix I is the presence of an additional dimerization interface in NUC-1 protein. Earlier observations indicated that c-Myc dimer formation, but not tetramer formation, can be accomplished in the absence of the HLH domain as long as there is an intact leucine zipper (12). To investigate more carefully the possible existence of a zipper domain, we introduced mutations into the carboxy-terminal side of the helix II domain. Some of the mutations abolished both dimerization and DNA binding (Fig. 7), indicating that this region is essential for function. The homology which the NUC-1 zipper domain bears to other zipper domains in bHLH/Z proteins is limited (Fig. 6). This region does not contain a heptad repeat of leucine residues, not even one in which leucine is the predominant amino acid, as is characteristic of well-studied bHLH/Z proteins like Max and c-Myc. Instead, it has a heptad repeat of methionine and alanine in the same position as the leucine repeat in bHLH/Z proteins (Fig. 6). Since no proline residues are present in the helix II domain or in the alanine-methionine repeat domain, these regions may form a continuous helix. The presence of a weakly hydrophobic amino acid such as alanine as the predominant residue in a zipper-like domain suggests that hydrophobic interactions may not be the only factor involved in NUC-1 dimer formation and that other factors such as interhelical electrostatic interactions and van der Waals interactions (2, 20, 45) should be considered as well. Mutation either in the helix II or the zipper domain abolished dimerization and DNA binding (Fig. 7). This implies that both the helix II and the zipper domains are essential for dimer formation, which in turn is necessary for DNA binding.

The helix I domain in the NUC-1 protein is likely to participate in a formation of a parallel, four-helix bundle in a way similar to that observed in the crystal structure of other members of the HLH family of proteins, Max (20), USF (19), E47 (17), and MyoD (37). However, our results indicate that dimerization *in vitro* can be accomplished without an intact helix I domain (Fig. 8) and that it is mediated by the helix II and the zipper domains. The differences observed in the extent of inhibition of the wild-type DNA binding (Fig. 8) may be explained by a higher affinity for homodimer (i.e., mutant-mutant complex) formation by the NUC1-14 protein rather than heterodimer (i.e., mutant-wild type complex) formation compared with the helix I mutant protein, T28M1. It is likely that the architecture of NUC-1 protein is similar to that of Max

rather than to that of E47 (17). The latter is characterized by an extension of the helix I domain in one α -helix turn compared with the helix I domain of Max (20). In addition to the presence of a zipper domain in the NUC-1 protein, the helix I domain of NUC-1 terminates with proline in the same position as in Max, c-Myc, and other HLH proteins (Fig. 6). The helix II tyrosine residue in Max which packs against the proline at the end of helix I (20) is also in the same position as in NUC-1 (Fig. 6). An interesting difference between NUC-1 and other HLH proteins is the presence of the helix-breaker amino acid glycine on the carboxy-terminal side of the NUC-1 basic region (Fig. 6). This might indicate that the positioning of the NUC-1 basic region on the DNA is different from that observed for other HLH proteins. It is clear from the present work that the helix I domain is essential for NUC-1 DNA-binding activity, since all of the helix I mutant proteins failed to bind DNA. This finding indicates that the helix I domain in NUC-1 may be important in proper alignment and positioning of the basic domain with the DNA.

A major structural difference between the NUC-1 protein and other HLH transcription factors was found by analysis of the amino acid sequences of their DNA-binding domains (Fig. 6). Such an analysis revealed that the amino acid sequence separating the two putative helices in NUC-1 is longer than any other loop region previously described for bHLH and bHLH/Z proteins. The loop structure was originally characterized as a novel category of protein secondary structure (36). Later the term loop was used to denote the sequences separating the two putative helices in HLH proteins (42). Many of the loop regions in HLH proteins, including that in NUC-1, do not necessarily satisfy all of the criteria initially established for an idealized omega-shaped loop (36). The presence of a large loop which consists of many charged amino acids separating the two putative helices of the NUC-1 protein might indicate that this region has a role in NUC-1 function, presumably through electrostatic interactions. It has been suggested that very long loops might contact the minor or the major grooves and provide DNA-binding selectivity in regions flanking the core sequence, CACGTG (20). However, deletion of most of the loop region (to only nine residues), which removed most of the charged residues, did not affect dimerization, DNA-binding, or *in vivo* functions of the NUC-1 protein. Several other loop regions belonging to other members of the HLH family of proteins have been subjected to mutational analysis. Deletion of the MyoD eight-residue loop region to only four residues resulted in loss of DNA-binding activity (52). Insertion of several alanine residues into the MyoD loop (14) resulted in failure to oligomerize, probably because this imposed an α -helical structure on the loop. It has been demonstrated that the loop region is crucial for the dominant negative activity of Id1 protein (46). The recent study on the crystal structure of the E47 bHLH protein bound to DNA (17) suggests that a glutamine residue within the loop region helps to stabilize the protein structure by hydrogen bonds connecting the loop with other glutamine residues present in the two helices. The function of the unusually long loop region of NUC-1 remains to be elucidated.

