5184–5191 Nucleic Acids Research, 1993, Vol. 21, No. 22

A single amino-acid substitution in the Ets domain alters core DNA binding specificity of Ets1 to that of the related transcription factors Elf1 and E74

Rémy Bosselut, Jonathan Levin[1], Elisabeth Adjadj[2] and Jacques Ghysdael*
CNRS URA 1443, Institut Curie, Section de Biologie, Bâtiment 110, Centre Universitaire, 91405 Orsay Cedex, [1]Unité d'Ingénierie des Protéines, Biotechnologies, Institut National de la Recherche Agronomique, Domaine de Vilvert, 78352 Jouy en Josas and [2]INSERM Unité 350, Institut Curie, Section de Biologie, Bâtiment 112, Centre Universitaire, 91405 Orsay Cedex, France

Received July 12, 1993; Revised and Accepted October 7, 1993

ABSTRACT

Ets proteins form a family of sequence specific DNA binding proteins which bind DNA through a 85 aminoacids conserved domain, the Ets domain, whose sequence is unrelated to any other characterized DNA binding domain. Unlike all other known Ets proteins, which bind specific DNA sequences centered over either GGAA or GGAT core motifs, E74 and Elf1 selectively bind to GGAA corecontaining sites. Elf1 and E74 differ from other Ets proteins in three residues located in an otherwise highly conserved region of the Ets domain, referred to as conserved region III (CRIII). We show that a restricted selectivity for GGAA core-containing sites could be conferred to Ets1 upon changing a single lysine residue within CRIII to the threonine found in Elf1 and E74 at this position. Conversely, the reciprocal mutation in Elf1 confers to this protein the ability to bind to GGAT core containing EBS. This, together with the fact that mutation of two invariant arginine residues in CRIII abolishes DNA binding, indicates that CRIII plays a key role in Ets domain recognition of the GGAA/T core motif and lead us to discuss a model of Ets proteins – core motif interaction.

INTRODUCTION

The *ets* gene family encodes a novel class of sequence specific DNA binding proteins, the prototype of which is *cets1*, the cellular homolog of the *v-ets* oncogene of avian leukaemia virus E26 (for reviews, see Refs.1–3). In addition to the *c-ets1* protooncogene product (4, 5), Ets1, several Ets proteins have been identified in vertebrates, including Ets2, Fli1, Erg, GABPα, PEA3, ER81, ER71, Elf1, Elk1, SAP1, Spi1/Pu1, SpiB, and *D.melanogaster* Elg, Yan/Pok, E74A and E74B (Refs. 6–21). The *ets* genes are involved in a variety of essential biological processes including cell growth, differentiation and development, and three of them (*v-ets*, *fli1* and *spi1/Pu1*) are implicated in oncogenic processes (for reviews, see Refs. 1–3). Ets proteins bind *in vitro* to specific DNA sequences referred to as Ets binding sites (EBS). These sequences are about 10 nucleotides long and are centered over a GGAA/T core motif, flanked by more variable but non random 5′ and 3′ sequences (11, 22–30). EBS act as *cis* acting responsive elements for several Ets proteins, including Ets1, Ets2, Fli1, Spi1/Pu1 (16, 31–33). Transcriptional activation by Ets proteins appears to be mediated through synergy with other, unrelated, transcription factors which bind to specific DNA sequences located in close proximity to EBS (33–35).

Sequence homology among all known Ets proteins is limited to a domain of about 85 aminoacids, referred to as the Ets domain. The Ets domain is necessary and sufficient for these proteins to specifically bind to EBS *in vitro* (36–39). Two distinct subdomains have been individualized in the Ets domain. The carboxyterminal half, which does not share sequence homology with any other characterized protein DNA binding motif, includes several conserved basic and hydrophobic aminoacids. The amino terminal half is characterized by a repetition of three tryptophan residues spaced by 17–18 aminoacids, which resembles that observed in the 51–53 aminoacids repetitions which constitute the DNA binding domain of Myb proteins (40). To date no information exists about which residues within the Ets domain are responsible for specific contacts with DNA, nor about the nature of the protein secondary or tertiary structure(s) involved in these interactions.

Unlike all other members of the Ets family characterized so far, which bind to EBS centered over a GGAA or a GGAT core sequence (11, 22, 23, 28), the closely related E74 and Elf1 proteins bind selectively to GGAA core-containing EBS (29, 30). In this report, we took advantage of this property to identify a single aminoacid difference between Elf1/E74 and all other members of the Ets family in an otherwise highly conserved region of eleven residues in the Ets domain, referred to as conserved region III (CRIII). We show that conversion of the

* To whom correspondence should be addressed

corresponding residue of Ets1 to that found in Elf1/E74 is sufficient to impose upon Ets1 a restricted selectivity for GGAA core containing EBS. Conversely, the reciprocal mutation in Elf1 confers to this protein the ability to bind to GGAT core containing EBS. These results, together with the fact that mutation of two invariant arginine residues in CRIII abolishes Ets1 specific DNA binding, support the notion that this region of the Ets domain is directly involved in core sequence recognition of the EBS.

MATERIAL AND METHODS

In vitro mutagenesis

In vitro site specific mutagenesis was carried out on the chicken *c-ets1* cDNA (5, 6) following a two steps PCR mediated procedure (41), using one mutagenic primer, a 5′ flanking primer spanning the *c-ets1* HindIII unique site (nucleotide 1090) and a 3′ flanking primer including the *c-ets1* AatII unique site (nucleotide 1300). The resulting *c-ets1* mutagenic fragments were gel purified, digested with HindIII and AatII and subcloned into a HindIII−AatII digested *c-ets1* cDNA inserted at the BamH1 site of pBluescript SK+. Conversion of Elf1 threonine 263 (T54 in Ets domain coordinates, see Fig. 2) into lysine was carried out on the *elf1* cDNA (12; a gift of Dr J.M.Leiden) inserted between XbaI and KpnI sites of pBluescript SK+, using a one step PCR procedure with a 3′ primer including the *elf1* BglII unique site (nucleotide 1207) and a 5′ mutagenic primer spanning the BsaI cut position (nucleotides 1103−1106). The BsaI to BglII fragment of *elf1* was exchanged by the BsaI−BglII digested PCR amplified fragment. All PCR were performed using Vent DNA polymerase (Biolabs) following the manufacturer's instructions. Dideoxy sequencing of PCR amplified sequences was performed on all mutant plasmids to ensure both for the presence of desired mutations and for the absence of additional ones.
Primer oligonucleotides were as follows:

