The EMBO Journal vol.6 no.10 pp.3005–3013, 1987

One cell-specific and three ubiquitous nuclear proteins bind *in vitro* to overlapping motifs in the domain B1 of the SV40 enhancer

J.H.Xiao, I.Davidson, D.Ferrandon, R.Rosales, M.Vigneron, M.Macchi, F.Ruffenach and P.Chambon

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, Strasbourg, France

Communicated by P.Chambon

We have used the gel retardation assay to investigate the binding of nuclear proteins to the domain B1 of the SV40 enhancer, which contains the GT-II motif. Four proteins (GT-IIA, GT-IIBα, GT-IIBβ and GT-IIC) were detected, three of which were present in nuclear extracts from several cell lines. The fourth protein (GT-IIC) showed a clear cell-specificity, being absent from the lymphoid cell extracts tested. The results of methylation interference assays and of the binding of the proteins to mutated templates indicate that the domain B1 contains three distinct, but overlapping, protein-binding motifs (GT-IIA, B and C). The cell-specific binding of protein GT-IIC *in vitro* correlates with the *in vivo* enhancer activity of its cognate motif, strongly suggesting that this protein acts as a positive *trans*-acting enhancer factor. Two of the proteins also recognize other enhancer motifs; protein GT-IIBα binds to the μE3 motif present in the immunoglobulin heavy chain enhancer; protein GT-IIC binds to an enhancer motif of the polyomavirus mutant PyEC9.1 adapted to growth in F9 embryonal carcinoma cells, but not to the corresponding wild-type sequence.

Key words: SV40 enhancer/domain B1/nuclear proteins

Introduction

Enhancers are important *cis*-acting control elements which stimulate transcription *in vivo* from RNA polymerase class B(II) promoters, irrespective of their orientation and to some extent of their location (Moreau *et al.*, 1981; Banerji *et al.*, 1981; for reviews see Yaniv, 1984; Chambon *et al.*, 1984; Serfling *et al.*, 1985; Picard, 1985; Sassone-Corsi and Borrelli, 1986; Wasylyk, 1986). *In vivo* competition and 'footprinting' experiments have shown that their activity is mediated by *trans*-acting factors (Schöler and Gruss, 1984; Borrelli *et al.*, 1984; Mercola *et al.*, 1985; Ephrussi *et al.*, 1985; Schöler *et al.*, 1986; Wasylyk *et al.*, 1987). Several enhancers such as those of the immunoglobulin heavy chain (IgH) (Banerji *et al.*, 1983; Gillies *et al.*, 1983) and kappa light chain genes (Queen and Baltimore, 1983; Picard and Schaffner, 1984; Atchinson and Perry, 1987), lymphotropic papovavirus (Mosthaf *et al.*, 1985) and the insulin gene (Walker *et al.*, 1983) exhibit cell-type specificity, suggesting that some of the *trans*-acting factors necessary for their activity are absent from some cell types. Others, typically the simian virus 40 (SV40) enhancer, are active in a variety of cell-types (Schirm *et al.*, 1987 and references therein, and see reviews above). *In vivo* studies have shown that the SV40 enhancer is composed of multiple sequence motifs which act synergistically to generate enhancer activity in HeLa cells (Zenke *et al.*, 1986; Herr and Clarke, 1986). These seqence motifs, GT-II, GT-I, TC-II, Sph-I, Sph-II and P (see Figure 1) are organized in two domains, A and B, whose multimerization generates enhancer activity (Zenke *et al.*, 1986). The B domain can be further subdivided into two apparently independent subregions B1 and B2 (Figure 1; Zenke *et al.*, 1986; Wildeman *et al.*, 1986).

The activity of several enhancers, including that of SV40, has also been partially reproduced *in vitro* (Sassone-Corsi *et al.*, 1984, 1985; Wildeman *et al.*, 1984; Sergeant *et al.*, 1984; Schöler and Gruss, 1985; Augereau and Chambon, 1986; Dougherty *et al.*, 1986). *In vitro* DNase I footprinting studies have shown that each of the SV40 sequence motifs binds a protein(s) present in HeLa cell nuclear extracts (Wildeman *et al.*, 1986; Davidson *et al.*, 1986). The binding of these proteins is prevented by mutations which are detrimental to enhancer activity *in vivo*, suggesting that they correspond to *trans*-acting factors responsible for the *in vivo* stimulation. Subsequently, it has been demonstrated that the enhancer activity is mediated by a different set of motifs and proteins in HeLa and mouse myeloma MPC11 cells (Davidson *et al.*, 1986; Nomiyama *et al.*, 1987). This cell-type specific recognition is in part achieved by the differential use of the overlapping Sph and octamer motifs in enhancer domain A. A differential pattern of protein binding was also noticed in the enhancer B1 domain, which contains the GT-II motif (Davidson *et al.*, 1986, see Figure 1). We have studied here the binding of proteins present in nuclear extracts from several cell-types to the B1 domain. We show that this region contains several overlapping sequences which interact with distinct nuclear proteins. One of these sequences binds a cell-type specific protein whose presence correlates with the activity of the B1 domain *in vivo*.

Results

Several proteins bind to the enhancer B1 domain

To study the interaction of nuclear proteins with the region of the SV40 enhancer containing motif GT-II, synthetic oligonucleotides containing the wild-type sequence from nucleotides 257–294 (BBB numbering system, Tooze, 1982) were cloned between the *Kpn*I and *Bam*HI sites of a derivative of pUC18 (pGT2-0, Figure 2 and Materials and methods). The SV40 sequence was mutated at positions 268–270 (pGT2-1) or 281–283 (pGT2-2) or simultaneously at both positions (pGT2-3, Figure 2). The mutations in pGT2-1 and pGT2-2 correspond to those present in recombinants pA4 and pA211 (Zenke *et al.*, 1986, see Figure 1). The mutation in pA4, but not in pA211, has a marked deleterious effect on enhancer activity *in vivo* in HeLa cells, while both prevent the binding *in vitro* of different proteins present in HeLa cell extracts (Wildeman *et al.*, 1986).

The [^{32}P]-5′ end-labelled *Kpn*I–*Bam*HI fragments from the pGT2 series plasmids were incubated with HeLa cell nuclear extracts in the presence of the synthetic duplexes poly[dI–dC] or poly[dA–dT] to suppress 'non-specific' interactions (see Materials and methods; hereafter pGT2-0 to pGT2-3 refer to the labelled *Kpn*I–*Bam*HI templates). The resulting labelled DNA–protein complexes were resolved on a non-denaturing poly-

