Nucleic Acids Research, 1998, Vol. 26, No. 7 **1681–1688**

Mutational analysis of p50E4F suggests that DNA binding activity is mediated through an alternative structure in a zinc finger domain that is regulated by phosphorylation

Robert J. Rooney*, Kristen Rothammer and Elma R. Fernandes+

Department of Biochemistry, St Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105, USA

Received December 5, 1997; Revised and Accepted February 7, 1998

ABSTRACT

p50E4F is a cellular transcription factor whose DNA binding activity is stimulated in a phosphorylation-dependent manner by products of the adenovirus E1A oncogene. Although p50E4F does not contain a bZIP DNA binding motif, it binds a tandemly repeated palindromic sequence in the adenovirus E4 promoter that is recognized by a large number of bZIP proteins, but with much greater stability. Analysis of deletions in the p50E4F sequence identified the regions that are responsible for its unique DNA binding properties. Sequence-specific DNA binding and factor dimerization were localized to a C-terminal region containing two C_2H_2 and one CCHC zinc finger motifs; the phosphorylation site critical for DNA binding activity was also localized to this domain. The high stability of p50E4F binding also required residues within the first 83 amino acids of the N-terminus. Analysis of single and double amino acid substitutions in the C-terminal zinc finger domain demonstrated that while the second C_2H_2 zinc finger was required for DNA binding activity, the putative structures of the first C_2H_2 and the CCHC zinc fingers were not. Instead, residues from these other zinc finger motifs appeared to participate in an alternative structure that mediates DNA binding activity and is regulated by phosphorylation.

INTRODUCTION

One of the striking features of eukaryotic sequence-specific transcription factors is the modular design of their functional domains. Distinct and separate regions mediate DNA recognition and binding, factor dimerization, transcriptional activation or repression or other regulatory interactions and can often be grouped into distinct classifications based on common structural motifs that, in principle, function in the same way in different proteins. Thus far a relatively limited number of motifs, including the basic region/leucine repeat (bZIP), basic region/helix–loop–helix (bHLH), winged helix–turn–helix (WHtH), homeobox, POU and zinc finger domains, have been found to mediate sequence-specific DNA binding and, in some cases, requisite factor dimerization (1–9 and references therein). Often the interaction of transcription factors with DNA regulatory elements is a major control point in regulation of gene expression. Therefore, it is important to define and understand the domains within a factor that mediate or influence DNA binding, since the structural characteristics of these motifs will dictate the types of regulatory mechanisms the cell can use to regulate the factor.

p50E4F is a low abundance cellular factor that is regulated by adenovirus E1A oncoproteins. During early lytic infection E1A-induced phosphorylation of p50E4F protein stimulates its binding to a tandemly repeated element in the adenovirus E4 promoter to help activate high level transcription (10–14). In uninfected cells ectopic expression of p50E4F inhibits colony and focus formation in E1A-transformed NIH 3T3 and primary rat embryo fibroblasts respectively (E.R.Fernandes and R.J.Rooney, submitted for publication), suggesting a role in cell growth control. However, the cellular genes normally regulated by p50E4F are presently unknown.

The sequence element recognized by p50E4F, RTGACGT-$^{C}/_{A}$AY, is also recognized by a number of relatively abundant bZIP factors, including ATF/CREB family members, AP-1, C/EBP, E4BP4 and NF-IL6 (13,15–22). One of these, ATF-2, can also mediate E1A-induced stimulation of the E4 promoter, through an independent mechanism that does not involve a change in ATF-2 DNA binding activity (23,24). However, mutational studies of E1A and the E4F binding site have indicated that a large portion of E1A-induced E4 promoter activity is p50E4F dependent (10,14,25). Experiments measuring protein–DNA complex dissociation have demonstrated that the stability of a p50E4F–DNA complex is one to two orders of magnitude greater than the stability of a CREB/ATF-1–DNA complex (14). We hypothesize that this greater stability enables p50E4F to outcompete more abundant factors for occupancy of its binding sites and, together with its stimulation by E1A, accounts for the apparent prominence of p50E4F in mediating E1A transactivation of the E4 promoter.