Ets1 5′ flanking: 5′ gatggctgggagttcaagctttccgatcc 3′
Ets1 3′ flanking: 5′ catctggcttgacgtccagcatggcgtgc 3′
Ets1M1 (−strand): 5′ aatagtaaagcagaccaaggctcagcttctcatag 3′
Ets1M2 (−strand): 5′ agtaacgcagagcacggcccatcgtctcatagttcatcttg 3′
Ets1M3 (−strand): 5′ ccacggctcagcgtctcatagttcatcttggg 3′
Ets1M4 (−strand): 5′ cagcttctgataggtcatcttgggcttgtttttcc 3′
Elf1T54K (+strand):5′ ggcactcgggtctcaattatgagaaaatgggaagagcactc 3′
Elf1 3′ flanking: 5′ tatatatataagatcttttggcatttc 3′

Recombinant baculoviruses and preparation of cell lysates

Recombinant AcNPV baculoviruses expressing either Elf1 or the mutant Ets1 proteins were obtained using the Pharmingen Baculogold transfection kit, following the manufacturer's instructions. Briefly, pVL941 derivatives containing the respective cDNAs were cotransfected in Sf9 cells together with linearized AcNPV viral DNA containing a deletion of the polyhedrin coding sequences. Recombinant viruses obtained from the supernatants of transfected cells were expanded by serial infection of Sf9 cells. The wild type AcNPV Ets1 was described previously (31).

Cell extracts were made from 10^7 Sf9 cells infected (m.o.i. of 20) with the respective recombinant viruses. Control extracts were made from non infected Sf9 cells. Forty hours after infection, cells were lysed by resuspension in four volumes of 10 mM HEPES, pH 7.9, 0.3 M NaCl, 1mM dithiothreitol (DTT), 0.1 mM EGTA, 10% glycerol, 0.2% Triton X100, 1% aprotinin, 0.1 mg/ml PMSF and 10mg/ml leupeptin. Cell lysates were centrifuged at 10000g for 15 minutes and supernatants were used as a source of protein for DNA binding experiments. Western blot analysis was carried out as previously described (27).

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSAs) using either Sf9 cell extracts or rabbit reticulocyte lysates were performed as described previously (31). For competition EMSAs, unlabelled competitor oligonucleotides were included in the binding assay as indicated in figures legends.

In vitro transcription/translation in reticulocyte lysates

Elf1 and Elf1T54K were expressed in rabbit reticulocyte lysates with a combined *in vitro* transcription/translation system (Promega), using T7 RNA polymerase. Reactions carried out in the presence of [^{35}S] methionine were subjected to SDS polyacrylamide gel electrophoresis to analyze expression of the *in vitro* synthetized proteins. Parallel reactions performed in the absence of labelled amino-acid were used as a source of protein in EMSAs. In both conditions, unprogrammed cell lysates were used as controls.

RESULTS

Ets protein binding sites are about 10 nucleotides long and contain a central core motif GGA(A/T), flanked by degenerate but non random 5′ and 3′ sequences which contribute, at least in part, to the DNA binding selectivity of individual members of the family (11, 22−30, 42; see Fig. 1A). The strict conservation of the core sequence in EBS and the fact that methylation interference of the core guanine residues invariably abolishes DNA binding of Ets proteins to EBS (16, 23, 27, 39, 43) suggest that this sequence is directly contacted by a highly conserved motif within the Ets domain. This motif, in turn, would be expected

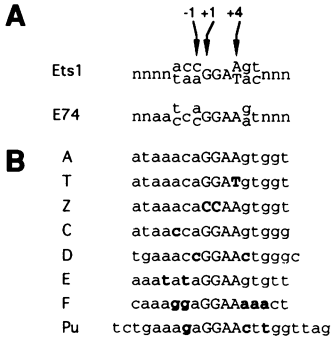

Figure 1. A. DNA binding sites of Ets1 and E74, as deduced from *in vitro* oligonucleotide selection/amplification experiments, for Ets1 (22−24) and E74 (30). Ets binding site numbering as used in the text is indicated. **B.** Oligonucleotides used in this study. Only the GGAA/T containing strand of the oligonucleotides is shown, with the EBS core depicted in capitals. In addition to sequences shown here, each oligonucleotide includes a XhoI site at either its 3′ (oligonucleotide E) or 5′ side (other oligonucleotides) and is flanked by AvaI restriction sites. Oligonucleotide A is an optimized high affinity *in vitro* binding site for Ets1. Oligonucleotides T and C are derived from A. Oligonucleotide Z differs from A by a GG→CC transversion at positions +1 and +2, and is therefore unable to bind Ets proteins. Oligonucleotide D is a high affinity binding site for Fli1 (42) and Ets1 (our unpublished data). Oligonucleotide E spans nucleotides −83 to −97 from the interleukin2 receptor α chain gene promoter (63). Oligonucleotide F contains the NF-IL2B site from the interleukin 2 gene promoter (12, 64). Between positions −5 and +7, nucleotides differing from the corresponding positions in oligonucleotide A are shown in boldface type.

to contain conserved aminoacid residues at positions directly contacting invariant bases or the phosphodiester DNA backbone, as previously shown e.g. among members of the HTH family of procaryotic DNA binding proteins (44) and eucaryotic homeodomain proteins (45).

As shown in Fig. 2 one such highly conserved sequence, referred to as conserved region III (CRIII), is found in the DNA binding domain (Ets domain) of all Ets family members. CRIII extends between positions 50 and 60 of the Ets1 DNA binding domain (throughout this paper, Ets domain residues are numbered using Ets1 as a reference; position 1 is defined as the first conserved hydrophobic residue within the Ets domain; see Fig. 2). To analyze whether CRIII played any role in EBS core sequence recognition, specific mutations were introduced in the *c-ets1* cDNA using PCR mediated site specific *in vitro* mutagenesis (see Fig. 3A for a schematic of the mutants). To produce proteins suitable for DNA binding studies, we used recombinant baculoviruses to express wild type and mutant forms of the protein in insect cells. Previous analyses with wild type Ets1 have shown that whole cell extracts from infected cells generate a single specific complex in electrophoretic mobility shift assays using a variety of EBS oligonucleotide probes (22, 27, 31; see also Fig. 3D). All mutant proteins analyzed in this study were found to be expressed at levels similar to that of wild type Ets1, as shown by Western Blot analysis using an Ets specific antiserum (Fig. 3B).