© IRL Press Limited, Oxford, England

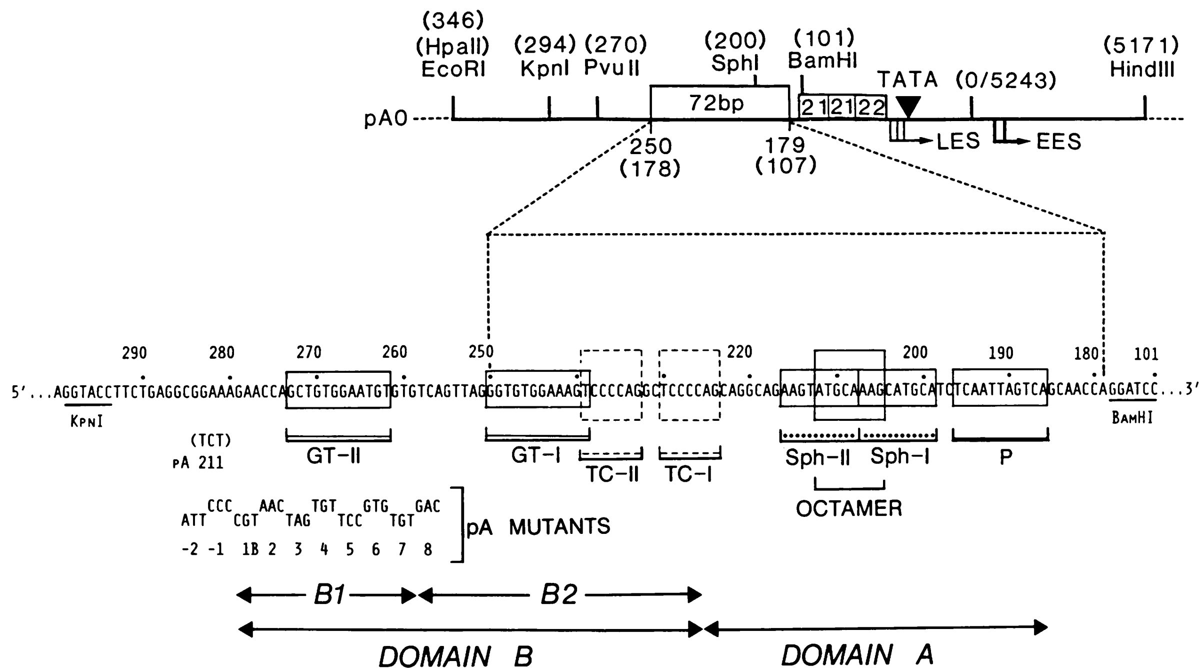

Fig. 1. Organization of the SV40 early promoter. The top line shows the structure of the SV40 early promoter in plasmid pA0, containing a single copy of the 72 bp sequence (Zenke *et al.*, 1986). The position of the essential elements of the promoter, the TATA box, the 21 bp repeat element [22,21,21] and 72 bp enhancer element are indicated along with the coordinates (BBB system, Tooze, 1982) of several natural or engineered restriction enzyme recognition sites. EES and LES indicate the position of the early-early and late-early mRNA start sites, respectively. The second line shows the sequence of the late coding strand of the 72 bp element (indicated by the vertical broken lines) and the 5'-flanking sequences. The location of the sequence motifs which have been identified in the enhancer region (see text and Zenke *et al.*, 1986) is indicated by the boxes, along with the positions of the A, B1 and B2 enhancer domains. 'pA mutants' refers to the late coding strand sequences present in the mutated recombinants pA-2 to pA8 and pA211 (see Zenke *et al.*, 1986; Nomiyama *et al.*, 1987).

acrylamide gel (Fried and Crothers, 1981; Garner and Revzin, 1981; Strauss and Varshavsky, 1984). Incubation of the wild-type template pGT2-0 with the HeLa cell nuclear extract and poly[dA−dT] generated four major retarded complexes (A−D in Figure 3A). The mutation in pGT2-1 prevented formation of complexes B and C, while complex A was dramatically reduced by the mutation in pGT2-2 (complex B consists of two closely migrating species α and β, and for the sake of simplicity is referred to as B unless the two components show different behaviours, see below). Using pGT2-3, only complex D was efficiently formed. These results indicate that three specific complexes (A−C) are formed with the B1 domain, while D corresponds to a 'non-specific' complex (i.e. a complex not affected by the mutations in pGT2-3). In addition, the proteins forming complexes B and C appear to interact with the same or closely overlapping sequences as both are affected by the mutation in pGT2-1 (compare pGT2-0 with pGT2-1 in Figure 3A). Using poly[dI−dC] as competitor, complexes A, C and D were efficiently formed; however complex B was dramatically reduced (compare HeLa cell results in Figure 3A and B).

Complexes analogous to A and B were observed using extracts from non-differentiated mouse F9 embryonal carcinoma cells [F9(ND) cells], the human Burkitt lymphoma-derived lymphoblastoid cell line BJA-B (Klein *et al.*, 1975), the human T cell line Molt-4 (Minowada *et al.*, 1972), the mouse myeloma MPC11, the mouse pre-B cell line 70Z/3 (Paige *et al.*, 1978) and the simian kidney cell line CV1 (Figure 3A and B and data not shown). A minor difference however was observed using MPC11 extracts, where complex A had an altered electrophoretic mobility. Interestingly, complex B was not found in extracts from retinoic acid-treated (differentiated) F9 cells (data not shown). In addition to these specific complexes, several complexes not affected by the mutations in pGT2-3 were formed in the different cell extracts. In each cell extract the formation of complexes A and B exhibited the same sequence requirements as observed in HeLa cell extracts (compare pGT2-0 to pGT2-3 in each case).

Complexes analogous to C were formed using non-differentiated and differentiated F9 or CV1 cell extracts (Figure 3A and B and data not shown). In non-differentiated F9 cell extracts, complex C is slightly obscured by a 'non-specific' band (broken arrow in Figure 3A and B). The presence of complex C was, however, confirmed using the [^{32}P]-5' end-labelled template OGT2-50 (Figure 2, see below and data not shown). In contrast to complexes A and B, no complex C was observed with the lymphoid cell extracts (BJA-B, MPC11, Molt-4 and 70Z/3, Figure 3A and data not shown). These results suggest that this region of the SV40 enhancer may interact with three distinct nuclear proteins, one of which (C) exhibits cell-specificity. In order to test this hypothesis, we attempted to resolve these proteins by ion-exchange chromatography.

HeLa or BJA-B cell nuclear extracts were applied to a heparin agarose column. The retained proteins were eluted using a linear 0.2−0.6 M KCl gradient (Materials and methods) and fractions analysed using the pGT2-0 template. The peak of HeLa cell proteins forming complexes A and C eluted between 0.30−0.35 and 0.33−0.36 M KCl, respectively (Figure 4A). Complex B was eluted as two separate peaks (α and β) at 0.36−0.40 and 0.31−0.36 M KCl, respectively. In addition, several other minor complexes (Figure 4A) were observed. These complexes are 'non-specific' as their formation was not prevented by the mutations in pGT2-3 (data not shown).

The BJA-B proteins forming complexes A and B eluted at the same KCl concentration as their HeLa cell counterparts indicating that they are probably identical in the two cell-types (compare