The complementation tests are in harmony with the *in vitro* results. Mutant alleles that did not complement the *nuc-1* mutation gave rise to proteins that did not bind to DNA and, in several cases, did not dimerize either (Fig. 7). However, two of the mutants coding for proteins which failed to dimerize or to bind DNA *in vitro* did complement the *nuc-1* mutation, albeit weakly (Fig. 7, alleles 9 and 13). Both of these had changes in the zipper domain. We suggest that the *in vivo* test is far more sensitive than the *in vitro* dimerization and

DNA-binding tests and that these mutant proteins are able to function in vivo to some very limited degree and bind to DNA, even though the in vitro experiment does not detect it. It has been proposed for the bHLH/Z protein TFEB (21) that the zipper is important for high-affinity DNA binding but that low-affinity binding can occur even in the absence of an intact zipper.

We were unable to demonstrate DNA-binding activity with the full-length NUC-1 protein synthesized in *E. coli* or in vitro. It is possible that NUC-1 contains inhibitory sequences which mask its dimerization or DNA-binding domains. Such inhibitory domains have been found in several eukaryotic transcription factors (30, 31, 49, 53). However, we cannot exclude the possibility that other events such as misfolding or inactivation of the bacterially produced protein or interaction of the in vitro-synthesized protein with some inhibitory substance in the reticulocyte lysate may prevent the full-length NUC-1 from binding DNA.

Our results underscore the need for caution in inferring secondary and tertiary structure by comparing primary protein structure with those of model systems. If function is not studied directly in the system of interest, important domains may be overlooked.

ACKNOWLEDGMENTS

We are grateful to Laura Kiessling, Patricia Kiley, and Catherine Royer for helpful discussions. We also thank members of our laboratory, Jeff Grotelueschen, Rodolfo Aramayo, Tom Randall and Wayne Versaw, for help and for numerous suggestions.

This work was supported by PHS grant GM08995 to R.L.M.