CRIII can be further subdivided into two regions (see Fig. 2). The first region (region B) includes two arginine residues (R57 and R60 in the Ets1 DNA binding domain) conserved in all Ets family members. Consistent with a role of CRIII in binding to EBS core, mutation of both arginines abolishes Ets1 binding to an oligonucleotide containing a high affinity binding site for this protein (oligonucleotide A, see Fig. 1B for sequence of the oligonucleotides used in this study) (Fig. 3C). The second region within CRIII (motif A) extends over seven aminoacid residues which are identical or similar in all Ets proteins, except for the related *D.melanogaster* E74 and human Elf1. Interestingly, unlike other Ets proteins, which bind EBS containing either a GGAA or a GGAT core sequence (11, 22, 23, 28), both Elf1 and E74 bind selectively to GGAA core-containing EBS (29, 30; see also Fig. 4B). To analyze whether these aminoacid differences between Elf1 and Ets1 were involved in the discrimination between GGAA and GGAT sequences, we generated Ets1 mutants in which selected residue(s) within the A motif were replaced by the corresponding Elf1 amino-acids (see Fig. 3A for a schematic of the mutants). We first replaced Ets1 A motif by the corresponding Elf1 residues, thereby generating mutant Ets1M2. Of note, only one residue within the CRIII region A motif (K54 in Ets1) is invariant in all Ets family members, except for Elf1 and E74 where it is replaced by a threonine residue. To investigate the role of this residue in the specific discrimination between GGAA and GGAT core-containing EBS, we therefore generated mutant Ets1M3, in which K54 was changed into threonine. As a control, we also used mutant Ets1M4, in which Ets1 residues N51 and E53 were changed into T and Q, respectively (which are observed in the corresponding position in the Spi1/Pu1 Ets domain). Both Ets1 and mutants Ets1M2, Ets1M3, Ets1M4 were found to bind efficiently to oligonucleotide A (Fig. 3D, lanes 1, 4, 7 and 10). This binding is specific since it is competed out by a 25 fold molar excess of unlabelled oligonucleotide A, but not by the same excess of a mutant oligonucleotide (oligonucleotide Z, Fig. 1B) containing a GG to CC transversion in the core sequence (Fig. 3D).

The binding of Ets1 and Ets1M2 to a set of three matched oligonucleotide probes (see Fig. 1B for sequence of the oligonucleotides) were compared by electrophoretic mobility shift assay. Oligonucleotides A and T were of identical sequence, except for the presence of a GGAA core in oligonucleotide A and a GGAT core in oligonucleotide T. Oligonucleotide C differs from oligonucleotide A by a single base change outside of the core sequence, at position 2. Ets1 was found to bind with similar efficiency to either GGAA- (probe A) or GGAT-based EBS (probe T) (Fig. 4A, lanes 4–5). In contrast, Ets1M2 was found to inefficiently bind to oligonucleotide T, as compared to its

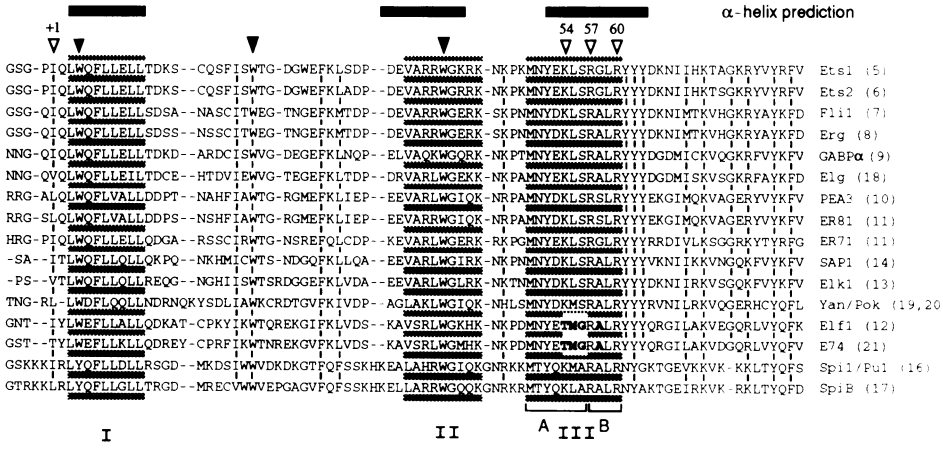

Figure 2. Ets domain sequence of representative mammalian and *D.melanogaster* Ets proteins. Residues are numbered with respect to the sequence of the Ets domain of Ets1. Positions +1 (Ile), +54 (Lys), +57 (Arg) and +60 (Arg) are indicated by a white arrowhead above the Ets1 sequence. Isolated conserved residues are indicated by vertical lines. The three conserved tryptophans which constitute the tryptophan repeat motif are indicated by black arrowheads. Regions of strong homology (Conserved Regions [CR] I, II and III) are indicated by grey boxes. Elf1 and E74 residues which differs from Ets1 within CRIII are shown in boldface type within a white box. References are indicated for each sequence. All these sequences, but those of ER71, ER81, Yan/Pok and SpiB, were used for secondary structure prediction with the 'mpred' algorithm. Three regions predicted to fold as an α-helix are indicated by thick black bars at the top of the figure.

