Alternating purine–pyrimidine tract activates transcription from the Rouse sarcoma virus LTR lacking promoter and enhancer elements

Wladyslaw A. Krajewski*

Laboratory of Biochemistry, Institute of Developmental Biology, Russian Academy of Science, Vavilova str. 26, Moscow 117808, Russian Federation

Received 26 October 1994; revised version received 28 November 1994

Abstract The transcriptional control region of the Rouse sarcoma virus long terminal repeats (LTR) was shown to contain enhancer and promoter elements located within 200 base pairs upstream from the transcription initiation site [Cullen et al. (1985) Mol. Cell. Biol. 5, 438–447]. Deletion of these elements results in significant loss of LTR transcriptional activity. In the present paper it is shown that a short alternating purine–pyrimidine sequence can restore the constitutive activity of the Rouse sarcoma virus LTR in the absence of upstream elements when inserted in close proximity to the transcription initiator site. The possible molecular bases of this phenomena are discussed.

Key words: Rouse sarcoma virus; Transcriptional control region; Enhancer; Gene regulation; Histone; Nucleosome; Z-DNA

1. Introduction

Mutational analysis of a number of eukaryotic transcriptional control regions (TCR) has led to significant advances in the study of the DNA elements that modulate activity of eukaryotic genes. At present, one can identify at least three functionally distinct but cooperatively acting elements that are normally required for efficient and precise in vivo transcription (reviewed in [1,2]), namely: (i) the 'TATA' block located within the transcription initiation or cap site, (ii) the upstream sequence or promoter element normally located about 100 bp 5′ of the TATA block, and (iii) the third functional element, termed 'enhancer', capable of activating promoters regardless of orientation and from significant distances when present in *cis* on DNA.

Whereas the three described above functional elements can be identified within the majority of eukaryotic gene control regions, the participation of these elements in the control of gene expression still remains somewhat poorly understood. The possible molecular mechanisms presumably involve formation of interacting multi-enzyme complexes on all the TCR elements (Fig. 4B) and might be grouped into two general, and not mutually exclusive categories, namely (i) the mechanisms which mediate the rearrangement of the chromatin environment to establish a gene control region structure accessible by transcription components [3–5], and (ii) the mechanisms which facilitate formation of active transcription pre-initiation complexes by interacting with RNA pol II machinery [1,2,6]. To assess the relative importance of these mechanisms would be of significant advantage in understanding the molecular bases of regulation of gene activity.

*Corresponding author. Fax: (7) (095) 135 8012.

In this respect the avian retroviral long terminal repeats (LTR) could be of particular interest. These sequences, that flank either end of the integrated proviral genome, were shown to be strong promoters when introduced into a variety of eukaryotic cells. Cullen et al. have demonstrated that Rouse sarcoma virus LTRs alone contain all of the functional elements required for efficient transcription [7,8]. Using site-directed mutagenesis the promoter and enhancer elements were defined and were shown to locate within the 200 base pairs upstream from the transcription initiation site [8]. Deletion of these elements results in significant loss of LTR transcriptional activity.

In the present paper we have shown that a potential Z-DNA forming sequence, previously reported to be resistant to nucleosome formation [9], can restore the constitutive activity of the Rouse sarcoma virus LTR in the absence of upstream-activating elements when inserted in close proximity to the transcription initiator site. The possible molecular bases of this phenomenon are discussed and a conceivable model is presented, which illustrates the assembly of preinitiation complexes in vivo.

2. Materials and methods

Preparation of plasmid DNAs, DNA fragments, and bacterial transformation were performed as described elsewhere [10].

Eukaryotic cell transfection were carried out only by using calcium phosphate [11] and contained 5 μg of promoter construction, and pGEM3 to make a total of 20 μg of DNA per 60 mm culture dish. All experiments were performed in triplicate with different preparation of reporter plasmids.

Assays of CAT activity were performed essentially as described by Tsukada et al. [11], and chloramphenicol acetylated forms were resolved by ascending chromatography in a 95% chloroform/5% methanol system.

Core histones were associated with gel-purified supercoiled and relaxed forms of pZ^+ and pZ^- plasmids by a gradual salt dilution procedure described previously [12,13]; core histones were obtained by acid fractionation of isolated nuclei [14]. Micrococcal nuclease digestion was performed to yield monomeric nucleosomes [14,15], and dot hybridization with ^{32}P-labeled initiator site oligonucleotide (5′cgagctaggcacttaaattacatactct3′) or with multiprimer-labeled plasmids pZ^+ and pZ^- was performed by a standard dextran sulfate procedure as described elsewhere [15].