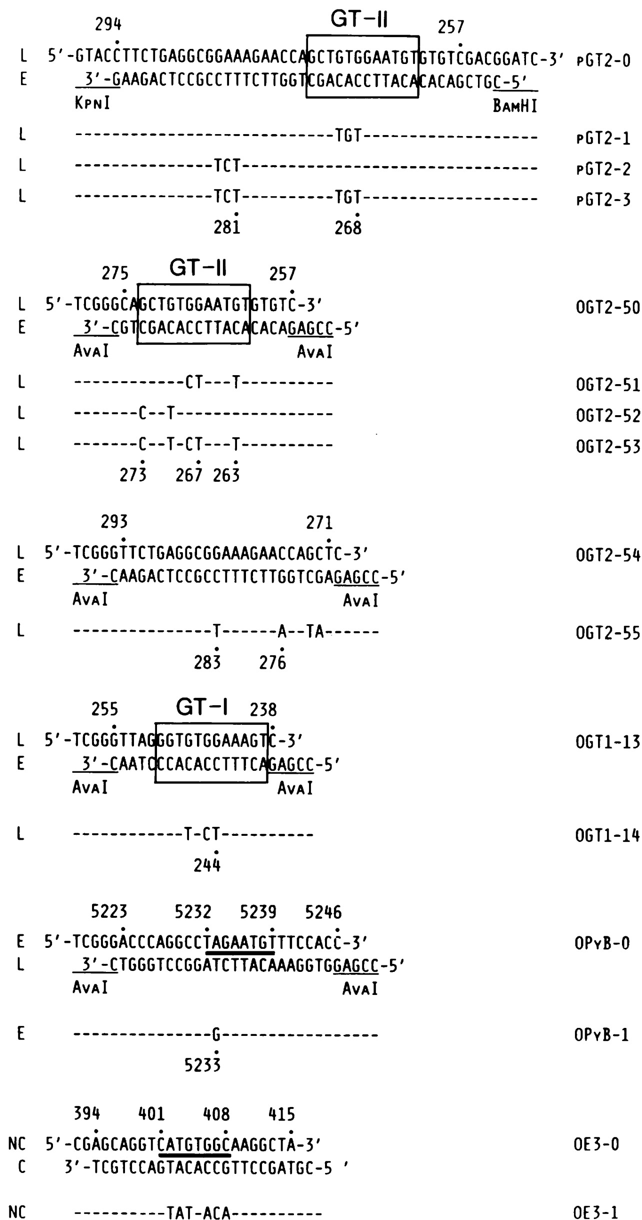

Fig. 2. Sequences of enhancer motif-containing templates. The top line shows the wild type SV40 sequence (bases 257–294) with the addition of the *Bam*HI and *Kpn*I restriction sites (see text), which constitutes the templated pGT2-0. E and L refer to the early and late coding strands, and the numbers above the sequence, to the coordinates in the SV40 early promoter. The positions, sequences and coordinates of the point mutations which are present in templates pGT2-1, -2 and -3 are indicated below the pGT2-0 sequence. The sequences of the other wild-type SV40 templates (OGT2-50, OGT2-54, OGT1-13) and their corresponding mutants (OGT2-51 to OGT2-53, OGT2-55 and OGT1-14) are indicated in the same way as that of pGT2-0. In each case, the locations of the GT-II or GT-I motifs are indicated. OPyB-0 and OPyB-1 contain, respectively, the wild-type and mutant EC9.1 polyoma virus enhancer B domain sequences from positions 5223–5246. The solid line between the two strands indicates the position of the homology with the SV40 GT-IIC motif. E and L refer to the polyoma virus early and late mRNA coding strands. OE3-0 and OE3-1 contain the murine IgH enhancer sequences from positions 394 to 415 (numbering system of Ephrussi *et al.*, 1985) in either a wild-type or mutated form. The solid line indicates the position of the μE3 motif (Ephrussi *et al.*, 1985; Sen and Baltimore, 1986a). NC and C refer to the heavy chain mRNA coding and non-coding strands, respectively.

A and B in Figure 4A and B). In agreement with the results obtained using the total extracts, no complex C was formed in any of the eluted fractions (Figure 4B and data not shown). Similar results were obtained with MPC11 extracts (data not shown).

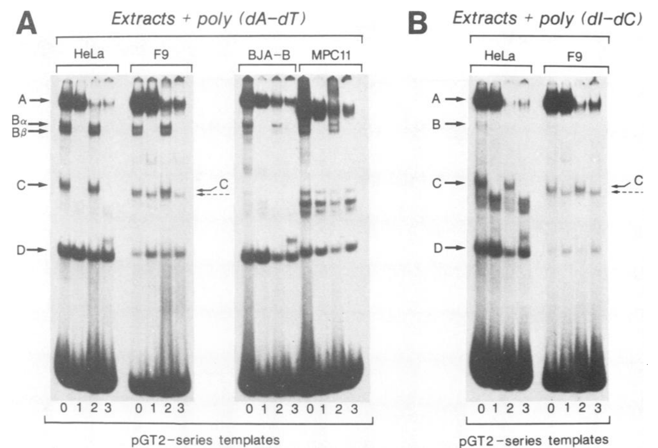

Fig. 3. *In vitro* protein binding to the enhancer B1 domain. **Panel A** shows the set of labelled protein–DNA complexes formed with nuclear extracts from various cell types and the pGT2-series templates in the presence of poly(dA–dT). The origin of the cell extracts is shown above the lanes and the template of the pGT2 series below them. Each lane contains a constant amount of DNA template with 16 μg HeLa extract, 8 μg F9 (ND) extract, 12 μg BJA-B extract and 16 μg MPC11 extract, and a constant amount (2 μg) of synthetic poly[dA–dT]. The positions of complexes A–D are indicated by the arrows to the left of the figure. The broken arrows indicate the position of a 'non-specific' complex found in the F9 cell extracts which migrates close to the complex C (see text). Similar results were reproducibly obtained using independently prepared nuclear extracts from each cell type. **Panel B** shows the results of experiments similar to those displayed in panel A, except that poly[dI–dC] replaced poly[dA–dT] as a non-specific competitor. Nomenclature, as in panel A.

These results confirm that complexes A, B and C are formed by distinct nuclear proteins which apparently differ both by their mol. wt and charge. Furthermore, the protein forming complex C could not be detected in BJA-B and other lymphoid cell extracts tested, even after chromatography and thus appears to be cell-specific. As the formation of complexes B and C, but not A, was prevented using pGT2-1 (Figure 3A), this suggests the possibility that two of these proteins may recognize the same or overlapping sequences. To investigate this possibility, dimethylsulfate (DMS) methylation interference experiments were performed.

The proteins forming complexes A, B and C bind to overlapping sequences

The DMS methylation interference assay (Siebenlist *et al.*, 1980 and Hendrickson and Schleif, 1985) was first used to determine which guanine residues are important for the formation of the three complexes. The guanine (G) residues in the [^{32}P]-5′ end-labelled pGT2-0 template were partially methylated by treatment with DMS. After electrophoresis, the A, B, and C complexes formed using these modified templates were excised from the retardation gel, and the eluted DNA was cleaved at the methylated G residues by the method of Maxam and Gilbert (1980). G residues which are underrepresented in the complexed DNA relative to the unbound and the input DNA (B, F and S, respectively, in Figure 5A and B) are those whose methylation interfered with complex formation.

Methylation of residues G273, G276 and G283 prevented efficient formation of complex A in both HeLa and BJA-B nuclear extracts (Figure 5A, B and data not shown). Complex formation was affected to a lesser extent by methylation at G272 and G275. Formation of complex B was prevented by methylation at G270, G272, G273 and to a lesser extent at G275, whereas

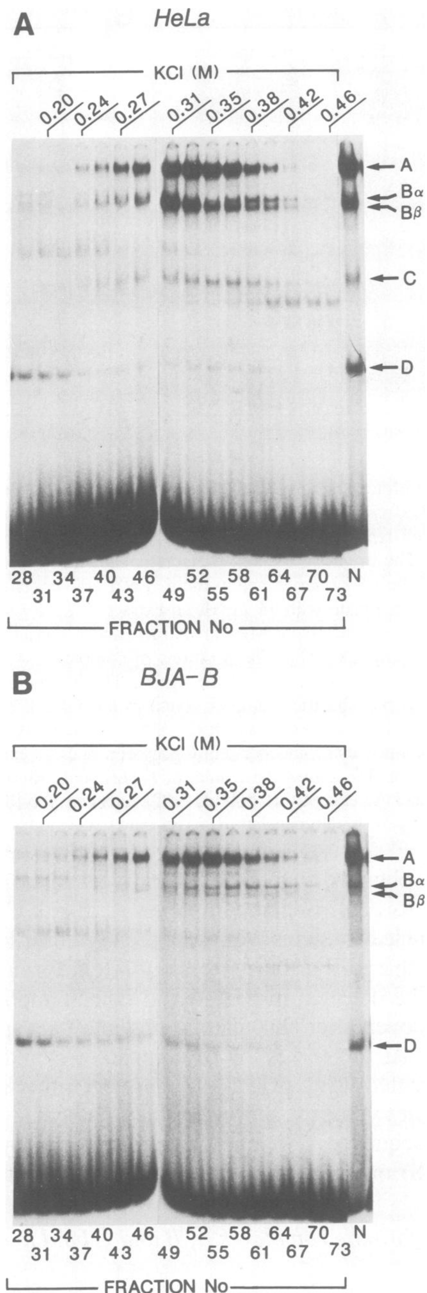

Fig. 4. Chromatographic fractionation of the HeLa and BJA-B cell nuclear extracts. **Panel A** shows the labelled protein−DNA complexes formed using the pGT2-0 template with the HeLa cell proteins eluted from the heparin-agarose column by a linear 0.2−0.6 M KCl gradient. The concentration of KCl (M) in every sixth fraction is indicated above the lanes and the fraction number below. Each lane contains a constant 6 μl of protein fraction and 2 μg poly(dA−dT). Only the relevant fractions from the gradient containing the proteins forming complexes A−C are shown. The positions of complexes A−D in the crude HeLa cell nuclear extract (N) are shown on the right hand side of the figure. Bα and Bβ refer to the two subspecies which constitute complex B (see text). Similar results were obtained using two independent nuclear extract preparations. **Panel B** shows the labelled protein−DNA complexes formed by the BJA-B cell proteins eluted from the heparin-agarose column as described in panel A. Nomenclature, as in panel A. The positions of the A and Bα, Bβ complexes formed using the crude BJA-B cell nuclear extract (N) are indicated on the right hand side.

formation of complex C (which is not formed in lymphoid extracts) was prevented by methylation at G263, G267 and G268, and to a lesser extent at G270 and G272 (Figure 5A and B). Similar results were obtained with complex B formed in BJA-B extracts (data not shown). Note, however, that as the Bα and Bβ complexes migrate too closely to be eluted separately, the results represent the cumulative effect of the two proteins. Thus, we do not know whether they interact with identical sequences. These results which are summarized in Figure 6 suggest that the proteins forming complexes A, B and C interact with three different, but overlapping sequence motifs, GT-IIA, IIB and IIC (271−284, 269−275 and 262−272, respectively, see legend to Figure 6).