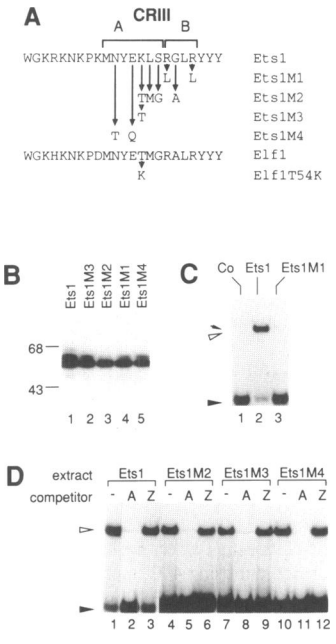

Figure 3. A. Schematic of Ets1 and Elf1 CRIII mutants used in this study. For each mutant, only those residues differing from the wild type sequence are shown. **B.** Expression of Ets1 and mutants derived thereof. An equal amount from extracts of Sf9 cells infected with the respective recombinant baculovirus were subjected to SDS-PAGE, and proteins were transferred to nylon membrane. The membrane was probed with an Ets specific antiserum (4), followed by incubation with [^{125}I] *S.aureus* protein A. Position of the 68 kd and 43 kd molecular weight markers is shown. Note that the slowly migrating components represent phosphorylated forms of the proteins which are unable to bind DNA (65; B.Rabault and J.G., unpublished data) **C.** Ets1M1 does not bind to DNA. EMSAs were carried out using 200 fmol of [^{32}P] labelled oligonucleotide A as a probe and either control extracts (lane 1), or extracts containing Ets1 (lane 2), or Ets1M1 (lane 3). Positions of the free probe and of the retarded complex are indicated by black and white arrowheads, respectively. A thin black arrow indicates a faint non specific complex observed with control extracts. **D.** Ets1 mutants bind specifically to DNA. EMSAs were carried out using 400 fmol of oligonucleotide A as a probe and either Ets1 (lanes 1−3), Ets1M2 (lanes 4−6), Ets1M3 (lanes 7−9) or Ets1M4 (lanes 10−12) containing extracts. Specific competitors were added as follows : no competitor (lanes 1, 4, 7 and 10); a 25 fold molar excess of oligonucleotide A (lanes 2, 5, 8 and 11); a 25 fold molar excess of oligonucleotide Z (lanes 3, 6, 9 and 12). Positions of the free probe and of the retarded complex are indicated by black and white arrowheads, respectively. Exposure time of the autoradiogram was three times longer for lanes 4−12 than for lanes 1−3.

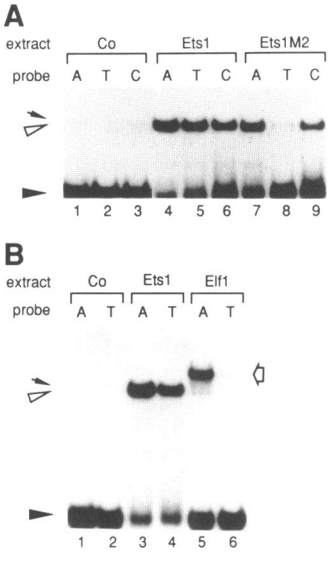

Figure 4. A. Effects of mutations within the A motif of CRIII. EMSAs were carried out using 200 fmol of either [^{32}P] labelled oligonucleotide A (lanes 1, 4, and 7) or T (lanes 2, 5 and 8) or C (lanes 3, 6, and 9) and either control extracts (lanes 1−3), or extracts containing Ets1 (lanes 4−6), or Ets1M2 (lanes 7−9). Positions of the free probe and of the retarded complex are indicated by black and white arrowheads, respectively. A thin black arrow indicates a faint non specific complex observed with control extracts. **B.** Elf1, unlike Ets1, binds poorly to a GGAT core containing EBS. EMSAs were carried out using 200 fmol of either [^{32}P] labelled oligonucleotide A (lanes 1, 3 and 5), or T (lanes 2, 4 and 6) and either control (lanes 1 and 2), Ets1 (lanes 3 and 4), or Elf1 (lanes 5 and 6) cell extracts. Positions of the free probe and of the Ets1−probe complex are indicated by black and white arrowheads, respectively. A thin black arrow indicates a faint non specific complex observed with control extracts. A thick white arrow indicates the position of the Elf1 specific complex.

binding to oligonucleotide A (Fig. 4A, lanes 7, 8). In that respect, Ets1M2 was found to behave similarly to Elf1 (Fig. 4B, lanes 5, 6). Of note, mutation M2 does not alter binding to oligonucleotide C (Fig. 4A, lanes 6 and 9). These results suggest that the residues within the CRIII motif A are responsible for the discrimination between GGAA and GGAT core-containing EBS.

The ability of Ets1M3 and Ets1M4 to bind to either a GGAA or a GGAT corecontaining EBS was next analyzed by direct electrophoretic mobility shift assays using oligonucleotides A and T as probes. Both Ets1 and Ets1M4 were found to bind in a similar way to oligonucleotide A and oligonucleotide T (Fig. 5A, lanes 1, 2 and 5, 6). In contrast, Ets1M3 bound inefficiently to oligonucleotide T although its binding to oligonucleotide A was similar to that of wild type Ets1 (Fig. 5A, lanes 3 and 4). To confirm these observations and to obtain relative estimates of the affinities of the respective proteins for each type of probes, competition binding studies were performed using oligonucleotide A as a probe. Binding of Ets1 to oligonucleotide A was efficiently competed out by a 5 fold molar excess of either oligonucleotides A or T (Fig. 5B lanes 1, 2 and 4). A 25 fold molar excess of either competitor resulted in an essentially complete disappearance of the specific Ets1−probe complex (Fig. 5B, lanes 3 and 5). Similar results were observed with Ets1M4 (Fig. 5B, lanes 11−15). In contrast, and consistent with the direct binding assay, binding of Ets1M3 to oligonucleotide A probe, although efficiently competed out by a 5 fold molar excess of unlabelled oligonucleotide A, was barely affected by the same amount of oligonucleotide T (Fig. 5B, compare lanes 7 and 9). Moreover, unlike what is observed for Ets1, a 25 fold molar excess of oligonucleotide T acted as a weaker competitor for the binding of Ets1M3 to oligonucleotide A than a 5 fold molar excess of oligonucleotide A (Fig. 5B, compare lanes 10 to 7, and 5 to 2), indicating that Ets1M3 has a 5 to 10 fold lower affinity for GGAT-based oligonucleotide T than for a matched GGAA EBS. Similar results were observed in competition assays using oligonucleotide C as a probe (data not shown). Of note, the decrease in binding affinity for oligonucleotide T was similar for Ets1M2 and Ets1M3, as compared to Ets1 (Figs. 4A and 5A, and data not shown). We conclude from these experiments that the K54T substitution in the CRIII region of Ets1 is sufficient to confer to that protein the ability to discriminate between GGAT and GGAA core-containing EBS *in vitro*.