Nuclear extracts were prepared essentially as was previously described [16], except protease inhibitors (PMSF, leupeptin, pepstatin, chymostatin, triasylol and TPCK) were added in appropriate concentrations to all the solutions and 0.2% Nonidet P40 was added in the cell lysis buffer.

Transcription of assembled templates was assayed by mixing of equal volumes of assembled templates (0.5–1 μg of template DNA) and HeLa cell nuclear extracts (20–40 μg of total protein) in a final volume of 10–20 μl supplemented with 500 μM (each) ATP, GTP, CTP, and UTP; 5 U of RNAsin. The reaction mixture was then incubated for 5 min at 26°C. RNA products were purified, analyzed in urea-containing 5% polyacrylamide gels and visualized by hybridization with a ^{32}P-labeled CAT gene structural sequence as probe as described elsewhere [15].

All autoradiograms were quantified either by scintillation counting

SSDI 0014-5793(94)01377-2

of isolated radioactive spots and/or by densitometer reading of the films.

3. Results

Plasmids pZ^+/pZ^- are the pRSVΔECAT [11] derivatives containing the bacterial chloramphenicol acetyltransferase gene structural sequence (CAT; acetyl-CoA:chloramphenicol 3-O-acetyltransferase; EC 2.3.1.28) fused to the Rouse sarcoma virus (RSV; Schmidt Ruppin D strain) initiator region fragment, consisting of 51 bp 5′-flanking and 39 bp transcribed sequences [12], modified with two sites for interaction with the B-cell specific activator protein (BSAP; described by Adams et al., [17]) inserted further upstream, similarly as described by Krajewski and Lee [18]. Double-stranded oligonucleotides of the same length (14 bp) but differing in (CG) pairs content were subcloned immediately 5′ upstream from the transcription initiator site (*Sac*I restriction site), yielding the plasmids pZ^+ and pZ^- with respect to their ability to a form Z-DNA region: plasmid pZ^- contains a 14 bp random sequence insertion, whereas pZ^+ contains a perfect 14 bp tract of alternating CGs, the sequence that easily adopts a left-handed conformation in response to superhelical stress.

To compare the in vivo expression level of the obtained constructs we assayed CAT activity in extracts from cells transfected with the corresponding CAT fusion. Fig. 1 shows measurements of CAT activity in JEG3 extracts made 48 h after the introduction of plasmid DNAs. The results clearly demonstrate that there was considerably more CAT enzymatic activity in cells transfected with TCR constructs containing the 14 bp alternating CG motif. This plasmid yielded more than 50 times the level of CAT activity that was induced by the plasmid with a random sequence promoter insertion.

To assess the molecular mechanism of the described phenomenon and to dissect possible effects from its constitutive

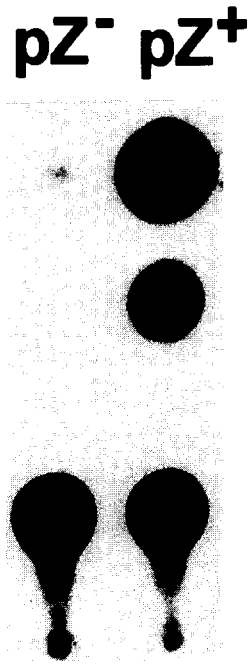

Fig. 1. CAT activity in JEG3 cells transfected with the pZ^+/pZ^- plasmids; activity was assayed 48 h after introduction of plasmid DNAs.

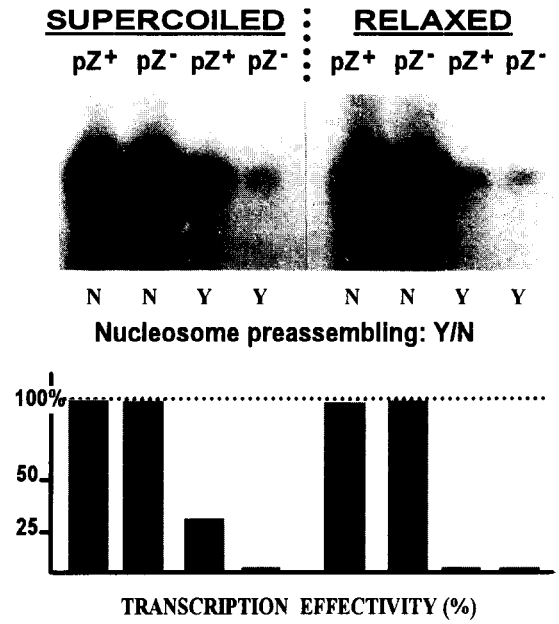

Fig. 2. Transcriptional repression of DNA templates assembled in chromatin. Nucleosomes were reconstituted on supercoiled and relaxed forms of plasmids pZ^+ and pZ^-, as indicated, and portions of the assembly reaction were incubated in the transcription mixture supplemented with extracts from HeLa cells. The RNA products derived from each template were visualized by hybridization with a ^{32}P-labeled probe corresponding to the CAT structural sequence. Unfortunately, the relative band intensities were perverted upon reproduction of the figure.