*To whom correspondence should be addressed. Tel: +1 901 495 3424; Fax: +1 901 525 8025; Email: robert.rooney@stjude.org

+Present address: Department of Virology and Molecular Biology, St Jude Children's Research Hospital, Memphis, TN 38105, USA

Previous experiments demontrated that a fragment containing the first 262 amino acids encoded by the E4F cDNA (E4F262) was functionally equivalent to endogenous p50E4F with respect to DNA binding, E1A regulation and transactivation of the E4 promoter (13). A number of motifs often found in transcription factors, including various zinc fingers, an amphipathic helix–turn–helix motif and a proline-rich region, are present in the E4F262 fragment. However, unlike the other factors that bind to the E4F site, there is no bZIP DNA binding domain. In this report a zinc finger region near the C-terminus of p50E4F was shown to mediate both phosphorylation-dependent sequence-specific DNA binding and factor dimerization. A separate region in the N-terminus was required for increased p50E4F DNA binding stability. Furthermore, mutational analysis of the DNA binding/dimerization region suggested that a central C_2H_2 zinc finger motif and portions of two other surrounding zinc finger motifs assume a phosphorylation-dependent alternative structure that mediates binding to the E4F recognition sequence.

MATERIALS AND METHODS

Plasmid DNA

Truncated sections of the E4F cDNA contained in pGST–E4F fusion constructs (as specified in the text) were synthesized by PCR and subcloned into pGEX-1 or pGEX-2T (Pharmacia). E4F-N1 and E4F-N4 truncations used for *in vitro* transcription/translation were cloned into pCITE-3b (Novagen). Single and double amino acid substitutions were created by PCR using 5′ primers that contained point mutations in the specified residues and a 3′ primer containing wild-type E4F cDNA sequence to residue 262 followed by a termination codon and introduced into the pGST-E4FN2 sequence by overlapping PCR. All cloned PCR products were sequenced in their entirety.

Protein preparation

To produce glutathione S-transferase (GST)–E4F fusion proteins pGST-E4F fusion plasmids were transformed into BL21(DE3)pLysS or NovaBlue (Novagen). Bacterial cultures were grown overnight at 30°C, diluted 1:10 in LB broth, grown for 1 h at 37°C and induced with 0.1 mM IPTG for 4 h. Induced cultures were pelleted and resuspended in 3 ml TSE buffer/l culture (TSE: 25 mM Tris, pH 8.0, 50 mM sucrose, 10 mM EDTA, pH 8.0, 0.5 mM PMSF, 100 µg/ml each aprotinin, leupeptin and pepstatin; Boehringer Mannheim). Bacteria were lysed by addition of 15 mg lysozyme/l culture and 5 min incubation at room temperature, adjustment to 10 mM $MgCl_2$, 1 mM $MnCl_2$, 10 µg/ml DNase I and a 15 min incubation at 37°C, followed by adjustment to 1× PBS, 1% Tween-20, 1% Triton X-100, 10 mM DTT and thorough mixing. After removal of cell debris by centrifugation at 4°C GST–E4F proteins were purified from the lysates by glutathione–Sepharose affinity chromatography as per the supplier's instructions (Pharmacia), dialyzed in buffer A (10 mM HEPES, pH 7.9, 50 mM KCl, 1.5 mM $MgCl_2$, 10% glycerol, 0.1 mM PMSF, 100 µg/ml each aprotinin, leupeptin and pepstatin) and stored frozen at –85°C. Protein concentrations were determined by the Bradford colorimetric assay (BioRad).

For heterodimerization analysis 2 µg GST–WT and GST–N5 protein were mixed together or separately with 2 µg GST protein, diluted with an equal volume of 6 M guanidine HCl/$NaPO_4$, pH 8.0, for 5 min and dialyzed for 12 h in buffer A. Dialyzed proteins were concentrated using a Centricon P-20 (Amicon), measured by Bradford assay and stored frozen at –85°C until use. All protein preparations were analyzed by 12% SDS–PAGE and silver staining (BioRad).

For *in vitro* synthesis of E4F-N1 and E4F-N4 proteins pCITE-E4F-N1 and pCITE-E4F-N4 plasmids were transcribed and translated in the STP (Novagen) rabbit reticulocyte lysate system containing T7 RNA polymerase and [^{35}S]methionine (NEN). Products were analyzed by 10 or 12% SDS–PAGE.