To support this hypothesis, a set of complementary oligonucleotides containing the wild-type SV40 sequences from positions 257−275 (OGT2-50) or 271−293 (OGT2-54) were synthesized (see Figure 2). In agreement with the predictions which could be made from the methylation interference results, only complex A was formed using [^{32}P]-5′ end-labelled OGT2-54 as template, while complexes B and C, but not complex A, were formed using OGT2-50 (data not shown). These two sequences were then mutated at positions where, based on the methylation interference experiments, one would expect to selectively prevent the formation of either complex A, B or C. Thus mutation in OGT2-50 at positions 270 and 273 (OGT2-52) should selectively prevent formation of complex B, whereas mutation at positions 263, 267 and 268 (OGT2-51) should prevent only the formation of complex C. Similarly, mutation in OGT2-54 at positions 272, 273, 276 and 283 (OGT2-55) should prevent the formation of complex A. These predictions were fulfilled (data not shown) and, in agreement with the results obtained using the labelled oligonucleotide templates, complexes B and C formed using pGT2-0 were competed by OGT2-50, whereas only formation of complex A was unaffected (Figure 7A). Using OGT2-51 only complexes Bα and Bβ were competed, while OGT2-52 competed complex C only. As expected, OGT2-53, containing a combination of the mutations present in OGT2-51 and OGT2-52, competed neither complexes B nor C. Complex B was efficiently competed only at high competitor concentration suggesting that the corresponding proteins may be abundant in the extract, and Bβ was reproducibly competed more efficiently than Bα. Similarly, OGT2-54, but not OGT2-55, competed complex A (Figure 7B). These results confirm that the A, B and C complexes result from the interaction of three distinct nuclear proteins (possibly four, considering Bα and Bβ) with the three different partially overlapping motifs GT-IIA, IIB and IIC.

Interaction of proteins A, B and C with other enhancer motifs

The proteins forming complexes A, B and C interact specifically with the region (domain B1) of the SV40 enhancer containing the GT-II motif. Using the GT-II motif-containing template pGT2-0, no competition for the formation of complexes A, B, or C was observed with the template OGT1-13 which contains the closely related SV40 GT-I motif (Figures 1 and 2 and Figure 7B; see also Zenke *et al.*, 1986; OGT1-14 is a mutant control for OGT1-13). In addition, no competition was observed with the other SV40 enhancer motifs TC-II, Sph-I and Sph-II, octamer and P, nor with the Sp1 factor-binding motif GC-III from the 21 bp repeat region of the SV40 early promoter (data not shown; the sequences of the corresponding competing oligonucleotides are given in the legend to Figure 7C). Similarly, no competition was observed using the adenovirus type 2 (Ad2) binding sites for nuclear factors I (NF-I) and III (NF-III) (Rosenfeld and Kelly, 1986; Rosenfeld *et al.*, 1987; Pruijn *et al.*, 1986), and the herpes simplex virus thymidine kinase 'CAAT' box (Jones *et al.*, 1985, 1987) (data not shown).

Surprisingly, the B complex was efficiently competed by the wild type, but not the mutated, μE3 motif (Ephrussi *et al.*, 1985;

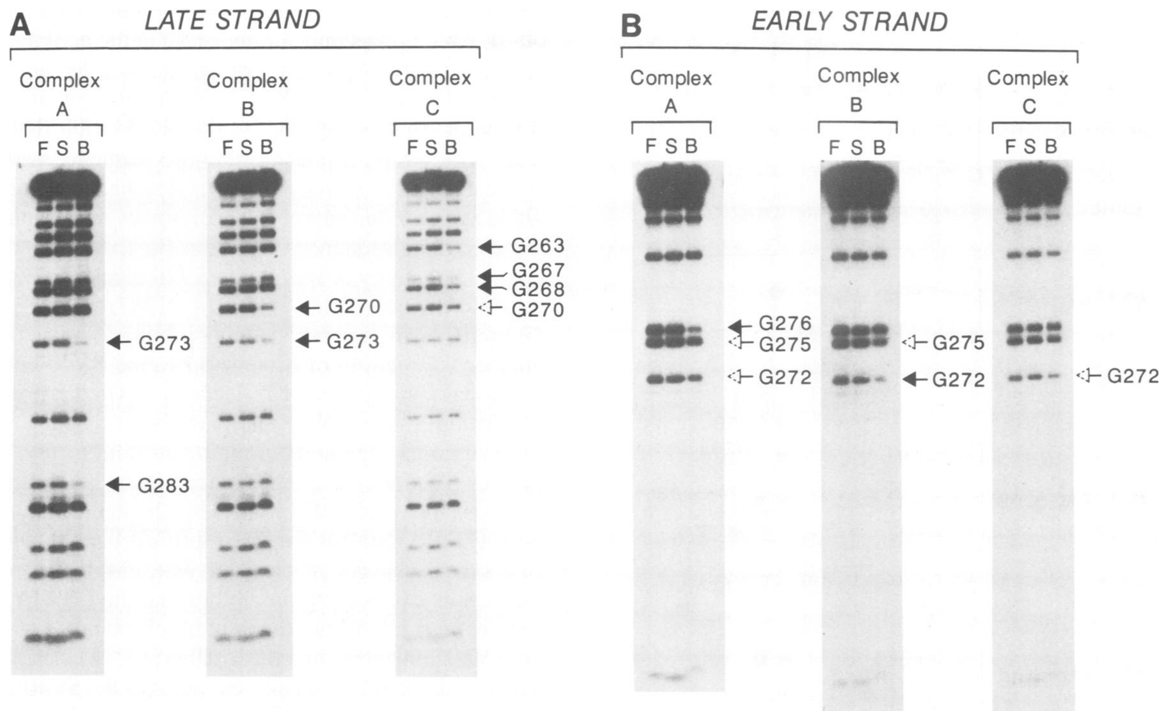

Fig. 5. Methylation interference in complexes A, B and C. **Panel A** shows the pattern of methylation interference from HeLa cell complexes A, B and C on the late coding strand of the pGT2-0 template. F indicates the free (i.e. 'unbound' DNA) and B the bound DNA (i.e. the DNA present in the respective A, B or C complexes). S refers to the DNA which was used as the starting material for complex formation. The location and coordinates of those guanine (G) residues, the methylation of which efficiently or partially prevented complex formation, are indicated by the filled or open triangles, respectively. **Panel B** shows the pattern of methylation interference from HeLa cell complexes A, B and C on the early coding strand of the pGT2-0 template. Nomenclature, as in panel A. The pGT2-0 template was [^{32}P]-5′ end-labelled on the early and late coding strands at the *Bam*HI and Asp718 (an iso-enzyme of *Kpn*I which generates a protruding 5′ end) restriction sites, respectively.