In order to determine whether mutation K54T in the Ets1 DNA binding domain results in additional changes in DNA binding specificity, binding of Ets1 and Ets1M3 to an oligonucleotide

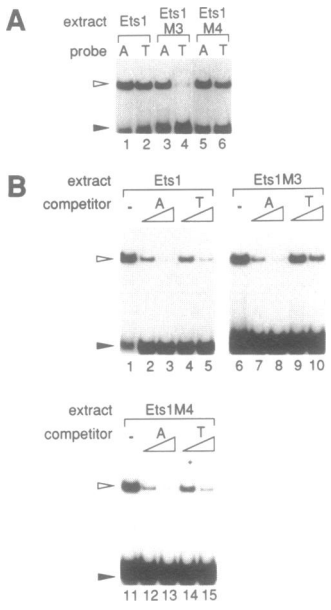

Figure 5. Mutation K54T impairs binding of Ets1 to GGAT core EBS. **A.** EMSAs were carried out using 200 fmol of either [^{32}P] labelled oligonucleotide A (lanes 1, 3 and 5), or T (lanes 2, 4 and 6) and extracts containing either Ets1 (lanes 1 and 2), Ets1M3 (lanes 3 and 4) or Ets1M4 (lanes 5 and 6). Positions of the free probe and of the retarded complex are indicated by black and white arrowheads, respectively. **B.** Competition binding assays. EMSAs were carried out using 400 fmol of oligonucleotide A as a probe and cells extracts containing either Ets1 (lanes 1−5), Ets1M3 (lanes 6−10) or Ets1 M4 (lanes 11−15). Specific competitors were added as follows: no competitor (lanes 1, 6 and 11); oligonucleotide A (lanes 2 and 3, 7 and 8, 12 and 13); oligonucleotide T (lanes 4 and 5, 9 and 10, 14 and 15). Binding reactions were performed in the presence of a 5 fold molar excess of competitor in lanes 2, 4, 7, 9, 12 and 14, and of a 25 fold molar excess in lanes 3, 5, 8, 10, 13 and 15. Positions of the free probe and of the retarded complex are indicated by black and white arrowheads, respectively. Exposure time of the autoradiogram was three times longer for lanes 6−15 than for lanes 1−5.

A probe was competed by a series of GGAA based EBS which differ from oligonucleotide A in sequences flanking the 5′ (oligonucleotides C and E) or both the 5′ and 3′ sides of the core (oligonucleotide D, F and Pu; see Fig. 1B for sequences). As shown in Fig. 6, the competition pattern for both Ets1 and Ets1M3 were similar, irrespective of the nature of the competing oligonucleotide used. Of note, oligonucleotide F which has been shown to bind Elf1 but not Ets1 because of the presence of a purine at position +6 (29) was also found unable to bind to Ets1M3 (Fig. 6, lanes 8 and 18). We conclude from these experiments that the K54T mutation in the Ets1 CRIII domain specifically affects the recognition of position +4 of the core sequence.

To further confirm the role of residue 54 of the Ets domain in the discrimination between GGAA and GGAT core-containing EBS, we replaced Elf1 threonine 54 by a lysine residue, thereby generating mutant Elf1T54K. Both wild-type Elf1 and the mutant protein were expressed at similar levels in rabbit reticulocyte lysates following *in vitro* transcription/translation of the respective cDNAs (Fig. 7A). Both proteins were found to bind with the same efficiency to an oligonucleotide A probe (Fig. 7B, lanes 2 and 7) and this binding was competed in a dose dependent manner by addition of unlabelled oligonucleotide A as competitor. In contrast, binding of Elf1T54K was more efficiently competed by the oligonucleotide T competitor than that of wild type Elf1.

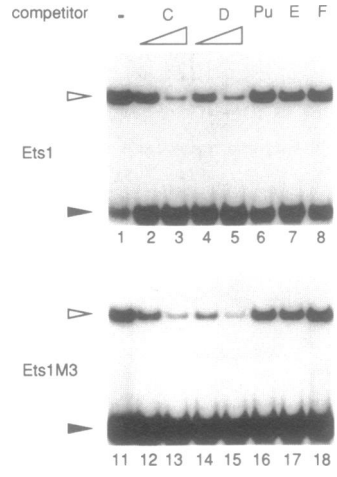

Figure 6. Mutation K54T does not affect Ets1 binding to several GGAA-based EBS. EMSAs were carried out using 400 fmol of oligonucleotide A as a probe and extracts containing either Ets1 (lanes 1−8), or Ets1M3 (lanes 11−18). Specific competitors were added as follows: no competitor (lanes 1 and 11); oligonucleotide C (lanes 2, 3, 12 and 13); oligonucleotide D (lanes 4, 5, 14 and 15); oligonucleotide Pu (lanes 6 and 16), oligonucleotide E (lanes 7 and 17); oligonucleotide F (lanes 8 and 18). A 5 fold molar excess of competitor was used in reactions shown in lanes 2, 4, 12 and 14, and a 25 fold molar excess in reactions shown in lanes 3, 5−8, 13 and 15−18. Positions of the free probe and of the retarded complex are indicated by black and white arrowheads, respectively. Exposure time of the autoradiogram was three times longer for lanes 11−18 than for lanes 1−8.

Specifically, unlike the situation observed for Elf1, a 20 fold molar excess of oligonucleotide T acted as a more efficient competitor for the binding of Elf1T54K to the probe than a 5 fold molar excess of oligonucleotide A (Fig. 7B, compare lanes 2−6, to 8−11). In line with these results, only Elf1T54K was able to bind efficiently to an oligonucleotide T probe (data not shown). We conclude from these experiments that the T54K mutation in the CRIII region of Elf1 impairs its ability to discriminate between GGAA and GGAT core containing EBS.

DISCUSSION

To determine which region(s) of the Ets domain are actually involved in DNA recognition, we have generated *in vitro* mutagenized versions of Ets1 and examined their ability to bind DNA. With respect to the identification of residues closely contacting DNA, DNA binding defective mutants are only poorly informative since point mutation can abolish DNA binding not only by affecting residues involved in specific protein−DNA contacts but also through indirect effects related to disruption of aminoacid side chain interactions essential to proper folding of the protein. In contrast, change of specificity mutants which result in a differential selectivity among related binding sites, are more valuable. Indeed, although such mutations could affect residues which contact backbone phosphate groups, as in the case of bZip proteins (46), comparison of genetic and structural data collected from the study of either procaryotic (see Ref. 47 for review) or eucaryotic (45, 48−55) protein−DNA interactions indicates that change of specificity mutations usually affect residues interacting with specific base pairs in the recognition sequence. Results obtained in these structurally characterized systems therefore provide a rationale to use change of specificity mutants as a