To dephosphorylate GST–E4F proteins 1 µg protein (in 30 µl buffer A) was incubated with 10 µl alkaline phosphatase coupled to acrylic beads (Sigma) at 30°C for 20 min, after which the beads were removed by centrifugation (12,13). A second procedure, in which 1 µg protein (in 20 µl WCE dialysis buffer) was adjusted to a final volume of 50 µl in 1× gel shift binding buffer (20 mM HEPES, pH 7.9, 60 mM KCl, 1 mM $MgCl_2$, 0.5 mM DTT, 4% Ficoll) containing 10 µM $ZnCl_2$ and incubated with 5 U calf intestinal alkaline phosphatase at 30°C for 90 min, followed by addition of 2 µl 1 M NaF, was equivalent in reducing GST–E4F DNA binding activity.

Gel shift assays

The structure of the E4wt, E4pm1, E4pm2 and E4pm4 probes and competitor DNAs and the binding reactions and gel shift assays for E4F and ATF were as previously described (13,14). Reactions with GST fusion proteins contained 1 µg protein and 1 µg poly(dI·dC)·poly(dI·dC) (Pharmacia). Reactions with *in vitro* synthesized proteins contained 7 µl reticulocyte lysate and 3 µg poly(dI·dC)·poly(dI·dC). Competitor DNA amounts given in the figure legends are relative to 1 ng ^{32}P-labeled probe/reaction.

Dissociation rate assays

Binding reactions containing GST–E4F proteins and ^{32}P-labeled E4wt probe were scaled up 3-fold, incubated for 40 min to allow DNA–protein complexes to form and then challenged by addition of a 500-fold molar excess of unlabeled E4wt DNA. At the times indicated 10 µl samples were removed and loaded onto a 4% polyacrylamide gel running at 150 V. After addition of the last samples the gels were run at 280 V to completion (14).

RESULTS

The p50E4F zinc finger domain mediates sequence-specific DNA binding activity

Previously p50E4F sequence-specific DNA binding activity was localized to amino acids 84–262 of the E4F peptide sequence (13). Contained within this region are several motifs that could potentially serve as DNA binding domains (Fig. 1). Residues 120–153 have the potential to form two amphipathic α-helices separated by a triple glycine turn, reminiscent of helix–turn–helix motifs found in prokaryotic DNA binding proteins (26). Immediately preceding this motif is a proline-rich sequence (residues 82–113), a characteristic noted in a number of transcriptional regulatory domains (27–35). Residues 191–247 harbor two consensus C_2H_2 zinc finger motifs with Krüppel-like linker regions (36,37). Also, immediately following the two consensus C_2H_2 motifs in the E4F sequence is an abortive C_2H_2 motif or an atypical CCHC zinc finger motif (38). However, the C-terminal half of this putative structure is deleted by the C-terminal truncation to residue 262, which does not interfere with DNA binding and thus is not required.