Sen and Baltimore, 1986a) from the IgH enhancer (see OE3-0 and OE3-1 in Figures 2 and 7C). This result indicates that the same protein(s) can bind to the GT-IIB and μE3 motifs, although the μE3 sequence shows no obvious homology to the GT-IIB motif (see Figure 2). However, in contrast to the SV40 GT-II motif the μE3 motif competed for the formation of complex GT-IIBα more efficiently than complex GT-IIBβ (Figure 7C), suggesting that the corresponding proteins do not recognize identical sequences. In agreement with these results, the [^{32}P]-5′ end-labelled μE3 template generated a complex with an electrophoretic mobility similar to that of complex Bα, and the corresponding protein eluted during heparin–agarose chromatography of HeLa cell and lymphoid cell extracts at the same ionic strength as the protein responsible for the formation of the GT-IIBα complex. However, complex Bβ was formed less efficiently than complex Bα using the OE3-0 template, in contrast to the results obtained with the SV40 enhancer motif GT-II, but in agreement with the above competition results (data not shown). A comparison of the kinetics of competition and of the amounts of retarded complexes formed using the two motifs indicates that the μE3 motif present in OE3-0 is in fact a higher affinity binding site than the SV40 GT-IIB motif for protein Bα (Figure 7 and data not shown). It is noteworthy that the μE3 motif template does not compete for the formation of complex C, although its homology is higher for the GT-IIC than for the GT-IIB motif.

The polyoma virus (Py) enhancer contains in its domain B (Herbomel *et al.*, 1984) a sequence related to the GT-IIC motif (see PyB-0 in Figure 2). The enhancer of the polyoma virus host range isolate PyEC9.1 which, in contrast to the wild-type virus, is active in non-differentiated F9 cells (Linney and Donerly, 1983; Herbomel *et al.*, 1984; Melin *et al.*, 1985) contains a mutation at position 5233 which generates the G-C base pair found in the GT-IIC motif (position 268, Figure 2), thereby increasing the homology. A template containing the PyEC9.1 sequence, OPyB-1, but not the corresponding wild-type Py template, OPyB-0, efficiently competed for the formation of complex C (Figure 7C). In agreement with this finding, a retarded complex with similar mobility to that of complex C was found using the OPyB-1, but not the wild-type template OPyB-0, (data not shown). Thus protein C also binds to the PyEC9.1, but not to the wild-type, Py enhancer. Furthermore, these results indicate that the octanucleotide homology 5′-TGGAATGT-3′ found in both the PyEC9.1 and SV40 templates, may constitute the minimal protein binding site. Note that the G residues, the methylation of which strongly interfered with formation of complex C (G267, G268, and G263, see Figure 5A and B), lie within this sequence.

Discussion

The data presented here indicate that the domain B1 of the SV40 enhancer consists of three overlapping sequence motifs GT-IIA, B and C. These three motifs interact with four nuclear proteins hereafter designated proteins GT-IIA, GT-IIBα, GT-IIBβ and GT-IIC. The interactions of each of these proteins with the B1 domain are highly specific as they are selectively prevented by point mutations, and competition with homologous or heterologous (in the cases of proteins GT-IIBα and GT-IIC) enhancer templates. Proteins GT-IIA, GT-IIBα and Bβ are found in extracts from all of the cell lines tested with the notable exception of differentiated F9 cells where GT-IIBα and Bβ were absent.

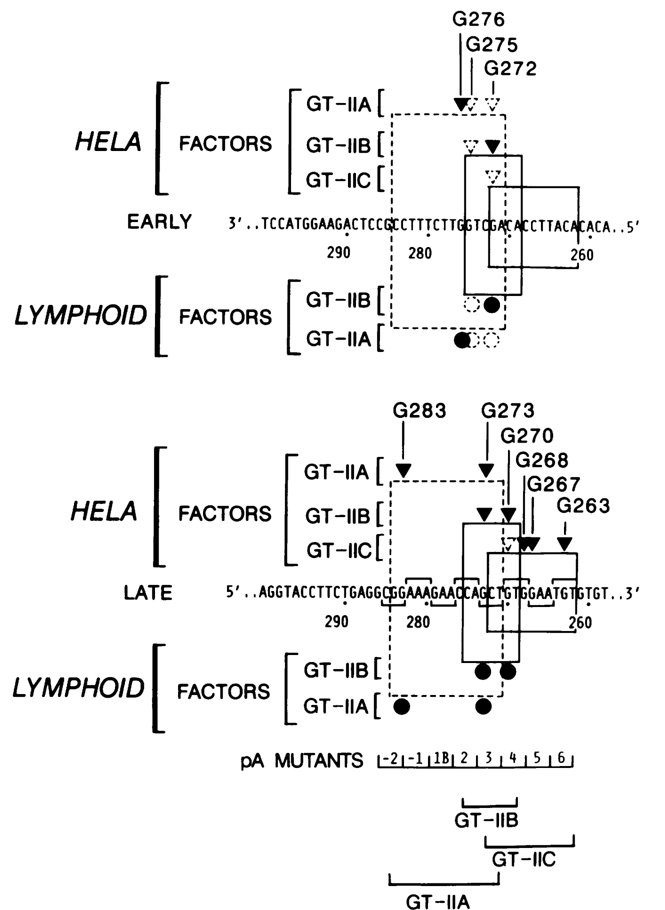

Fig. 6. The SV40 enhancer domain B1 contains three overlapping protein binding sites. The locations of the three sequence motifs GT-IIA, GT-IIB and GT-IIC are shown schematically. The top half shows the sequence of the SV40 early coding strand in the GT-II region. The three sequence motifs GT-IIA, -IIB and -IIC are boxed and indicated at the bottom of the figure. The filled and broken triangles indicate the locations of the G residues whose methylation strongly or partially interferred with the formation of the respective HeLa A, B or C complexes (see Figure 5). The filled and broken circles indicate the positions of the G residues which interfered with the formation of the BJA-B, A and B complexes. The bottom half of the figure shows the sequence of the late coding strand in the GT-II region. The GT-IIA, IIB and IIC motifs are boxed. The filled and broken arrows and circles represent the late strand G residues whose methylation interfered with the formation of the HeLa and BJA-B complexes, respectively. In each case, the borders of the motifs were determined by including the base adjacent to the first non-interfering G residue. The 5′ border of GT-IIA, however, includes G284, as partial protection of this base was observed in DMS methylation protection experiments using BJA-B nuclear extracts (Davidson et al., 1986). The locations of the mutated triplets present in the pA mutant plasmids pA-2 to pA6 are shown by the bars above and below the late strand coding sequences, and indicated in the lower part of the figure.

Protein GT-IIC exhibits a clear cell specificity, being absent from all of the lymphoid cell lines tested.

Possible functions of proteins GT-IIA and B

An indication as to the function of these proteins may be deduced by the comparison of the results obtained here *in vitro* with those obtained in *in vivo* studies. Mutations pA-2, pA-1 and pA1B (see Figures 1 and 6) specific to the binding site for protein GT-IIA cause a small decrease in the SV40 enhancer activity which is more pronounced in MPC11, non-differentiated and differentiated F9 cells than in HeLa cells (Nomiyama et al., 1987). The fact that these mutations have only a minor effect *in vivo* does not rule out the possibility that protein GT-IIA is in fact an enhancer factor, as the P motif (see Figure 1 and Zenke et al., 1986), which plays only a minor role in the activity of the SV40 enhancer *in vivo*, is a major determinant in the activity of other enhancers (Lee et al., 1987). That the same situation may apply to protein Bα is suggested by the observation that this protein binds with a high affinity to the μE3 sequence which has been implicated in the activity of the IgH enhancer (Ephrussi et al., 1985; Sen and Baltimore, 1986b); however, a mutation specific to the cognate sequence of protein Bα (pA3) has only little effect on the activity *in vivo* of the SV40 enhancer in the cell-types tested up to now (Nomiyama et al., 1987). As the μE3 and GT-IIB motifs share no obvious sequence homologies, it is not clear whether the binding of protein Bα to the SV40 enhancer has a functional significance or reflects an *in vitro* artefact of the gel retardation system. We note, however, that the yeast Hap 1 transcriptional regulatory protein can bind to two different sequences (Pfeifer et al., 1987), each of which is functionally relevant. Whether protein Bβ which has a higher affinity for the SV40 motif GT-IIB than for the IgH motif μE3 could also be a factor for another enhancer is unknown. Note also that none of our present results exclude the possibility that the GT-IIA, Bα and Bβ proteins may contribute much more efficiently to the activity of the SV40 enhancer in other cell-types. Alternatively, as the enhancer domain B1 contains elements of the SV40 late promoter (Vigneron et al., 1984; Ernoult-Lange et al., 1987), the GT-IIA, Bα and Bβ proteins may participate in the activity of this promoter. An ~3-fold decrease in late transcription has been observed when the 270−300 region, which contains the binding sites for proteins GT-IIA and GT-IIBα and Bβ, was deleted (see pSV4BK3 and pSV2Δ72 in Ernoult-Lange et al., 1987).