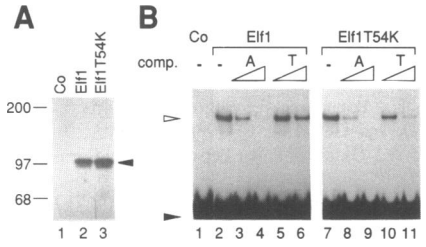

Figure 7. A. Expression of Elf1 and Elf1T54K synthetized in rabbit reticulocyte lysates. *In vitro* transcription/translation reactions were performed in the presence of 1 μCi/μl of L[^{35}S] methionine, and subjected to SDS polyacrylamide gel electrophoresis as described previously (4). Position of molecular weight markers is shown at the left of the figure. **B.** Mutation T54K in Elf1 impairs its selectivity for GGAA based EBS. EMSAs were carried out using 100 fmol of oligonucleotide A as a probe and either a non programmed *in vitro* transcription/translation reaction (lane 1), or *in vitro* synthetized Elf1 (lanes 2–6) or Elf1T54K (lanes 7–11). Specific competitors were added as follows: no competitor (lanes 1, 2 and 7); oligonucleotide A (lanes 3 and 4, 8 and 9); oligonucleotide T (lanes 5 and 6, 10 and 11). Binding reactions were performed in the presence of a 5 fold molar excess of competitor in lanes 3, 5, 8, 10, and of a 20 fold molar excess in lanes 4, 6, 9, 11. Positions of the free probe and of the retarded complex are indicated by black and white arrowheads, respectively.

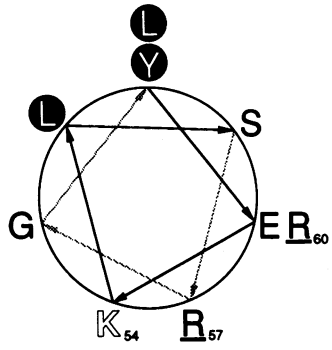

Figure 8. Helical wheel projection of aminoacid residues 52–60 of the Ets1 Ets domain. Hydrophobic residues are shown as white letters on a black background. Arginines 57 and 60 (underlined) and lysine 54 (outlined) are indicated. For clarity, the helix has been drawn with a 3.5 residues per turn path.

genetic approach towards the identification of residues in close contact with DNA, prior to structural analysis of a DNA–protein complex.

The results we report here describe the first change of specificity mutants for an Ets domain protein. Within a highly conserved subdomain of the Ets1 DNA binding domain, we have substituted a single residue (K54 in Ets1) which is conserved in all Ets proteins except Elf1 and E74 into the threonine residue found in the corresponding position in the latter proteins. This substitution converts the DNA binding properties of Ets1 from a GGAA/GGAT core mixed specificity to a highly preferential GGAA core specificity, a feature characteristic of both Elf1 and E74. This change of specificity appears to be specific to the K54T substitution since mutations located immediately upstream of position 54 (mutant Ets1M4) do not affect GGAA *vs* GGAT selectivity of the mutant protein. Furthermore, the change of specificity resulting from the K54T substitution appears to be restricted to position +4 of the EBS, since it does not alter recognition of other GGAA based EBS, differing one from another by changes outside of the GGAA core. Finally, the reciprocal substitution (T54K) in the Elf1 Ets domain releases its selectivity for GGAA core-containing EBS and allows the mutant protein to bind to GGAT core-containing EBS. These results suggest that residue 54 is actually in close proximity of base pair +4 of the EBS core and that the CRIII region mediates the specific interaction of the Ets domain with the EBS core. In line with this proposal, motif B of CRIII is highly conserved phylogenetically in all Ets proteins, a property expected for a protein domain involved in recognition of a conserved DNA sequence. Also consistent with this proposal, mutation of two invariant residues in CRIII motif B (R57 and R60) completely abolishes Ets1 specific DNA binding.

Although definitive elucidation of how Ets proteins interact with DNA will obviously require structural analysis of an EBS-bound Ets domain, our results and secondary structure prediction of the CRIII-containing subdomain of Ets proteins allow further speculation on this point to be made. We took advantage of the 'mpred' method (56) which allows an accurate prediction to be made through the combined analysis of several Ets protein sequences, assuming that the folding pattern of the CRIII region is conserved among all Ets proteins. Using this method, this region of the Ets domain was predicted to form an amphipathic α-helix, encompassing both the CRIII region and the three conserved tyrosine residues, located immediately downstream of CRIII (see Fig. 2). Essentially similar predictions were achieved when combining results provided by two distinct procedures (Chou-Fasman [57] and GOR [58] analyses) on five Ets proteins (Ets1, GABPα, Fli1, E74 and Spi1). As illustrated in the helical wheel projection shown in Fig. 8, Ets1 residues K54, R57 and R60 are predicted to lie on the same hydrophilic side of this putative α-helix, as would be expected if they were involved in direct interaction with DNA. On the other hand, methylation interference studies (16, 23, 27, 39, 43) have clearly established that methylation of G+1 and G+2 within the EBS core drastically inhibits Ets1, GABPα and Spi1/Pu1 DNA binding to their cognate sites, indicating that interactions with these bases occur in the major groove of the DNA helix. These major groove interactions could extend in 5′ flanking sequences up to nucleotide −2, at least in a subset of EBS, but are unlikely to extend further than nucleotide +4 at the 3′ side of the EBS (16, 23, 27, 39, 43). Together with these biochemical data, our results and secondary structure predictions lead to a model of Ets domain–EBS core interaction in which the putative α-helix centered over CRIII would be docked in the DNA major groove. This would allow direct contact between DNA base pairs and/or phosphate and CRIII residues positionned on the hydrophilic side of the helix. In this context, properties of mutant Ets1M3 suggest that residue 54 could interact with base pair +4. The strict conservation of both arginine residues 57 and 60 in all known Ets proteins would make them good candidates for interacting with either G+1 or G+2. Of note, arginine–guanine interactions have been observed in several structurally resolved DNA–protein complexes (see Ref. 59 for review). Alternatively, these residues could be involved either in major groove contacts with base pairs −2, −1 or +3, or interact with backbone phosphates which have been shown to be in close contact with DNA boundEts1 (23). That either R57 or R60 are directly involved in DNA binding is also consistent with the complete loss of DNA binding of our mutant Ets1M1, in which both arginines were changed into leucines. That leucines are often found in α-helices suggests that the R to L substitution does not abolish specific DNA binding through disruption of the putative CRIII α-helix.