REFERENCES

- Akins, R. A., and A. M. Lambowitz. 1985. General method for cloning *Neurospora crassa* nuclear genes by complementation of mutants. *Mol. Cell. Biol.* 5:2272–2278.
- Amati, B., M. W. Brooks, N. Levy, T. D. Littlewood, G. Evan, and H. Land. 1993. Oncogenic activity of the c-myc protein requires dimerization with max. *Cell* 72:233–245.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.). 1987. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York.
- Ayer, D. E., L. Kretzner, and R. N. Eisenman. 1993. Max: a heterodimeric partner for Myc that antagonizes Myc transcriptional activity. *Cell* 72:211–222.
- Batley, J., C. Moulding, R. Taub, W. Murphy, T. Stewart, H. Potter, G. Lenoir, and P. Leder. 1983. The human c-myc oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell* 34:779–787.
- Benezra, R., R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub. 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* 61:49–59.
- Blackwell, T. K., L. Kretzner, E. M. Blackwood, R. N. Eisenman, and H. Weintraub. 1990. Sequence-specific DNA-binding by the c-Myc protein. *Science* 250:1149–1151.
- Blackwood, E. M., and R. N. Eisenman. 1991. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with myc. *Science* 251:1211–1217.
- Brennan, T. J., T. Chakraborty, and E. N. Olson. 1991. Mutagenesis of myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. *Proc. Natl. Acad. Sci. USA* 88:5675–5679.
- Cai, M., and R. W. Davis. 1990. Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. *Cell* 61:437–446.
- Dang, C. V., C. Dolde, M. L. Gillison, and G. J. Kato. 1992. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *Proc. Natl. Acad. Sci. USA* 89:599–602.
- Dang, C. V., M. McGuire, M. Buckmire, and W. M. F. Lee. 1989. Involvement of the 'leucine zipper' region in the oligomerization and transforming activity of human c-myc protein. *Nature (London)* 337:664–666.
- Davis, L. J., and T. D. Halazonetis. 1993. Both the helix-loop-helix and the leucine zipper motifs of c-Myc contribute to its dimerization specificity with Max. *Oncogene* 8:125–132.
- Davis, R. L., P.-F. Cheng, A. B. Lassar, and H. Weintraub. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* 60:733–746.
- Davis, R. L., H. Weintraub, and A. B. Lassar. 1987. Expression of single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51:987–1000.
- Edmondson, D. G., and E. N. Olson. 1993. Helix-loop-helix proteins as regulators of muscle-specific transcription. *J. Biol. Chem.* 268:755–758.
- Ellenberger, T., D. Fass, M. Arnaud, and S. C. Harrison. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix. *Genes Dev.* 8:970–980.
- Ellis, H. M., D. R. Spann, and J. W. Posakony. 1990. *extramacrochaetae*, a negative regulator of sensory organ development in drosophila, defines a new class of helix-loop-helix proteins. *Cell* 61:27–38.
- Ferré-D'Amaré, A. R., P. Pognonec, R. G. Roeder, and S. K. Burley. 1994. Structure and function of the b/HLH/Z domain of USF. *EMBO J.* 13:180–189.
- Ferré-D'Amaré, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley. 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature (London)* 363:38–45.
- Fisher, D. E., C. S. Carr, L. A. Parent, and P. A. Sharp. 1991. TFEB has DNA-binding and oligomerization properties of a unique helix-loop-helix/leucine-zipper family. *Genes Dev.* 5:2342–2352.
- Grotelueschen, J., Y. Peleg, L. Glass, and R. L. Metzberg. 1994. Cloning and characterization of the *pho-2⁺* gene encoding a repressible alkaline phosphatase in *Neurospora crassa*. *Gene* 144:147–148.
- Guam, K., and J. E. Dixon. 1991. Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal. Biochem.* 192:262–267.
- Herlitz, S., and M. Koenen. 1990. A general and rapid mutagenesis using polymerase chain reaction. *Gene* 91:143–147.
- Johnson, P. F., and S. L. McKnight. 1989. Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* 58:799–839.
- Kadesch, T. 1992. Helix-loop-helix proteins in the regulation of immunoglobulin gene transcription. *Immunol. Today* 13:31–36.
- Kang, S. 1993. Functional domains of the transcription activator NUC-1 in *Neurospora crassa*. *Gene* 130:259–264.
- Kang, S., and R. L. Metzberg. 1990. Molecular analysis of *nuc-1⁺*, a gene controlling phosphorus acquisition in *Neurospora crassa*. *Mol. Cell. Biol.* 10:5839–5848.
- Kang, S., and R. L. Metzberg. 1993. Insertional mutagenesis in *Neurospora crassa*: cloning and molecular analysis of the *preg⁺* gene controlling the activity of the transcriptional activator NUC-1. *Genetics* 133:193–202.
- Kato, G. J., W. M. F. Lee, L. Chen, and C. V. Dang. 1992. Max: functional domains and interaction with c-Myc. *Genes Dev.* 6:81–92.
- Klein, E. S., D. M. Simmons, L. W. Swanson, and M. G. Rosenfeld. 1993. Tissue-specific RNA splicing generates an ankyrin-like domain that affects the dimerization and DNA-binding properties of a bHLH protein. *Genes Dev.* 7:55–71.
- Lamb, P., and S. L. McKnight. 1991. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci.* 16:417–422.
- Lassar, A. B., J. N. Buskin, D. Lockshon, R. L. Davis, S. Apone, S. D. Hauschka, and H. Weintraub. 1989. MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. *Cell* 58:823–831.
- Lassar, A. B., R. L. Davis, W. E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub. 1991. Functional activity of myogenic HLH proteins requires hetero-oligomeriza-

