1172–1178 Nucleic Acids Research, 1994, Vol. 22, No. 7

Characterization of the DNA-binding site repertoire for the Epstein – Barr virus transcription factor R

Henri Gruffat+ and Alain Sergeant*
U 412 INSERM, Unité de Virologie Humaine de L'Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

Received January 4, 1994; Revised and Accepted February 23, 1994

ABSTRACT

The Epstein – Barr virus gene BRLF1 encodes the transcription factor R, which is a sequence-specific DNA-binding protein important for the switch from latency to a productive cycle. We have defined a repertoire of specific R-binding sites using a GST-R fusion protein and a pool of 23 bp random DNA sequences. The R-bound sequences were selected by several rounds of Electrophoretic Mobility Shift Assay (EMSA) and amplification by PCR. Among the 45 sites selected, some positions in the sequences were highly conserved, i.e., 5′-G^{T}GCC N_7 GTGGTG-3′. The guanine methylation assay revealed that R simultaneously contacts guanines in the two conserved cores, defining the consensus binding site 5′-GNCC N_9 GGNG-3′, and 30 sites among the 45 selected have this sequence. This last result also suggests that R binds two adjacent major grooves of the DNA. As shown by EMSA assay, R binds to all the sites tested with a comparable affinity, and they all mediate R-induced transcriptional activation in a transient expression assay.

INTRODUCTION

In vitro infection of B lymphocytes by Epstein – Barr virus (EBV) is generally followed by deregulation of cell proliferation (immortalization) and persistence of the virus in a latent state (see reference 1 and the references therein). The switch from latency to the lytic cycle is thought to be mainly due to the expression of two EBV transcription factors EB1 (otherwise known as Z, Zta or ZEBRA), and R (Rta). Indeed, in latently infected B cells, the genes coding for EB1 and R are down-regulated, and transient expression of EB1 or R can induce transcription of the EBV genes required for progression to the lytic cycle (2–6). Accordingly, there are EB1 (ZREs) and R (RREs) responsive elements in many EBV early promoters (7–15). However, only EB1 is able to induce the lytic cycle (6), because only transactivates ORIlyt, the origin of replication active during the lytic cycle (16–18).

R is a protein of 605 aminoacids that can be functionally divided into a DNA-binding/dimerisation region localized in the 280 N-terminal aminoacids and an activation region located in the C-terminal part of the protein (19, 20). R probably activates transcription by interacting directly with specific DNA sequences called RREs. Indeed, *in vitro* R binds directly to RREs as a homodimer but such homodimers can also form and remain stable in the absence of a DNA-binding site (19). The activation domain of R is composed of an acidic region whose activity is potentialized in a cell-specific manner, by an adjacent region rich in prolines (21). *In vitro*, the acidic domain of R interacts directly with both the TATA box factor TBP, and TFIIB (21), suggesting that R activates transcription by recruiting both TFIID and TFIIB to the promoter.

We, and others, have characterized four binding sites for R in three EBV early promoters. Two, RRE-DR1 and RRE-DR2, were identified in the DR/DL promoter (11), one in the BMLF1 promoter (RRE-M) (14) and one in the BMHRF1 promoter (RRE-Mr) (15) (Figure 2A). As shown by guanine methylation interferences, R simultaneously contacts two core sequences separated by a random sequence, each binding-site covering 17 to 18 bp, (Figure 2A) (11, 14, 15). Comparison of these sites shows that the 5′-core sequences are relatively homologous, namely 5′-GTCC-3′. Although the 3′-core sequences are more degenerate it is possible to deduce the consensus sequence $G_ATGGT_{AC}G_C$ for the different sites (Figure 2A). In order to further characterize the R-binding sites, we have used the strategy of Cyclic Amplification and Selection of Targets (CASTing also called SSAB, Sequential Selection and Amplification of Binding sites) (For a review see reference 22 and the references therein). We have selected 45 R-binding sites, and the majority of them have the sequence 5′-G$^{G}/_{T}$CC $(N)_7$ GTGGTG-3′. A second CASTing done with a degenerated oligonucleotide containing the 5′-core sequence GTCCCT demonstrated that nearly all of the selected R-binding sites had the sequence 5′-GTCCCT $(N)_5$ GTGGTG-3′, confirming the high conservation of the 3′-core sequence. Guanine methylation interference on some of the selected sites confirmed that R simultaneously contacts two core sequences in which two pairs of guanines separated by 9 bp and

*To whom correspondence should be addressed

+Present address: Institut für Klinische Molekularbiologie und Tumorgenetik, GSF-Forschungszentrum für Umwelt und Gesundheit, GmbH Marchioninistrasse 25, D-81377 Munich, Germany

located on the opposite DNA strands, are crucial for R-binding. This last result also suggests that R binds two adjacent major grooves of the DNA. As shown by electrophoretic mobility shift assay, R binds to all of the sites tested with a comparable affinity, and they all mediate R-induced transcriptional activation in a transient expression assay.

MATERIAL AND METHODS

Production of R and mutant proteins *in vitro*

The R-expression plasmids pSPTR and pSPTRt356 have been previously described (19). They were used to produced the R protein (pSPTR) or the DNA-binding domain of R (pSPTRt356), in an *in vitro* transcription/translation system.

The R protein was also produced in bacteria as a glutathione-S-transferase (GST) fusion protein, by fusing an R cDNA (Aminoacids 1 to 605) to the glutathione-S-transferase gene in the expression vector pGEX-2T (Pharmacia). The GST-R fusion protein was expressed in *E.coli* and purified by affinity-chromatography on glutathione-agarose beads as described previously (21).