The GT-IIC protein may be a trans-acting factor for both the SV40 and the PyEC9.1 polyoma virus enhancers

Our present *in vitro* results show that the GT-IIC motif is recognized in a cell-specific manner by the protein GT-IIC. This motif (5′-CTGTGGAATGT-3′) is essentially the motif defined by mutations pA4, pA5 and pA6 (5′-GTGGAATGT-3′) which are detrimental to enhancer activity in HeLa cells *in vivo* (Zenke et al., 1986; see also Figures 1 and 6). Herr and Clarke (1986) have also found that a mutation located in the GT-IIC motif (dpm1, T-A base pair at position 262) is deleterious for the SV40 enhancer activity in HeLa cells. The competition experiments with the PyEC9.1 enhancer motif indicate also that the GT-IIC octanucleotide sequence, 5′-TGGAATGT-3′ which coincides with the genetically defined sequence motif, may correspond to the minimal protein binding site. The same pA4, pA5 and pA6 mutations are also highly detrimental to SV40 enhancer activity in non-differentiated and differentiated F9 cells, whereas they have no significant effect in MPC11 B cells (Nomiyama et al., 1987). Moreover, the multimerized GT-IIC motif exhibits a similar pattern of cell-specific activity *in vivo* (Schirm et al., 1987; Ondek et al., 1987; Fromental et al., 1987). That protein GT-IIC is present in the cell-types where the motif is active (HeLa and F9), but absent from the lymphoid B cells where the motif is inactive (MPC11), suggests very strongly that this protein is the positive *trans*-acting factor responsible for the activity of the B1 enhancer domain. Our results, however, do not exclude the possibility that protein GT-IIC is present in an inactive form in MPC11 cells, as has been suggested to be the case for the enhancer binding NFκB protein in other cell-types (Sen and Baltimore, 1986b).

The activity of the polyoma virus enhancer which is low in non-differentiated F9 EC cells (Linney and Donerly, 1983; Herbomel et al., 1984; Melin et al., 1985) is increased by the

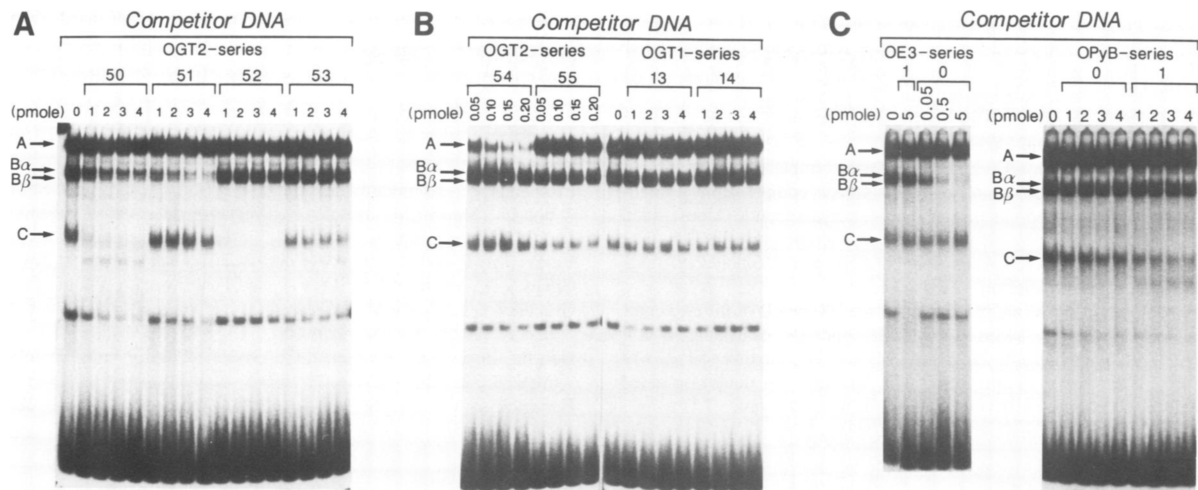

Fig. 7. Competition for the formation of the HeLa A, B and C complexes with the pGT2-0 template by homologous and heterologous enhancer motifs. **Panel A** shows the results of the competition experiments using increasing concentrations of OGT2-50 to OGT2-53 templates as competitors (see Figure 2). The quantity (in picomol) and origin of the competitor templates are shown above the lane. Each lane contains a constant amount of HeLa nuclear extract, [^{32}P]-5′ end-labelled pGT2-0 DNA template and poly[dA−dT]. The positions of the A, Bα, Bβ and C complexes in the absence of specific competitor DNA are shown on the left hand side. **Panel B** shows the results of competition experiments using templates OGT2-54 and 55, and OGT1-13 and 14 (see Figure 2) as competitors. The origin and quantity of competitor templates used are indicated above each lane. **Panel C** shows the competition of complexes Bα, Bβ and C by heterologous enhancer motifs. Competitors OE3-0 and OE3-1 contain the wild-type and mutated sequence of the μE3 motif from the murine IgH enhancer (see Figure 2 and Ephrussi *et al.*, 1985; Sen and Baltimore, 1986a). OPyB-0 and OPyB-1 contain the sequence, from positions 5223−5246, of the wild-type polyomavirus (Py) or the mutant isolate PyEC9.1 enhancer B domains, respectively. The positions of the HeLa A, Bα, Bβ and C complexes formed using the pGT2-0 template are indicated to the left of the figures. The quantity of the competitor DNA used is indicated above each lane. No competition was observed with oligonucleotides containing: the other SV40 enhancer motifs, TC-II [5′-(245)GGAAAGTCCCCAGG(232)-3′], P[5′-(201)GCATCTCAATTAGTCA-GCAACCA(179)-3′] and Sph [5′-(222)AGGCAGAAGTATGCAAAGCATGCATCT(196)-3′]; the Sp1 factor binding site GC-III of the SV40 21 bp repeats [5′-(78)GCCCCTAACTCCGCCCAGTTC(58)-3′]; the murine IgH enhancer motif μE1 [5′-(343)TTGAGTCAAGATGGCCGATCAG(364)-3′]; the adenovirus-2 (Ad2) binding site for the nuclear factor I (NF-I) [5′- (15)TACCTTATTTTGGATTGAAGCCAATATGATAA(46)-3′]; the Ad2 binding site for the nuclear factor III (NF-III) [5′-(29)TTGAAGCCAATATGATAATGAGGGGGTGGAG(59)-3′]; or the HSV thymidine kinae (TK) promoter 'CAAT' box [5′-(−95)CGTCTTGTCATTGGCGAATTCGAACACG(−68)-3′]. In each case only the motif (underlined) and the immediately flanking sequence are shown. Each template was cloned in the modified pUC18 vector (Materials and methods) and the *Hin*dIII−*Eco*RI fragment containing that motif was used in the competition experiments.

mutation present in the PyEC9.1 isolate. It is therefore very interesting that the protein GT-IIC binds to the mutated motif present in the domain B of the PyEC9.1 enhancer, but not to the corresponding wild-type motif. As non-differentiated F9 cells contain the protein GT-IIC, this differential binding may account for the activity of the mutant PyEC9.1 enhancer in these cells. The binding of the positively acting factor GT-IIC may also in part be responsible for the inability of the adenovirus-2 E1A gene products to efficiently repress this mutant Py enhancer (Borrelli *et al.*, 1984; Hen *et al.*, 1986).