- tion with E12/E47-like proteins in vivo. *Cell* **66**:305–315.
35. Legrain, M., M. De Wilde, and F. Hilger. 1986. Isolation, physical characterization and expression analysis of the *Saccharomyces cerevisiae* positive regulatory gene PHO4. *Nucleic Acids Res.* **14**:3059–3073.
 36. Leszczynski, J. F., and G. D. Rose. 1986. Loops in globular proteins: a novel category of secondary structure. *Science* **234**:849–855.
 37. Ma, P. C. M., M. A. Rould, H. Weintraub, and C. O. Pabo. 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77**:451–459.
 38. Mann, B. J., R. A. Akins, A. M. Lambowitz, and R. L. Metzenberg. 1988. The structural gene for a phosphorus-repressible phosphate permease in *Neurospora crassa* can complement a mutation in positive regulatory gene *nuc-1*. *Mol. Cell. Biol.* **8**:1376–1379.
 39. Mann, B. J., B. J. Bowman, J. Grotelueschen, and R. L. Metzenberg. 1989. Nucleotide sequence of *pho-4⁺*, encoding a phosphate-repressible phosphate permease of *Neurospora crassa*. *Gene* **83**:281–289.
 40. Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**:560–564.
 41. Metzenberg, R. L., and W. Chia. 1979. Genetic control of phosphorus assimilation in *Neurospora crassa*: dose-dependent dominance and recessiveness in constitutive mutants. *Genetics* **93**:625–643.
 42. Murre, C., P. S. McCaw, and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD* and *myc* proteins. *Cell* **56**:777–783.
 43. Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore. 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**:537–544.
 44. Ogawa, N., and Y. Oshima. 1990. Functional domains of a positive regulatory protein, PHO4, for transcriptional control of the phosphatase regulon in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **10**:2224–2236.
 45. O'Shea, E. K., J. D. Klemm, P. S. Kim, and T. Alber. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**:538–544.
 46. Pesce, S., and R. Benezra. 1993. The loop region of the helix-loop-helix protein Id1 is critical for its dominant negative activity. *Mol. Cell. Biol.* **13**:7874–7880.
 47. Prendergast, G. C., D. Lawe, and E. B. Ziff. 1991. Association of Myc, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA-binding and ras cotransformation. *Cell* **65**:395–407.
 48. Prendergast, G. C., and E. B. Ziff. 1992. A new bind for Myc. *Trends Genet.* **8**:91–96.
 49. Rabindran, S. K., R. I. Haroun, J. Clos, J. Wisniewski, and C. Wu. 1993. Regulation of heat shock factor trimer formation: role of a conserved leucine zipper. *Science* **259**:230–234.
 50. Selker, E. U. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu. Rev. Genet.* **24**:579–613.
 51. Smith, D. B., and K. S. Johnson. 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**:31–40.
 52. Starovasnik, M. A., T. K. Blackwell, T. M. Laue, H. Weintraub, and R. E. Klevit. 1992. Folding topology of the disulfide-bonded dimeric DNA-binding domain of the myogenic determination factor MyoD. *Biochemistry* **31**:9891–9903.
 53. Sun, X.-H., and D. Baltimore. 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. *Cell* **64**:459–470.
 54. Torres, R., N. Schreiber-Agus, S. D. Morgenbesser, and R. A. DePinho. 1992. Myc and Max: a putative transcriptional complex in search of a cellular target. *Curr. Opin. Cell Biol.* **4**:468–474.
 55. Versaw, W., and R. L. Metzenberg. Unpublished data.
 56. Vollmer, S. J., and C. Yanofsky. 1986. Efficient cloning of genes of *Neurospora crassa*. *Proc. Natl. Acad. Sci. USA* **83**:4869–4873.
 57. Voronova, A., and D. Baltimore. 1990. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proc. Natl. Acad. Sci. USA* **87**:4722–4726.
 58. Wright, W. E., D. A. Sassoon, and V. K. Lin. 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell* **56**:607–617.
 59. Yoshida, K., Z. Kuromitsu, N. Ogawa, and Y. Oshima. 1989. Mode of expression of the positive regulatory genes PHO2 and PHO4 of the phosphatase regulon in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **217**:31–39.