DNA fragments and oligonucleotides

All DNA oligonucleotides were synthesized with an Applied Biosystems Synthesizer (380A). The structure of the degenerated oligonucleotide D_{23} is as follows: 5'-AGACGGATCCAG $(N)_{23}$ CCGAATTCGATA-3'. The structure of the degenerated oligonucleotide D_A is as follows: 5'-AGACGGATCCAG $(N)_4$ GTCCCT $(N)_{15}$ CCGAATTCCTAC-3'. The sequences of the 20 base oligonucleotides used as primers for PCR amplification are as follows: primer A: 5'-TGTCTCGAGTAGGAATTC-GG-3'; primer B: 5'-TAGCCTCGAGACGGATCCAG-3'. The *Eco*RI and *Bam*HI recognition sites, which allowed cloning of individual fragments, are underlined. Double-stranded templates were generated by annealing D_{23} or D_A to a ten fold molar excess of primer A, synthesizing the complementary DNA strand with the Klenow fragment of *E.coli* DNA polymerase and purifying the template on a 10% polyacrylamide gel. The templates were 5'-end-labeled by kinase reactions. The RRE-DR1 and RRE-DR1* oligonucleotides have been described elsewhere (11).

DNA binding analysis

Electrophoretic mobility shift assays were performed by incubating 5×10^4 cpm of 5'-end-labeled double-stranded D_{23} or D_A templates with 100 ng of purified GST or GST-R fusion protein, or with 2µl of *in vitro* translated R for 30 min at room temperature in 10 mM HEPES (pH 7.9), 100 mM KCl, 0.5 mM $MgCl_2$, 1 mM EDTA, 0.5 mM DTT, 8% glycerol and 1µg of poly(dI-dC) in a final volume of 20 µl. For the selection rounds, the GST-R fusion protein was used instead of *in vitro* translated R. After incubation, the mixture was loaded onto a 4.5% polyacrylamide gel (29 to 1 crosslinked), 0.2×TBE, and run at room temperature at 10 V/cm for 3 hr. The R-DNA complexes were visualized by autoradiography.

PCR amplification

To isolate the DNA from the R-DNA complexes, the dried gel was cut together with the 3MM (Whatman) paper and incubated at 37°C for 3 hours in 0.5 ml of 0.5 M NH_4Ac, 10 mM $MgCl_2$, 1 mM EDTA and 0.1% SDS. After addition of 5 µg of glycogen, the eluate was extracted with phenol−chloroform and then precipated with ethanol. One fifth of the recovered DNA was amplified by 20 cycles of PCR in a 100 µl reaction with primers A and B. PCR was performed under standard conditions. The PCR amplification products were purified on a 10% polyacrylamide gel and electroeluted. The products were either 5'-end labeled and used for another round of selection, or digested by *Eco*RI and *Bam*HI for cloning of individual fragments.

Sequencing of PCR products

PCR-amplified DNA fragments were cloned in the *Eco*RI and *Bam*HI sites of the pBluescript II cloning vector (pBSK, Stratagene) and individual clones sequenced.

Methylation interference

The 89 bp *Xho*I−*Xba*I fragments from the pBSK-D_{23} clones were 5'-^{32}P-labeled at the *Xho*I or *Xba*I site. Labeled DNA was purified by gel electrophoresis and 2×10^5 cpm was methylated using 1 µl of DMS during 3 mn at 18°C. The methylated probe was then incubated with 2µl of Rt356 protein produced in an *in vitro* transcription/translation system. The binding conditions were as described above. After electrophoresis, the gel was exposed to film overnight to reveal the R-DNA complexes. The retarded DNA probe (B) and the non-retarded DNA probe (F) were electroeluted and incubated in 100 µl of 1M piperidine for 30 mn at 90°C to cleave the DNA at the methylated guanosines. An equal amount of the radioactive B and F probes was analysed on 8% polyacrylamide sequencing gels and visualized by autoradiography.

Transfections

Selected R-binding sequences were cloned in the *Eco*RI and *Hin*dIII restriction sites of vector pG1 (23), carrying the rabbit β-globin gene. The plasmids used for transfection were prepared by the alkaline lysis method and purified through two CsCl gradients. The DNAs were in the same topological state as assayed by agarose gel electrophoresis. HeLa cells were grown in DMEM (Gibco) supplemented with 10% (v/v) foetal calf serum and were seeded at 5×10^5 cells per 100 mm Petri dish 8 h prior to transfection. Transfections were performed by the calcium precipitate method. Cells were mixed with the appropriate DNAs, and typically 15 µg of DNA was used which included: 0.1 µg of either R-expressing vector or Rt515-expressing vector (19), 5 µg of plasmid carrying the reporter gene, 1µg of plasmid pSV2β carrying the SV40 early gene as an internal control for transfection and pUC19 up to the 15 µg total.

RNA extraction and S1 nucleas mapping

The transfected cells were collected and lysed in LB buffer (10mM Tris pH 7.8; 10mM NaCl; 2mM $MgCl_2$; 5mM DTT) complemented with 0.5% NP40. Nuclei were pelleted and RNA phenol extracted from the cytoplasmic fraction. 20 µg of total cytoplasmic RNA was hybridized overnight at 30°C in 50% formamide, 0.4M NaCl, 0.04M Pipes pH 6.5 to a 5'-^{32}P-labeled synthetic single-stranded DNA probe as described previously (23). The hybrids were digested 2 hours at 25°C with 10 U of S1 nuclease per 10 µg of RNA. The S1-protected DNA was analysed on 8% (w/v) polyacrylamide sequencing gels. Quantification was made by cutting the specific S1-protected bands out of the gel and counting the radioactivity. The results were corrected as follows: (i) according to the efficiency of transfection as evaluated by counting the radioactivity present

in S1-protected DNA corresponding to specific SV40 early RNA expressed from plasmid pSV2β cotransfected as an internal control, and (ii) according to the activities of the different constructions in the presence of Rt515, an R mutant with the activation domain deleted.