Such a model would be consistent with two apparently general features which have been observed in most structurally analysed

protein–DNA complexes (see Ref. 59 for review): (i) aminoacids involved in direct interactions with either bases or phosphates often take place in a well defined secondary structure (most frequently an α-helix in examples analyzed so far) and (ii) these sequence specific interactions frequently result from positionning of such an helix in the DNA major groove, where more base-pair discriminatory information is available than in the minor groove (60). An α-helix positionned in the DNA major groove mediating interactions with specific base pairs is highly reminiscent of the 'recognition helix' found in both procaryotic HTH and eucaryotic homeodomain DNA binding motifs. In this respect, secondary structure predictions suggest that Ets domain residues 35 to 43 could also fold as an α-helix. Moreover, it has recently been suggested that residues 35 to 64 in the Spi1/Pu1 Ets domain could adopt a HTH like conformation, the second ('recognition') helix of which would be contributed by the CRIII motif described here (61).

Although we favour the above discussed model for the geometry of Ets domain–EBS interaction, one obviously cannot exclude other explanations for the phenotype induced by K54T substitution. Beside more indirect effects, residue 54 could specifically interact with EBS base pair +4 in a different way. In particular, Nye *et al.* observed a methylation interference affecting the adenine matched to thymidine +4 in a GGAT core EBS (23), therefore suggesting the possibility that Ets1 recognition of base pair +4 takes place in the minor groove. One should note however that such a methylation interference could be a consequence of a methylation-induced local DNA conformational change rather than indicating direct interaction between base pair +4 and the Ets1 protein in the minor groove. In line with this, methylation of the third (underlined) adenine in the sequence TAATGG has been shown to interfere with binding of two homeodomain proteins (Antp and Ftz) (62), although both NMR analysis of an Antp–DNA complex (55) and X-ray analysis of an Engrailed-DNA cocrystal (54) show that this base pair is involved in a major groove contact with residues located in the third helix of the homeodomain. Moreover, minor groove contacts provide poor discrimination between AT and TA base pairs, since in B DNA conformation, neither hydrogen bonding nor van der Waals interactions between aminoacids methyl groups and thymidine appear to be able to discriminate between these base pairs (60).

ACKNOWLEDGEMENTS

We thank Dr J.M.Leiden for the *elf1* cDNA, Drs E.Quiniou and C.Vesque for helpful discussions and Drs P.Charnay, G.Craescu, A.Gégonne and J.Nardelli for critical reading of the manuscript. This work was supported by funds from the Centre National de la Recherche Scientifique, Institut Curie, Ligue Nationale Française contre le Cancer, Agence Nationale de Recherches sur le SIDA and Association pour le Recherche sur le Cancer. During the earlier part of this work, R.B. was supported by a fellowship from the Agence Nationale de Recherches sur le SIDA.