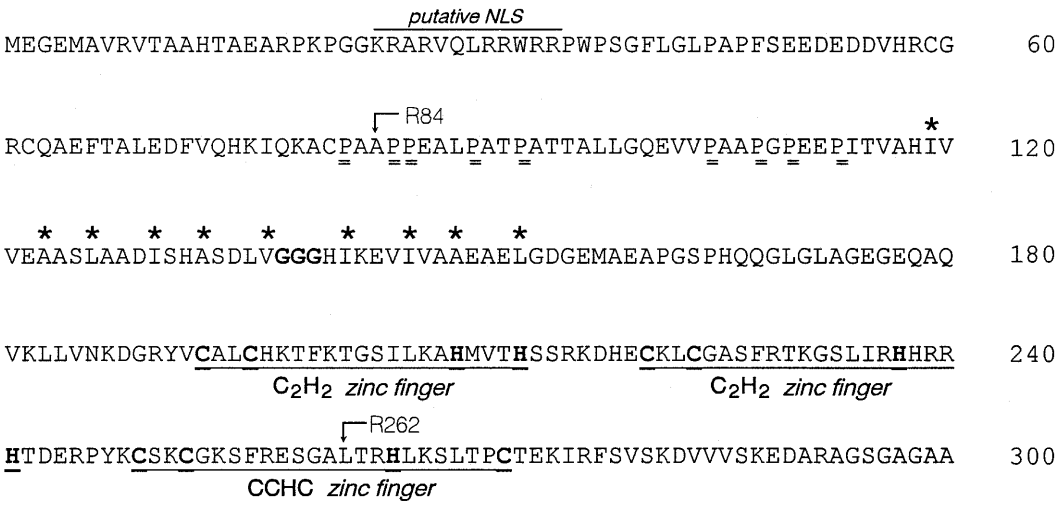


Figure 1. Potential DNA binding motifs in p50E4F. Structural motifs characteristic of transcription factors that are present in the N-terminal 300 amino acids encoded by the E4F cDNA are illustrated. C_2H_2 and CCHC zinc finger motifs located between residues 193 and 272 are underlined; conserved cysteines and histidines, required for zinc coordination, are in bold type. A potential amphipathic α-helix–turn–α-helix motif lies between residues 119 and 153; hydrophobic residues in a 4–3 repeat are denoted by *. Prolines clustered between residues 82 and 113 are double underlined. A putative nuclear localization signal (NLS) is overstruck. The positions of residues (R) 84 and 262 are denoted.

A series of GST fusion proteins containing further N- or C-terminally truncated E4F sequences were created to identify the regions responsible for specific binding to an E4F site-containing probe (Fig. 2). Proteins that were N-terminally truncated to residue 189 (GST–N1 to GST–N5) retained DNA binding activity and displayed the same sequence specificity previously demonstrated for HeLa cell p50E4F, thereby delimiting the N-terminal boundary to the zinc finger domain. C-Terminal truncation to residue 225 (GST–C1), which deletes most of the second C_2H_2 zinc finger motif, eliminated DNA binding activity, as did deletion of the entire domain (GST–C2).

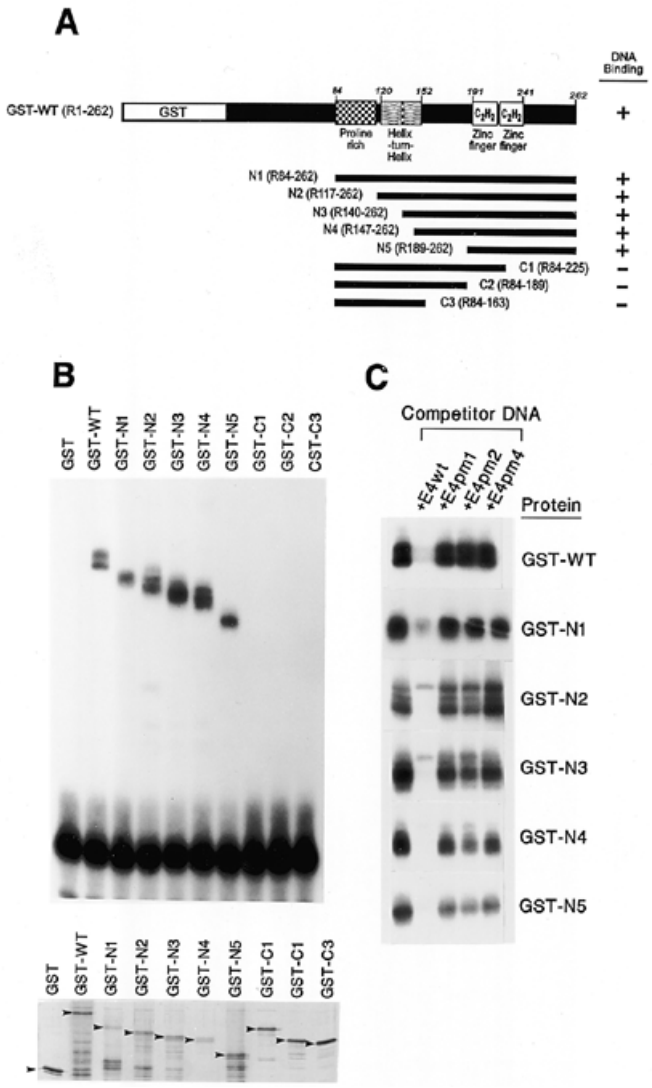
To better define involvement of the C_2H_2 zinc finger motifs, additional N-terminal truncations (Fig. 3) and substitution mutations (Fig. 4A) were introduced into this region. Truncation of the first five residues in the first C_2H_2 zinc finger motif had only a slight effect on DNA binding activity, whereas deletion of the entire first zinc finger motif abolished DNA binding (Fig. 3). Similarly, substitution of the conserved cysteines at positions 193 and 196 in the first zinc finger by serines did not diminish but instead increased DNA binding activity ~2- to 3-fold, whereas substitution of the conserved histidine at position 209 by alanine eliminated DNA binding (Fig. 4C). The increased DNA binding activity of the C193,6S mutant was consistently observed in multiple protein preparations. These results indicate that residues in the C-terminal half of the first C_2H_2 zinc finger motif are required for E4F site binding but that the zinc finger motif itself is not.