RESULTS

Selection of specific R-DNA complexes

In order to select a pool of specific R-DNA complexes, we used the CASTing method (Cyclic Amplification and Selection of Targets) (22). A 47 bp oligonucleotide (D_{23}) containing a random set of 23 bp flanked by PCR priming sequences containing restriction sites, was used as a starting template for the first selection round. The free template was separated from the R-DNA complexes by electrophoretic mobility shift assay (EMSA). The DNA−protein complexes were eluted from the dried gel, amplified by PCR using the A and B PCR primers and then used in further selection rounds. Four selection rounds were sufficient to isolate DNA sequence specifically bound by R.

Figure 1A shows the progressive enrichment for R-binding sites obtained with this method. After the first selection round almost no retarded complexes between the D_{23} template and the GST-R protein were detected (Fig 1A, lane 2), as compared to the complexes obtained with the GST protein (Fig 1A, lane 1). For lane 2 the DNA was therefore excised from the entire lane above the non-retarded band. In the second and third selection rounds, DNA was isolated from an apparent single DNA-GST-R protein complex (Fig 1A, lanes 3 and 4). In the fourth selection round, a larger amount of complexes was detected with the GST-R protein (Figure 1A, lane 6), whereas no complexes were detected with the control GST protein (Figure 1A, lane 5). To ascertain that no further selection rounds were necessary, DNA amplified from the protein−DNA complexes formed in the fourth selection round or a double stranded oligonucleotide containing a single R-binding site, were incubated with the N-terminal 356 amino-acids of R translated *in vitro*. This polypeptide (called Rt356) contains the DNA-binding domain of R (21). As shown in Figure 1B, about 20% of both probes was recovered as DNA−protein complexes (Figure 1B, compare lanes 2 and 6), suggesting that the selection was complete. Binding of R to the selected DNA was specific since: (i) it was competed by an oligonucleotide containing an R-binding site (Figure 1B, lane 3) but not by an oligonucleotide where the R-binding site was destroyed by mutagenesis (Figure1B, lane 4), (ii) the R-DNA complexes were supershifted by a monoclonal antibody (8C12) directed against the N-terminal part of R, but not by a monoclonal antibody directed against the GST protein (1C11) (Figure 1C, lane 3 and 4). These results demonstrate that the DNA amplified after the fourth round of selection represents a pool of specific R-binding sites.

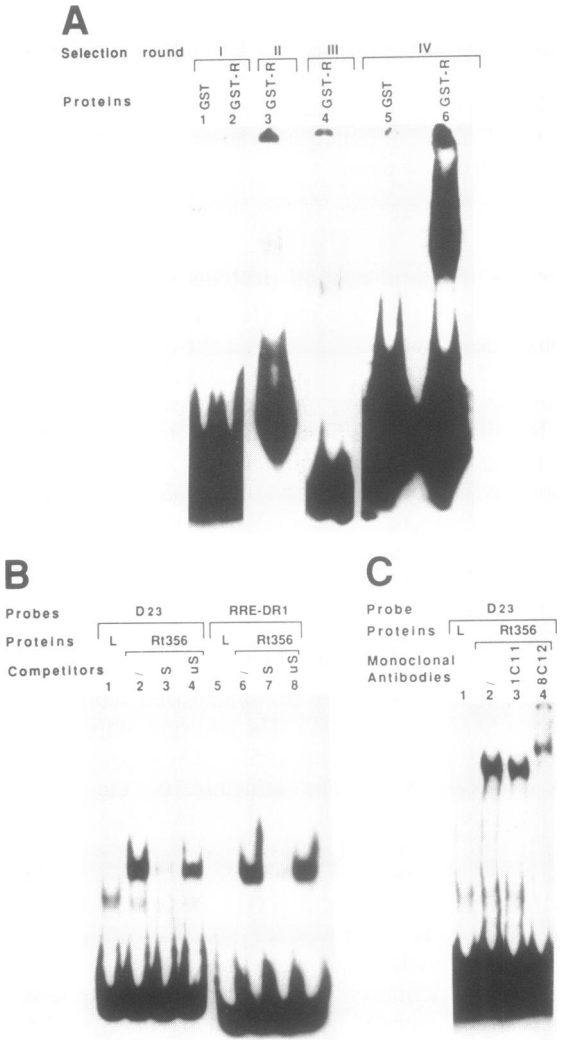

Figure 1. Selection of R-binding sites with oligonucletide D_{23}. (**A**). Gel mobility shift assay of labelled D_{23} oligonucleotides selected using the GST-R protein. I to IV: progressive enrichment in R-binding sites. GST: Gluthatione-S-Transferase protein. GST-R: fusion protein between GST and the R protein. **(B)** Competition assays. The selected D23 oligonucleotides (lanes 1 to 4) or an oligonucleotide carrying RRE-DR1 (lanes 5 to 8), 5′-^{32}P-labeled, were incubated with the Rabbit reticulocyte lysate (L, lanes 1 and 5), or with *in vitro* translated Rt356, an R mutant carrying the dimerization and DNA-binding domain but not the activation domain (lanes 2 to 4 and lanes 6 to 7). S: specific competition with an oligonucleotide carrying RRE-DR1 (lanes 3 and 7). uS: competition with a mutated RRE-DR1 oligonucleotide which does not bind R (lanes 4 and 8). **(C)** The selected D23 oligonucleotides, 5′-^{32}P-labeled, were incubated with the Rabbit reticulocyte lysate (L, lane 1), or with *in vitro* tranlated Rt356 (lanes 2 to 4). The Rt356-D_{23} complexes were supershifted by a monoclonal antibody directed against the N-terminal domain of R (8C12, lane 4), but not by a monoclonal antibody directed against the GST protein (1C11, lane 3).