The SV40 enhancer GT-I and GT-II motifs and the enhancer 'core' homology

The results obtained using point mutations within the GT-IIA, B, and C motifs suggest that the conservation of some bases is critical for these motifs to function as efficient protein binding sites. This conclusion is strongly supported by the observation that the SV40 GT-I motif and the wild-type Py sequence, each of which differs from the minimal protein binding region of the GT-IIC motif at only one position (TA to AT at position 264 in comparison with motif GT-I, and GC to AT at position 268 in comparison with Py, see Figures 1 and 2), do not compete for the formation of complex C. In fact, motif GT-I binds different nuclear proteins (Xiao *et al.*, 1987). Thus, in the case of SV40 enhancer, the two so-called 'core' sequences (5′-GTGG$^{AAA}_{TTT}$G-3′, Laimins *et al.*, 1982 and Weiher *et al.*, 1983) bind distinct nuclear proteins. That the only difference between motif GT-I and motif GT-IIC is the change of an A−T base pair to a T−A base pair indicates that the presence of a 'core' homology has little predictive value concerning the possible binding of a given factor.

The finding that both the A (Davidson *et al.*, 1986) and the B1 enhancer domains consists of cell-specifically recognized overlapping sequence motifs, shows that the SV40 enhancer has evolved to contain a large number of combinatorial regulatory possibilities within a short DNA segment. The purification of the proteins which bind to these motifs will help to elucidate the mechanism of enhancer activity.

Materials and methods

Preparation of nuclear extracts and chromatography

Nuclear extracts were prepared from HeLa, BJA-B and Molt-4 cells according to the method of Dignam *et al.* (1983) with the modifications of Wildeman *et al.* (1984) and dialysed at 4°C against buffer A [20 mM Hepes pH 7.9, 1 mM DTT, 1 mM $MgCl_2$, 20 mM KCl, 25% glycerol (w/v), and 0.5 mM phenylmethylsulfonyl-fluoride (PMSF)]. F9(ND), CV1, 70Z/3 (NI) and MPC11 cell nuclear extracts were prepared in the same way, except that all solutions contained the protease inhibitors, aprotinin (5 μg/ml), leupeptin (30 μg/ml), pepstatin (5 μg/ml) and 0.2 mM PMSF. Protein concentration was measured by the method of Bradford (1976).

Aliquots of 5 ml of nuclear extract (4 mg/ml) in buffer A were applied to a 20 ml heparin-agarose column, and the column was washed with three volumes of buffer A containing 0.1 M KCl. The retained proteins were eluted with a 60 ml linear 0.2−0.6 M KCl gradient. Fractions of 1 ml were collected and the KCl concentration determined by conductivity. The fractions were dialysed against buffer A, frozen in liquid nitrogen and stored at −70°C. The proteins were purified 3- to 5-fold by this chromatography step.

Construction of recombinant plasmids and preparation of DNA templates

The pGT2-series templates were cloned between the *Kpn*I and *Bam*HI sites of a modified pUC18 vector (Rosales *et al.*, 1987) by 'shotgun' ligation of synthetic complementary oligonucleotides (Grundström *et al.*, 1985), the sequences of which are described in Figure 2. The nucleotide sequence of the inserts was verified by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977).

All plasmids were prepared by standard methods (Zenke *et al.*, 1986) and purified over two CsCl gradients. The pGT2-series plasmids were cleaved at the *Bam*HI site, dephosphorylated with calf alkaline phosphatase, and labelled with [γ-^{32}P]-ATP (5000 ci/mmol, Amersham) using T4 polynucleotide kinase. The end-labelled templates were excised with *Kpn*I, purified on an 8% polyacrylamide gel and recovered by electroelution.

[^{32}P]-5′ end-labelled OGT2, OE3 and OPyB series templates were prepared as follows. An equal quantity (5 pmol) of two synthetic complementary oligonucleotides were [^{32}P]-5′ end-labelled with T4 polynucleotide kinase, and hybridized in the same buffer by incubation at room temperature for 2 h. The resulting double-stranded DNA templates were purified on an 8% polyacrylamide gel, and recovered by electroelution.

To prepare competitor DNA, an equal quantity ($\geq$500 pmol) of the two complementary oligonucleotides was added to a hybridization buffer containing 50 mM Tris−HCl (pH 7.9), 5 mM $MgCl_2$ and 1 mM EDTA. The mixture (50 μl) was incubated for 2 h at room temperature. The quantity of double-stranded oligonucleotide was determined by OD measurement at 260 nm.

Gel retardation and DMS methylation interference assays

Gel retardation assays were performed essentially as described by Strauss and Varshavsky (1984) with the modifications of Singh *et al.* (1986). Aliquots of 4−16 μg of protein were mixed with 2 μg of either poly[dA−dT] or poly[dI−dC] (Pharmacia) in 10 μl of 10 mM Hepes pH 7.9, 1 mM DTT, 1 mM $MgCl_2$, 30 mM KCl and 12% glycerol. Pre-incubation was carried out in ice for 15 min, before adding 10−20 fmol (4 $\times$ 10^4 c.p.m.) of the [^{32}P]-5′ end-labelled DNA template to the above mixture. Incubation was continued at 20°C for 15 min and the DNA-protein complexes were loaded onto a low ionic strength 7% polyacrylamide gel (30:1 cross-linking ratio) containing 6.7 mM Tris−HCl pH 7.5, 3.3 mM sodium acetate and 1 mM EDTA. Electrophoresis was performed at 25 mA until a suitable separation had been achieved. The gel was then dried and subjected to autoradiography.

The DMS methylation interference assays were carried out essentially as previously described (Hendrickson and Schleif, 1985). 2 $\times$ 10^6 c.p.m. of [^{32}P]-5′ end-labelled DNA fragments were partially methylated at the guanine (G) residues by treatment with DMS (Rosales *et al.*, 1987). The methylated DNA was ethanol-precipitated in the presence of 10 μg tRNA washed with 70% ethanol, and 2 $\times$ 10^5 c.p.m. used in the gel retardation assay as described above. After electrophoresis, the wet gel was autoradiographed at 4°C overnight. The complexed and 'unbound' DNA were excised from the gel and electroeluted. After two precipitations with ethanol, the DNA fragments were cleaved at the methylated G residues as described by Maxam and Gilbert (1980), and electrophoresed on a 15% polyacrylamide−7.5 M urea gel. The gel was then fixed, dried and autoradiographed.

Gel retardation competition assays

Binding reactions and gel electrophoresis were carried out as described above, except that the cold competitor DNA was added together with the labelled DNA following the pre-incubation period.

Acknowledgements

We would like to thank A.Staub for his help in the preparations of the oligonucleotides. We also thank C.Werlé and B.Boulay for the illustrations, the secretarial staff for typing the manuscript and M.Gilbert for technical assistance with cell culture. ID, MM, RR and J-HX were supported by fellowships from the Royal Society of Great Britain, the American Cancer Society (PF 2640), the Université Louis Pasteur, and the French Government. This work was supported by grants from the CNRS (ATP 6984), the INSERM, the Fondation pour la Recherche Médicale, the Ministère de la Recherche et de la Technologie (84V0803) and the Association pour le Développement de la Recherche sur le Cancer.

References

Atchison,M.L. and Perry,R.B. (1987) *Cell*, **48**, 121−128.

Augereau,P. and Chambon,P. (1986) *EMBO J.*, **5**, 1791−1797.

Banerji,J., Rusconi,S. and Schaffner,W. (1981) *Cell*, **27**, 299−308.

Banerji,J., Olson,L. and Schaffner,W. (1983) *Cell*, **33**, 729−740.

Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248−254.

Borrelli,E., Hen,R. and Chambon,P. (1984) *Nature*, **312**, 608−612.

Chambon,P., Dierich,A., Gaub,M.P., Jakowlev,S., Jongstra,J., Krust,A., LePennec,J.P., Oudet,P. and Reudelhuber,T. (1984) In Greep,R.O. (ed.), *Recent Progress in Hormone Research*. The Proceedings of the Laurentian Hormone Conference, vol. 40, Academic Press, NY, pp. 1−42.