Consistent with a requirement for the second C_2H_2 zinc finger motif, as implied by the effect of the C1 truncation, a double substitution of Arg236 by leucine (in the 'Z' position for specific base contact; 39–41) and conserved His237 by asparagine eliminated DNA binding activity, whereas double substitution of Lys231 by methionine and Gly232 by serine, located in non-conserved and non-DNA contacting positions, had no effect (Fig. 4B). Outside the C_2H_2 zinc finger motifs no effects on DNA binding activity were seen with mutations of Ser215 or Thr242 in the first and second Krüppel-like linker regions respectively or with mutation of the more C-terminal Lys254 (Fig. 4B). However, a double substitution of Lys248 by methionine and Cys249 by serine, at the N-terminal end of the truncated CCHC zinc finger motif, eliminated DNA binding (Fig. 4B). Taken together these data indicate that the second C_2H_2 zinc finger motif is required for sequence-specific DNA binding, whereas the putative zinc finger structures of the first consensus C_2H_2 motif and the non-consensus CCHC motif are not. Instead, the requirement for His209 in the first motif and Lys248 and Cys249 in the CCHC motif suggests that surrounding residues on both sides of the second C_2H_2 zinc finger motif participate in an alternative domain structure that mediates DNA binding activity.

p50E4F DNA binding activity requires dimerization through the zinc finger domain

The palindromic nature of the E4F recognition site suggests that p50E4F may bind as a dimer. As shown in Figure 5A, incubation of full-length GST–WT protein with the truncated GST–N5 protein did not produce any additional protein–DNA complexes other than those seen with each protein individually. However, when the mixed proteins were treated with guanidine HCl and subsequently renatured by dialysis prior to the DNA binding reaction protein–DNA complexes of intermediate mobility were observed, indicating heterodimerization. Similar treatment of each protein mixed with GST protein did not produce the intermediate complex. Complexes of intermediate mobility were seen with renatured mixtures of GST–WT + GST–N2 proteins and GST–N2 + GST–N5 proteins respectively (data not shown).

Evidence of dimerization was also observed with E4F proteins synthesized *in vitro*. E4F-N1 and E4F-N4 proteins were translated individually or together in reticulocyte lysates and analyzed by gel shift assay (Fig. 5B). Although DNA binding activity was low, a single gel shift complex with a distinct mobility was observed for each protein translated individually. In contrast, three gel shift complexes resulted from co-translation of the two proteins, two


A
DNA Binding
GST-WT (R1-262)
GST
84
120
152
191
241
262
Proline rich
Helix -turn- Helix
Zinc finger
Zinc finger
+
N1 (R84-262)
N2 (R117-262)
N3 (R140-262)
N4 (R147-262)
N5 (R189-262)
C1 (R84-225)
C2 (R84-189)
C3 (R84-163)
+
+
+
+
+
-
-
-
B
GST
GST-WT
GST-N1
GST-N2
GST-N3
GST-N4
GST-N5
GST-C1
GST-C2
CST-C3
GST
GST-WT
GST-N1
GST-N2
GST-N3
GST-N4
GST-N5
GST-C1
GST-C1
GST-C3
C
Competitor DNA
*E4wt
*E4pm1
*E4pm2
*E4pm4
Protein
GST-WT
GST-N1
GST-N2
GST-N3
GST-N4
GST-N5