Analysis of the selected R-binding sites

The DNA amplified from the protein−DNA complexes formed in the fourth selection round was digested by *Eco*RI and *Bam*HI, purified on a 10% polyacrylamide gel and cloned into the *Eco*RI and *Bam*HI sites of the pBluescript II cloning vector (Stratagene). Figure 2A shows a compilation of the 45 sites that were cloned and sequenced. The fixed sequences that permitted PCR amplification, contributed to 9 of the R-binding sites selected (cl1, cl7, cl8, cl21, cl27, cl35, cl41, cl54 and cl70), which might therefore be overrepresented. Among the sites selected two were identical except for the flanking sequences (cl3 and cl9). As shown in Figure 2B, alignment of the selected sites indicates that at positions 1 to 4 and 12 to 17, two sequences are highly conserved, G^{T}GCC and GTGGTG respectively. Between the two core elements, sequences are relatively random except at positions 5, 6, 10 and 11 where a G, a C, an A and a C are present with the respective frequencies of 45%, 67%, 56% and 60% respectively. We therefore next evaluated if a fixed 5′-core sequence with a C and T replacing the G and the C found at positions 5 and 6 would influence the composition of the R-

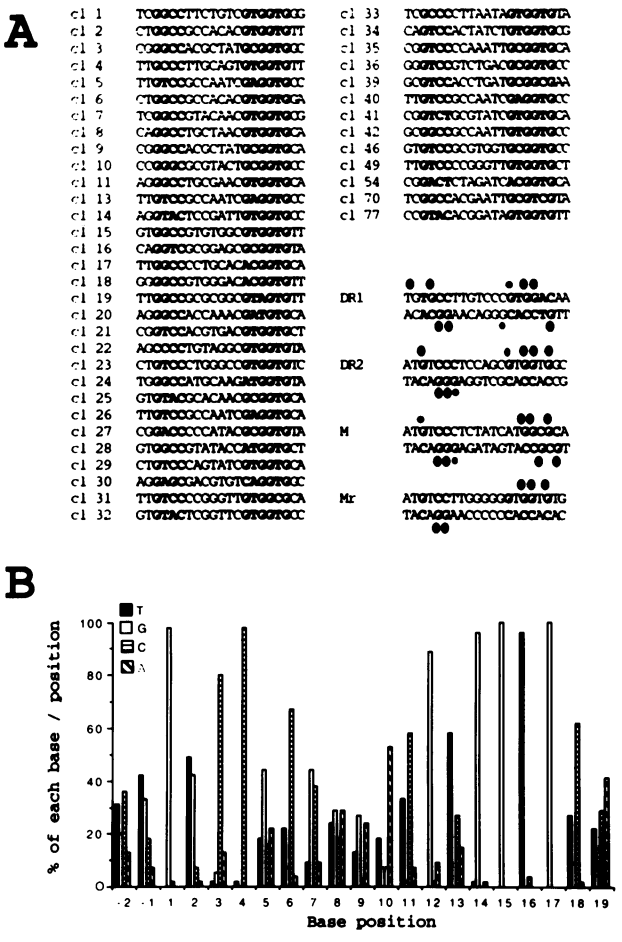

Figure 2. Sequence analysis of the selected R-binding sites using oligonucleotide D_{23}. (**A**) The sequence of 45 independent R-binding sites are presented together with the sequence of the known sites DR1, DR2, M and Mr. The most conserved sequences are indicated in bold letters. The black points over the DR1, DR2, M and Mr sequences represent methylation of guanines interfering with R-binding. Their size is relatively proportional to the degree of interference. (**B**) Percentage of each base (Y axis) at a particular position in the site (X axis). The first highly conserved base of the site is designated as position 1.

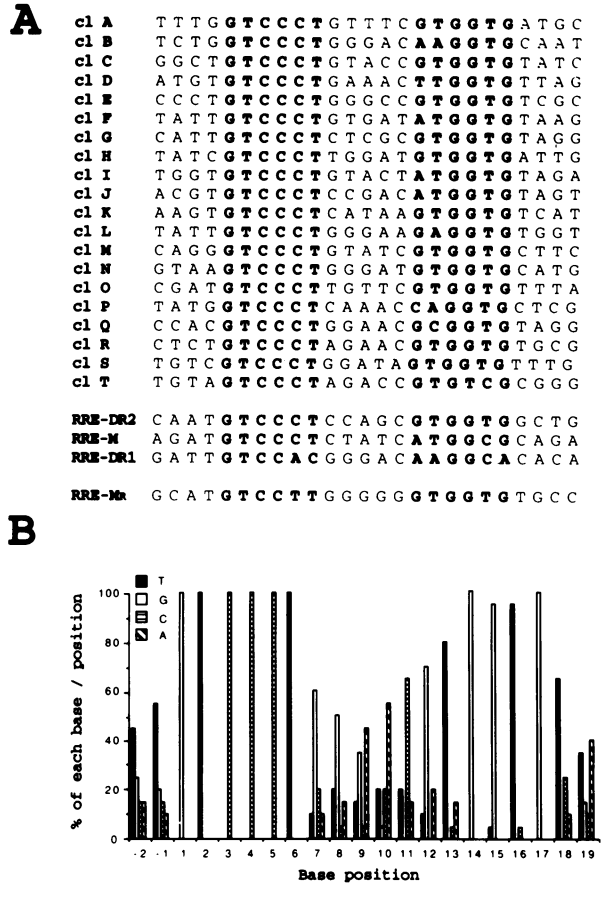

Figure 3. Sequence of R-binding sites selected with oligonucleotide D_A. (**A**) The sequence of 20 independent R-binding sites are presented together with the sequence of the known sites DR1, DR2, M and Mr. The fixed 5′ core sequence GTCCCT was found in all of the sites. The most conserved positions in the sequences are represented in bold letters. (**B**) Percentage of each base (Y axis) at a particular position in the site (X axis). The first highly conserved base of the site is designated as position 1.

binding sites. We did a second CASTing analysis with a partially degenerated oligonucleotide (D_A) containing the 5′ core element GTCCCT, which is also found in RREs DR2 and M (Figure 2A). Twenty R-binding sites were selected and sequenced (Figure 3A). Alignment of these sequences confirmed a high frequency for the 3′ core sequence GTGGTG (Figure 3B), and demonstrated that a particular 5′ core element did not influence the composition of the 3′ core element. At positions 7 to 11 the distribution of bases was relatively random, but again an A and a C were more frequent at positions 10 and 11 respectively.