REFERENCES

1. Ghysdael, J. and Yaniv, M. (1991) *Curr. Opin. Cell. Biol.*, 3, 484–492.
2. Wasylyk, B., Hahn, S.L. and Giovane, A. (1993) *Eur. J. Biochem.*, 211, 7–18.
3. Watson, D.K., Smyth, F.E., Thompson, D.M., Cheng, J.Q., Testa, J.R., Papas, T.S. and Seth A. (1992) *Cell Growth & Differentiation*, 3, 705–713.
4. Ghysdael, J., Gégonne, A., Pognonec, P., Dernis, D., Leprince, D. and Stehélin, D. (1986) *Proc. Natl. Acad. Sci. USA*, 83, 1714–1718.
5. Duterque-Coquillaud, M., Leprince, D., Flourens, A., Henry, C., Ghysdael, J., Debuire, B. and Stéhelin, D. (1988) *Oncogene Res.*, 2, 335–344.
6. Boulukos, K., Pognonec, P., Begue, A., Galibert, F., Gesquière, J.C., Stéhelin, D. and Ghysdael, J. (1988) *EMBO J.*, 7, 697–705.
7. Ben-David, Y., Giddens, E.B., Letwin, K. and Bernstein, A. (1991) *Genes & Dev.*, 5, 908–918.
8. Reddy, E.S.P., Rao, V.N. and Papas, T.S. (1987) *Proc. Natl. Acad. Sci. USA*, 84, 6131–6135.
9. LaMarco, K., Thompson, C.C., Byers, B.P., Walton, E.M. and McKnight, S.L. (1991) *Science*, 253, 789–792.
10. Xin, J.H., Cowie, A., Lachance, P. and Hassell, J.A. (1992) *Genes & Dev.*, 6, 481–496.
11. Brown, T.A. and McKnight, S.L.(1992) *Genes & Dev.*, 6, 2502–2512.
12. Leiden, J.M., Wang, C.Y., Petryniak, B., Markovitz, D.M., Nabel, G.J. and Thomspon, C.B. (1992) *J. Virol.*, 66, 5890–5897.
13. Rao, V.N., Huebner, K., Isobe, M., Ar-Rushdi, A., Croce, C.M. and Reddy, E.S.P. (1989) *Science*, 244, 66–70.
14. Dalton, S. and Treisman, R. (1992) *Cell*, 68, 597–612.
15. Moreau-Gachelin, F., Tavitian, A. and Tambourin, P. (1988) *Nature*, 331, 277–280.
16. Klemsz, M.J., McKercher, S.R., Celada, A., Van Beveren, C. and Maki, R.A. (1990) *Cell*, 61, 113–124.
17. Ray, D., Bosselut, R., Ghysdael, J., Mattei, M.G., Tavitian, A. and Moreau-Gachelin, F.(1992) *Mol. Cell. Biol*, 12, 4297–4304.
18. Pribyl, L.J., Watson, D.K., Schulz, R.A. and Papas, T.S. (1991) *Oncogene*, 6, 1175–1183.
19. Lai, Z.C. and Rubin, G.M. (1992) *Cell*, 70, 609–620.
20. Tei, H., Nihonmatsu, I., Yokokura, T., Ueda, R., Sano, Y., Okuda, T., Sato, K., Hirata, K., Fujita, S.C. and Yamamoto, D. (1992) *Proc. Natl. Acad. Sci. USA*, 89, 6856–6860.
21. Burtis, K.C., Thummel, C.S., Jones, C.W., Karim, F.D. and Hogness, D.S. (1990) *Cell*, 61, 85–99.
22. Woods, D.B., Ghysdael, J. and Owen, M.J. (1992) *Nucleic Acids Res.*, 20, 699–704.
23. Nye, J.A., Petersen, J.M., Gunther, C.V., Jonsen, M.D. and Graves, B.J. (1992) *Genes & Dev.*, 6, 975–990.
24. Fisher, R.J., Mavrothalassitis, G., Kondoh, A. and Papas, T.S. (1991) *Oncogene*, 6, 2249–2254.
25. Wasylyk, C., Gutman, A., Nicholson, R. and Wasylyk, B.(1991) *EMBO J.*, 10, 1127–1134.
26. Ho, I.C., Bhat, N.K., Gottschalk, L.R., Lindsten, T., Thompson, C.B., Papas, T.S. and Leiden, J.M. (1990) *Science*, 250, 814–818.
27. Gitlin, S.D., Bosselut, R., Gégonne, A., Ghysdael, J. and Brady, J.N.(1991) *J. Virol.* 65, 55135523.
28. Treisman, R., Marais, R. and Wyneen, J. (1992) *EMBO J.*, 11, 4631–4640.
29. Wang, C.Y., Petryniak, B., Ho, I.C., Thompson, C.B. and Leiden, J.M. (1992) *J. Exp. Med.*, 175, 1391–1399.
30. Urness, L.D. and Thummel, C.S. (1990) *Cell*, 63, 47–61.
31. Bosselut, R., Duvall, J.F., Gégonne, A., Bailly, M., Hemar, A., Brady, J. and Ghysdael, J. (1990) *EMBO J.*, 9, 3137–3144.
32. Wasylyk, B., Wasylyk, C., Flores, P., Begue, A., Leprince, D. and Stehélin, D. (1990) *Nature*, 346, 191–193.
33. Gégonne, A., Bosselut, R., Bailly, R.-A. and Ghysdael J. (1993) *EMBO J.* 12, 1169–1178.
34. Hill, C.S., Marais, R., John, S., Wynne J., Dalton, S. and Treisman, R. (1993) *Cell*, 73, 395–406.
35. Pongubala, J.M.R., Nagulapalli, S., Klemsz, M.J., McKercher, S.R., Maki, R.A. and Atchison, M.L. (1992) *Mol. Cell. Biol.*, 12, 368–378.
36. Gégonne, A., Punyammalee, B., Rabault, B., Bosselut, R., Seneca, S., Crabeel, M. and Ghysdael, J. (1992) *New Biol.*, 4, 512–519.
37. Janknecht, R. and Nordheim, A. (1992) *Nucleic Acids Res.*, 20, 3317–3324.
38. Lim, P., Kraut, N., Frampton, J. and Graf, T. (1992) *EMBO J.*, 11, 643–652.
39. Thompson, C.C., Brown, T.A. and McKnight, S.L. (1991) *Science*, 253, 762–768.
40. Anton, I.A. and Frampton, J. (1988) *Nature*, 336, 719.
41. Landt, O., Grunert H.P. and Hahn, U.(1990) *Gene*, 96, 125–128.
42. Klemsz, M.J., Maki, R.A., Papayannopoulou, T., Moore, J. and Hromas, R. (1993) *J. Biol. Chem.*, 268, 5769–5773.
43. Gunther, C.V., Nye, J.A., Bryner, R.S. and Graves, B.J. (1990) *Genes & Dev.*, 4, 667–679.
44. Pabo, C.O., Aggarwal, A.K., Jordan, S.R., Beamer, L.J., Obeysekare, U.R. and Harrison, S.C. (1989) *Science*, 247, 1210–1213.
45. Wollberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O. (1991) *Cell*, 67, 517–528.

46. Suckow, M., Wilcken-Bergmann, B. and Müller-Hill, B. (1993) *EMBO J.*, 12, 1193–1200.
47. Harrison, S.C. and Aggarwal, A.K. (1990) *Ann. Rev. Biochem.*, 59, 933–969.
48. Nardelli, J., Gibson, T.J., Vesque, C. and Charnay, P. (1991). *Nature*, 349, 175–178.
49. Nardelli, J., Gibson, T. and Charnay, P. (1992) *Nucleic Acids Res.*, 20, 4137–4144.
50. Desjarlais, J.R. and Berg, J.M. (1992) *Proc. Natl. Acad. Sci. USA*, 89, 7345–7349.
51. Pavletich, N.P. and Pabo, C.O. (1991) *Science*, 252, 809–817.
52. Hanes, S.D. and Brent, R. (1989) *Cell*, 57, 1275–1283.
53. Treisman, J., Conczy, P., Vashishta, M., Harris, E. and Desplan, C. (1989) *Cell*, 59, 553–562.
54. Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) *Cell*, 63, 579–590.
55. Otting, G., Qiu Qian Y., Billeter, M., Müller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1990) *EMBO J* , 9, 3085–3092.
56. Levin, J.M., Pascarella, S., Argos, P. and Garnier, J. (1993) *Protein Eng.*, in press.
57. Chou, P.Y., and Fasman, G.D. (1978) *Ann. Rev. Biochem.*, 47, 251–276.
58. Gibrat, J.F., Garnier, J., and Robson, B. (1987) *J. Mol. Biol.*, 198, 425–443
59. Pabo, C.O. and Sauer, R.T. (1992) *Ann. Rev. Biochem.*, 61, 1053–1095.
60. Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976) *Proc. Nat. Acad. Sci. USA*, 73, 804–808.
61. Galson, D.L., Hensold, J.O., Bishop, T.R., Schalling, M., D'Andrea, A.D., Jones, C., Auron, P.E. and Housman, D.E. (1993) *Mol. Cell. Biol., 1993*, 13, 2929–2941.
62. Percival-Smith, A., Müller, M., Affolter, M. and Gehring, W.J. (1990) *EMBO J.*, 9, 3967–3974.
63. Cross, S.L., Feinberg, M.B., Wolf, J.B., Holbrook, N.J., Wong-Staal, F. and Leonard W.J. (1987) *Cell*, 49, 47–56.
64. Durand, D.B., Shaw, J-P., Bush, M.R., Replogle, R.E., Belagaje, R. and Crabtree, R. (1988) *Mol. Cell. Biol.*, 8, 1715–1724.
65. Pognonec, P., Boulukos, K. and Ghysdael, J. (1989) *Oncogene*, 4, 691–697.