Davidson,I., Fromental,C., Augereau,P., Wildeman,A., Zenke,M. and Chambon,P. (1986) *Nature*, **323**, 544−548.

Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475−1489.

Dougherty,J.P., Augereau,P. and Chambon,P. (1986) *Mol. Cell. Biol.*, **6**, 4117−4121.

Ephrussi,A., Church,G.M., Tonegawa,S. and Gilbert,W. (1985) *Science*, **227**, 134−138.

Ernoult-Lange,M., Omilli,F., O'Reilly,D.R. and May,E. (1987) *J. Virol.*, **61**, 167−176.

Fried,M. and Crothers,D.M. (1981) *Nucleic Acids Res.*, **9**, 6505−6525.

Fromental,C., Nomiyama,H., Kano,M., Bornert,J.M., Davidson,I., Ferrandon,D., Macchi,M., Rosales,R., Vigneron,M., Xiao,J.H. and Chambon,P. (1987) in preparation.

Garner,M.M. and Revzin,A. (1981) *Nucleic Acids Res.*, **9**, 3047−3060.

Gillies,S.D., Morrison,S.L., Oi,V.T. and Tonegawa,S. (1983) *Cell*, **33**, 717−728.

Grundström,T., Zenke,M., Wintzerith,M., Matthes,H.W.D., Staub,A. and Chambon,P. (1985) *Nucleic Acids Res.*, **13**, 3305−3316.

Hen,R., Borrelli,E., Fromental,C., Sassone-Corsi,P. and Chambon,P. (1986) *Nature*, **321**, 249−251.

Hendrickson,W. and Schleif,R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3129−3133.

Herbomel,P., Bourachot,B. and Yaniv,M. (1984) *Cell*, **39**, 653−662.

Herr,W. and Clarke,J. (1986) *Cell*, **45**, 461−470.

Jones,K.A., Yamamoto,K.R. and Tjian,R. (1985) *Cell*, **42**, 559−572.

Jones,K.A., Kadonaga,J.T., Rosenfeld,P.J., Kelly,T.J. and Tjian,R. (1987) *Cell*, **48**, 79−89.

Klein,G., Giovanella,B., Westman,A., Stehlin,J.S. and Munford,D. (1975) *Intervirology*, **5**, 319−334.

Laimins,L.A., Khoury,G., Gorman,C., Howard,B. and Gruss,P. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6453−6457.

Lee,W., Haslinger,A., Karin,M. and Tjian,R. (1987) *Nature*, **325**, 368−372.

Linney,E. and Donerly,S. (1983) *Cell*, **35**, 693−699.

Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499−525.

Melin,F., Pinon,H., Kress,C. and Blangy,D. (1985) *J. Virol.*, **53**, 862−866.

Mercola,M., Goverman,J., Mirell,C. and Calame,K. (1985) *Science*, **227**, 266−270.

Minowada,J., Ohnuma,T. and Moore,G.E. (1972) *J. Natl. Cancer Inst.*, **49**, 891−895.

Moreau,P., Hen,R., Wasylyk,B., Everett,R., Gaub,M.P. and Chambon,P. (1981) *Nucleic Acids Res.*, **9**, 6047−6068.

Mosthaf,L., Pawlita,M. and Gruss,P. (1985) *Nature*, **315**, 597−600.

Nomiyama,H., Fromental,C. and Chambon,P. (1987) *Proc. Natl. Acad. Sci. USA*, in press.

Ondek,B., Shepard,A. and Herr,W. (1987) *EMBO J.*, **6**, 1017−1025.

Paige,C.J., Kincade,P.W. and Ralph,P. (1978) *J. Immunol.*, **121**, 641−647.

Pfeifer,K., Prezant,T. and Guarente,L. (1987) *Cell*, **49**, 19−27.

Picard,D. and Schaffner,W. (1984) *Nature*, **307**, 80−82.

Picard,D. (1985) In *Oxford Surveys on Eukaryotic Genes*. Vol. 2, Oxford University Press, pp. 24−48.

Pruijn,G.J.M., van Driel,W. and van der Vliet,P.C. (1986) *Nature*, **322**, 656−659.

Queen,C. and Baltimore,D. (1983) *Cell*, **33**, 741−748.

Rosales,R., Vigneron,M., Macchi,M., Davidson,I., Xiao,J.H. and Chambon,P. (1987) *EMBO J.*, in press.

Rosenfeld,P.J. and Kelly,T.J. (1986) *J. Biol. Chem.*, **261**, 1398−1408.

Rosenfeld,P.J., O'Neill,E.A., Wides,R.J. and Kelly,T.J. (1987) *Mol. Cell. Biol.*, **7**, 875−886.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.

Sassone-Corsi,P., Dougherty,J.P., Wasylyk,B. and Chambon,P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 308−312.

Sassone-Corsi,P., Wildeman,A. and Chambon,P. (1985) *Nature*, **313**, 458−463.

Sassone-Corsi,P. and Borrelli,E. (1986) *Trends Genet.*, **2**, 215−219.

Schirm,S., Jiricny,J. and Schaffner,W. (1987) *Genes and Development*, **1**, 65−74.

Schöler,H.R. and Gruss,P. (1984) *Cell*, **36**, 403−411.

Schöler,H.R. and Gruss,P. (1985) *EMBO J.*, **4**, 3005−3013.

Schöler,H., Haslinger,A., Heguy,A., Holtgreve,H. and Karin,M. (1986) *Science*, **232**, 76−80.

Sen,R. and Baltimore,D. (1986a) *Cell*, **46**, 705−716.

Sen,R. and Baltimore,D. (1986b) *Cell*, **47**, 921−928.

Serfling,E., Jasin,M. and Schaffner,W. (1985) *Trends Genet.*, **1**, 224−230.

Sergeant,A., Bohmann,D., Zentgraf,H., Weiher,H. and Keller,W. (1984) *J. Mol. Biol.*, **180**, 577−600.

Siebenlist,U., Simpson,R.B. and Gilbert,W. (1980) *Cell*, **20**, 269−281.

Singh,H., Sen,R., Baltimore,D. and Sharp,P.A. (1986) *Nature*, **319**, 154−158.

Strauss,F. and Varshavsky,A. (1984) *Cell*, **37**, 889−901.

Tooze,J. (1982) *Molecular Biology of Tumor Viruses, 2nd Edition, Part 2, Revised. DNA Tumor Viruses.* Cold Spring Harbor Laboratory Press, New York.

Vigneron,M., Barrera-Saldana,H.A., Baty,D., Everett,R.E. and Chambon,P.

(1984) *EMBO J.*, **3**, 2373–2382.
Walker,M.D., Edlund,T., Boulet,A.M. and Rutter,W.J. (1983) *Nature*, **306**, 557–561.
Wasylyk,B. (1986) In Reznikoff,W. and Gold,L. (eds), *Maximizing Gene Expression*. Boston: Butterworths, pp. 79–99.
Wasylyk,C., Imler,J.L., Perez-Mutul,J. and Wasylyk,B. (1987) *Cell*, **48**, 525–534.
Weiher,H., König,M. and Gruss,P. (1983) *Science*, **219**, 626–631.
Wildeman,A., Sassone-Corsi,P., Grundström,T., Zenke,M. and Chambon,P. (1984) *EMBO J.*, **3**, 3129–3133.
Wildeman,A., Zenke,M., Schatz,C., Wintzerith,M., Grundström,T., Matthes,H., Takahashi,K. and Chambon,P. (1986) *Mol. Cell. Biol.*, **6**, 2098–2105.
Xiao,J.H., Davidson,I., Macchi,M., Rosales,R., Vigneron,M., Staub,A. and Chambon,P. (1987) *Genes and Devel.*, in press.
Yaniv,M. (1984) *Biol. Cell.*, **50**, 203–216.
Zenke,M., Grundström,T., Matthes,H., Wintzerith,M., Schatz,C., Wildeman,A. and Chambon,P. (1986) *EMBO J.*, **5**, 387–397.

Received on May 20, 1987