Characterization of R-binding sites

To map precisely the site of R-binding within the selected sequences and to obtain some indication of the crucial R-DNA contacts, we used a guanine methylation interference assay. In this analysis, we compared a site containing the two core sequences found with the higher frequency (cl1), with sites where one or both core sequences diverged from the consensus. The guanine methylation interferences are shown for certain sites in Figure 4A, and are summarized for all the sites in Figure 4B.

Our results demonstrate that R simultaneously contacts crucial guanine residues in each core element of a binding site. Indeed, the methylation of guanines at positions 3 and 4 on one DNA strand (3′-C^{C}AGG-5′) and of Gs at positions 14, 15 and 17 on the other strand of the DNA (5′-GGTG-3′), interfered at hundred percent for the binding of R. If R contacts either of the other core element independently, we would have just observed fifty percent interference for each core element of a particular site. Furthermore, the results show that the sequences conserved between the different sites selected, are effectively implicated in the specific binding of R.

On all sites, we observed that methylation of guanines at positions −1, 1 and 2, in and flanking the conserved 5′ core element on the DNA strand 5′-G^{T}GCC-3′, also interfered with R-binding, although in general more weakly. For example in clones 14, 27 and 54, the methylation of guanines at positions −1 and 1, which are found with a frequency of 33% and 98% respectively, interfered only weakly with R-binding. However, in clones 49, 77 and in DR2 and M, we observed a stronger interference on methylation of the guanine at position 1. Finally, it is noteworthy that methylation of the guanine at position 1 always interfered, although weakly with R-binding suggesting that this guanine is part of the recognition site. On the other DNA

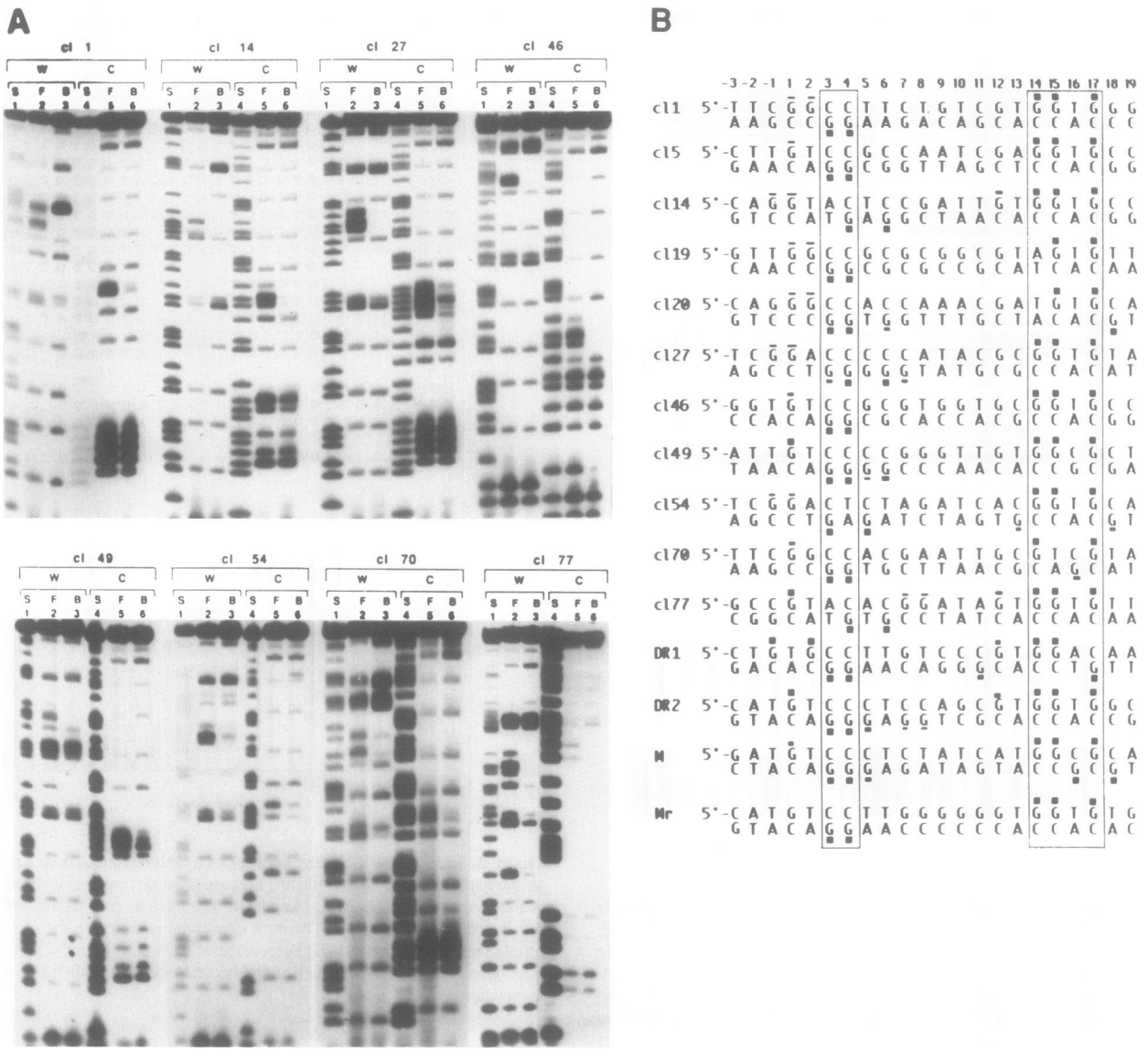

Figure 4. Identification of R-DNA contacts by guanine methylation interference. (**A**) DNA in which the W or the C strand was 5′-end labeled, was partially methylated by DMS and then incubated with R in a DNA-binding reaction. The bound (B) and free (F) DNA was separated by EMSA, cleaved at methylated guanine residues with piperidine, and analysed on a denaturating polyacrylamide gel. S: G+A sequence of the probe used in the binding assay. (**B**) Summary of the results presented in panel A. The W DNA strand is the upper one and the C DNA strand is the lower one in the sequences. The black squares over the guanines identify those positions where methylation interfered with R-binding. The size of the squares is relatively proportional to the degree of interference.

strand, although methylation of guanines when present at positions 5, 6, 7, 8 and 9, also interfered with R-binding, the interference was usually linked to a variation in the core sequence, and was in general weaker than in the core sequence.

For the 3′ core sequence on the DNA strand 5′-GTGGTG-3′, methylation of the Guanine at position 12, although present with a frequency of 90% in the selected sites, interfered weakly with R-binding and then only in clones 14 and 77. It should also be noted that in clone 1, the presence of a methylated Guanine at position 18 of the binding site seemed to facilitate the interaction between R and its DNA target sequence (Figure 4A, clone 1). In effect, in the DNA eluted from the R-containing complexes the frequency of Guanine residues at this position was higher than in the DNA eluted from non-complexed probe (Fig 3A, strand W, compare lane B with lane F). On the other DNA strand, methylation of the guanines when present at positions 11, 13, 16 and 11 or 18 also interfered with R-binding, but more rarely than in the core sequence.

In conclusion, the results of the guanine methylation experiments confirmed the importance of the two conserved core elements in R-binding, leading to the consensus binding sequence 5′-GNCC N_9 GGNG-3′. They also demonstrate that although in certain sites, R interacts with guanines flanking those core elements, in these cases, interference is often linked to variations in the core sequences.

Relative affinity of different selected sites for R

The fact that the majority of the selected sites have the sequence 5′-GNCC N_9 GGNG-3′ might be due to the higher affinity of R for these sites than for the other sequences selected. To test

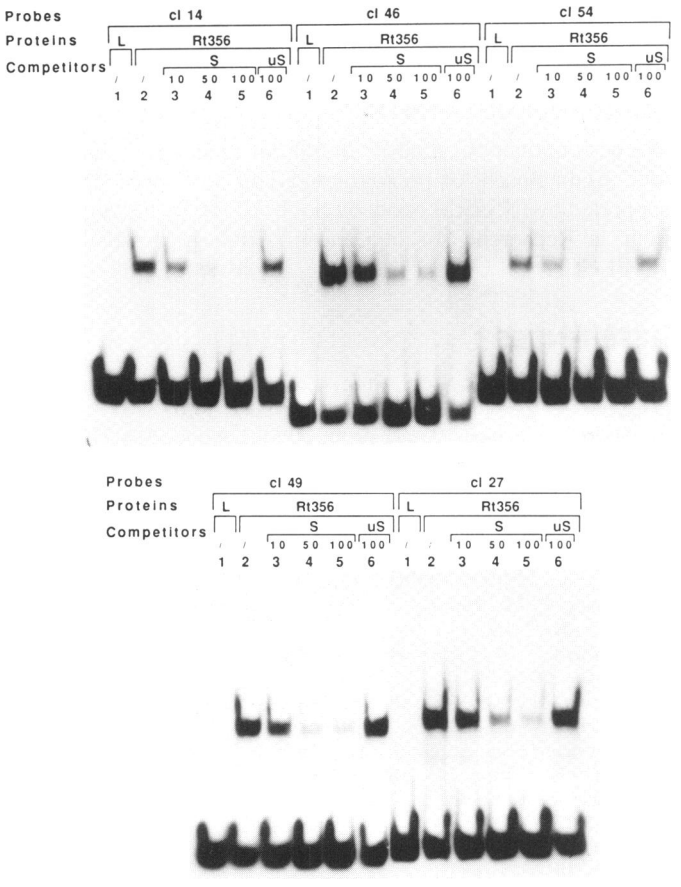

Figure 5. Relative affinity of R for D_{23}-selected sites. In this EMSA, the radiolabeled oligonucleotides cl14 , 46 , 54, 49 and 27 were incubated with the rabbit reticulocyte lysate (L, lanes 1), or with *in vitro* translated Rt356 (Lanes 2 to 6). The specific complexes formed (Lanes 2) were competed for with the cl1 unlabeled oligonucleotide (S, lanes 3 to 5)) or with the RRE-DR1* unlabeled oligonucleotide carrying a mutated R-binding site (uS, lanes 6). The excess of competitor added is indicated above each lane as a multiple of the concentration of the labeled probe.

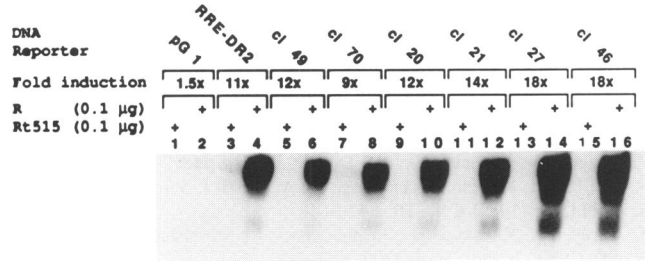

Figure 6. Selected R-binding sites mediate transcriptional activation by R. Plasmids carrying the indicated R-binding sites inserted upstream of the rabbit β-globin gene in plasmid pG1 (DNA reporter), were transfected in HeLa cells together with 0.1 µg of an Rt515-expressing vector (R deleted of the activation domain, lanes 1, 3, 5, 7, 9, 11, 13 and 15) or with an R-expressing vector (lanes 2, 4, 6, 8, 10, 12, 14 and 16). The level of specifically-initiated β-globin transcripts was evaluated by quantitative S1 mapping.

this possibility, several R-binding sites diverging from the consensus were assayed for binding in competition with cl1. As shown in figure 5, the cl1 sequence competed to about the same extent for the binding of R to the sites tested. With a ten fold excess of specific competitor, we observed a 20 to 30% decrease in the amount of R-DNA complexes formed on the binding sites tested (figure 5, lanes 3). With a 50 fold excess of specific competitor, we observed a decrease of approximately 70% in the amount of R-DNA complexes formed (Fig 5, lanes 4). Competition with a 100 fold excess of a non-specific oligonucleotide caused a very low decrease in complex formation (between 2 and 10%, Figure 5, lanes 6). These results strongly suggest that R binds to the selected sites with a similar affinity which is relatively independent of the composition of the core sequences.

The selected R-binding sites mediate the activation by R

To test if the selected R-binding sites were functional in transmitting R-mediated transcriptional activation, we cloned some of them upstream of the rabbit β-globin promoter in plasmid pG1 (23). The constructions were transfected into HeLa cells either with 0.1 µg of an R-expression vector or with 0.1 µg of an Rt515 expression vector. Rt515 is an R deletion mutant which contains the 515 N-terminal aminoacids carrying the DNA-binding/dimerization function but no activation domain (21). The transcriptional activity of the different β-globin constructs was evaluated by quantitative S1 analysis. Representative results are shown in figure 6. As expected, the rabbit β-globin promoter without an R-binding site had an undetectable basal activity in the presence of Rt515 and was not transactivated by R (Figure 6, lanes 1 and 2). Conversely, when linked to the RRE-DR2, the β-globin promoter was strongly activated by R as compared to its activity in the presence of Rt515 (Figure 6, lanes 3 and 4). Finally, the selected R-binding sites cloned upstream of the rabbit β-globin promoter were all strongly activated by R, with a comparable efficiency in their capacity to mediate transcriptional activation by R (Fig 6, compare lane 4 with lanes 6, 8, 10, 12, 14, 16).

Thus, at least in a transient expression assay, certain of the selected R-binding sites were able to mediate specific transcriptional activation by R, and this is likely to reflect the high affinity of R for these sites as visualized in the *in vitro* binding assay.

DISCUSSION

In order to determine the DNA sequences which are specifically recognized by the EBV transcription factor R, we performed a binding-site selection using a bacterially-expressed GST-R fusion protein and a pool of random DNA sequences. The DNA sequences of 45 selected binding sites was determined and their comparison revealed that some positions in the sequences were highly conserved, i.e., 5′-$G^G/_T$CC $(N)_7$ GTGGTG-3′. However, although our methylation interference data clearly demonstrated that R makes an intimate and simultaneous contact with the guanines in these two sequences, the R-binding site is more likely to be 5′-GNCC N_9 GGNG-3′ as 30 sites among the 45 sites selected have this sequence. In the other sites, variations such as GTAC, GGTC, CCCC, GAGC, GTCT and GACT are found in the 5′ sequence. Variations such as AGTG, TGTG and GTCG are also found in the 3′ sequence. This indicates that although the sequences directly contacted by R can tolerate some degree of variation, some positions in these sequences are highly conserved. Moreover, our transient expression assays indicate that when linked to the β-globin basal promoter, the R-binding

sites tested mediate a comparable transcriptional activation by R. This probably relates to the observation that R binds to these sites *in vitro* with a relative similar high affinity.

The guanine methylation interference data revealed the crucial contacts between DNA and R. Indeed, as shown in Figure 4B, methylation of the two guanines at position 3 and 4 on one DNA strand, and methylation of guanines at positions 14, 15 and 17 on the other DNA strand are crucial for R-binding. In addition, these two blocks of guanines are separated by 9 bp, suggesting that R binds two adjacent major grooves of the DNA. Some positions flanking the two core elements GNCC and GGNG in particular sites are also important for the interaction with R, but as yet we do not know if these interactions are critical both for the recognition and fixation of R to its target. However, in many of the sites examined, methylation of these sites interferes more weakly with R-binding. To exactly determine the contribution of each base in R-binding, site-directed mutagenesis is now necessary.

We have previously shown that R binds to DNA as a homodimer (19). From the guanine methylation studies we suggest that each R molecule in the dimer contacts a core element. In such cases the DNA-binding site normally has a palindromic structure as is the case for GCN4, Gal4, or Myc (24, 25, 26). The selected sites for R-binding have however a palindromic structure, since the guanines at positions 3 and 4 on one DNA strand and the guanines at positions 14 and 15 on the other DNA strand are crucial for R-binding. This allows a simpler representation of the prototype R-binding site: 5′-CC N_9 GG-3′. It could also be that the DNA-binding domain in the R protein is bipartite, with each sub-domain interacting with one core element, like it is the case for the Paired-domain (27). However, this hypothesis awaits further characterization of the R-DNA binding domain. It would be interesting to determine if related proteins such as the Herpesvirus Saimiri-R protein (HVS-R) (28) or the Bovine Herpesvirus 4 IE2 protein (BHV-4-IE2) (29) also bind to comparable sites. HVS-R and EBV-R are not functionally cross-reactive, since EBV-R does not transactivate a promoter inducible by HVS-R (28), but important homologies have been found between the DNA-binding domain of these proteins (28).

We searched in the EBV sequence (B95−8) for R-binding sites similar to those described here, and found more than 300 putatives sites. For example, using the sequence 5′-GTCC $(N)_9$ GGTG-3′ we have found 10 possible sites including of course RRE-DR2 and RRE-MR (see figure 2B). Nearly 30% of the putative sites are localized near a TATA box implicated in the transcription of early or late genes. As transient expression of R in latently infected B lymphocytes can induce the EBV genes required for progression to the lytic cycle, i.e., the early genes, it is reasonable to suppose that those promoters have one or more R-binding sites as in the case of the BMLF1 (14), the DL/DR (11) and the BMRF1 promoters (15). We did not find however, any putative R-binding site in the DNA polymerase gene promoter although it has been reported that transcription from this promoter can be activated by R (30). It has also been suggested that the promoter controlling the expression of mRNAs containing both the coding sequences for EB1 and R (pR promoter), can be activated by R in a transient expression assay, albeit at a very low level (31, 32). Again, we did not find any R-binding site in the pR promoter, and this was confirmed by a footprinting assay on the pR promoter using a nuclear extract which contained the R protein (H. Gruffat, unpublished results). It remains to be determined if the weak activation of the pR promoter is a direct or indirect effect of R, since R can weakly activate transcription from promoters that lack a specific binding site (33).

ACKNOWLEDGEMENTS

We wish to thank Conrad B. Bluink for reading the manuscript and. Alain Bosch for photography. This work was financially supported by INSERM (contract no. 871015), by the Association pour la Recherche sur le Cancer (ARC, no. 6810) and by FNCLCC.

REFERENCES

1. Kieff, E. and Liebowitz, D. (1990) in B. N. Fields and D. M. Knipe (ed.), Virology, Raven Press Ltd, New York, 1889−1920.
2. Miller, G., Rabson, M. and Heston (1984) J. Virol., **50**, 174−182.
3. Countryman, J. K. and Miller, G. (1985) Procd. Natl. Acad. Sci. USA, **82**, 4085−4089.
4. Countryman, J., Jenson, H., Grogan, E. and Miller, G. (1986) Cancer cells, **4**, 517−523.
5. Chevallier-Greco, A., Manet, E., Chavrier, P., Mosnier, C., Daillie, J. and Sergeant, A. (1986) EMBO J., **5**, 3243−3249.
6. Rooney, C. M., Rowe, D. T., Ragot, T. and Farrell, P. (1989) J. Virol., **63**, 3109−3116.
7. Hardwick, J. M., Lieberman, P. and Hayward, S. D. (1988) J. Virol., **62**, 2274−2284.
8. Buisson, M., Manet, E., Biemont, M. C., Durand, B. and Sergeant, A. (1989) J. Virol., **63**, 5276−5284.
9. Chevallier-Greco, A., Gruffat, H., Manet, E., Calender, A. and Sergeant, A. (1989) J. Virol., **63**, 615−623.
10. Kenney, S., Holley-Guthrie, E., Mar, E. C. and Smith, M. (1989) J. Virol., **63**, 3878−3883.
11. Gruffat, H., Manet, E., Rigolet, A. and Sergeant, A. (1990) Nucleic Acids Res., **18**, 6835−6843.
12. Cox, M. A., Leahy, J. and Hardwick, M. (1990) J. Virol., **64**, 313−321.
13. Holley-Guthrie, E. A., Quinlivan, E. B., Mar, E. C. and Kenney, S (1990) J. Virol., **64**, 3753−3759.
14. Gruffat, H., Duran, N., Buisson, M., Wild, F., Buckland, R. and Sergeant, A. (1992) J. Virol., **66**, 46−52.
15. Quinlivan, E. B., Holley-Guthrie, E. A., Norris, M., Gutsch, D., Bachenheimer, S. L. and Kenney, S. (1993) Nucleic Acids Res., **21**, 1999−2007.
16. Hammerschmidt, W. and Sugden, B. (1988) Cell, **55**, 427−433.
17. Fixam, E. D., Hayward, G. S. and Hayward, S. D. (1992) J. Virol., **66**, 5030−5039.
18. Schepers, A., Pich, D. and Hammerschmidt, W. (1993) EMBO J., **12**, 3921−3929.
19. Manet, E., Rigolet, A., Gruffat, H., Giot J. F. and Sergeant, A. (1991) Nucleic Acids Res., **10**, 2661−2667.
20. Hardwick, J. M., Tse, L., Applegren, N., Nicholas, J. and Veliuone, M. A. (1992) J. Virol., **66**, 5500−5508.
21. Manet, E., Allera, C., Gruffat, H., Mikaelian, I., Rigolet, A. and Sergeant, A. (1993) Gene Exp., **3**, 49−59.
22. Wright, W. E. and Funk, W. D. (1993) TIBS, **18**, 77−80.
23. Gruffat, H., Moreno, N. and Sergeant, A. (1990) J. Virol., **64**, 2810−2818.
24. Struhl, K. (1987) Cell, **50**, 841−846.
25. Carey, M., Kakidani, H., Leatherwood, J., Mostashari, F. and Ptashne, M. (1989) J. Mol. Biol., **209**, 423−432.
26. Blackwell, T. K., Kretzner, L., Blackwood, E. M., Eisenman, R. N. and Weintraub, H. (1990) Science, **250**, 1149−1151.
27. Czerny, T., Schaffner, G. and Busslinger, M. (1993) Genes and Dev., **7**, 2048−2061.
28. Nicholas, J., Coles, L. S., Newman, C. and Honess, R. W. (1991) J. Virol., **65**, 2457−2466.
29. van Santen, V. (1993) J. Virol., **67**, 773−784.
30. Furnari, F. B., Adams, M. D. and Pagano, J. (1992) J. Virol., **66**, 2837−2845.
31. Sinclair, A. J., Brimmel, M., Sanahan, F. and Farrell, P. J. (1991) J. Virol., **65**, 2237−2244.
32. Zalani, S., Holley-Guthrie, E. A., Gutsch, D. E. and Kenney, S. C. (1992) J. Virol., **66**, 7282−7292.
33. Giot, J. F., Mikaelian, I., Buisson, M., Manet, E., Joab, I., Nicolas, J. C. and Sergeant, A. (1991) Nucleic Acids Res., **19**, 1251−1258.