system elements we have remodeled this situation in a coupled in vitro nucleosome assembly and cell-free transcription system.

Transcription was assayed directly in crude HeLa cell nuclear extracts [16] after nucleosome assembly was complete. The autoradiogram demonstrates that the in vitro nucleosome assembly results in the disappearance of the corresponding RNA product for all the reconstituted plasmid templates except the supercoiled template containing the 14 bp alternating CG tract within the TCR region (Fig. 2). The autoradiograms revealed no marked differences in the mRNA levels produced by naked plasmid templates pZ^+ and pZ^-, implying that the observed differences in transcriptional activity of different forms of pZ^+ and pZ^- are presumably bound to the structural peculiarities of their association with nucleosomes.

Nucleosome core particles were assembled on supercoiled and relaxed forms of pZ^+ and pZ^- plasmids, and the transcription initiator region was tested for the presence of nucleosome organization by prolonged digestion with micrococcal nuclease (Fig. 3). DNA within the nucleosome core is protected by histones to micrococcal nuclease cleavage, and the presence of a histone octamer at a definite location can be detected by the accumulation of the nuclease-resistant fragments containing the corresponding DNA sequence. The nucleosome protection pattern of the promoter region revealed by dot hybridization (Fig. 3) clearly shows that, whereas an intact nucleosome has been assembled at the transcription start region in the supercoiled plasmid with the random sequence promoter insertion (pZ^-), in the supercoiled plasmid containing the 14 bp alternating CG motif within the promoter (pZ^+) the nucleosome abun-

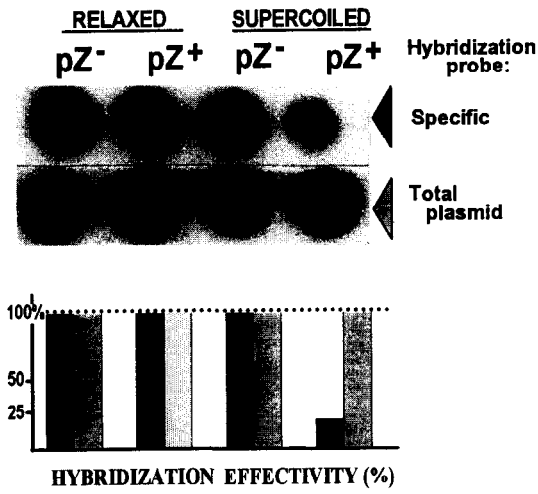

Fig. 3. Micrococcal nuclease analysis of reconstituted plasmid chromatin (micrococcal nuclease digestion was performed to yield monomeric nucleosomes). The nucleosome protection pattern of the Rouse sarcoma virus LTR transcription initiator region revealed by dot hybridization.

dance for the transcription initiator site was decreased up to four- to five-fold. In the case of relaxed forms of pZ^+ and pZ^- plasmids the nucleosome protection patterns produced were similar (Fig. 3). The topoisomerase activity of HeLa cell nuclear extracts we used was comparatively low: the detectable relaxation of supercoiled plasmids was observed only after 10 min of incubation (data not shown). The nucleosome exclusion, in principle, might be responsible for the observed difference in an in vivo transcriptional activity between pZ^+ and pZ^- plasmids.

4. Discussion

Using site-directed deletion mutagenesis, the Rouse sarcoma virus long terminal repeats were shown to be divided into the three following functional elements [8] (Fig. 4B): (i) the enhancer extending from position −219 to −139 (80 bp), (ii) the promoter extending from position −135 to −45 (90 bp), and (iii) the site of transcription initiation including the TATA block itself (−30 to −24). Deletion of the enhancer and promoter elements results in an about 50-fold decrease in in vivo transcriptional activity [8,35].

We have previously shown that in vivo transcription from the RSV initiator site could be restored in the absence of the RSV upstream functional domain if positive transcription factor binding sites are used instead of sequences lacking promoter and enhancer elements [18]. In this case transcription is fully dependent on the presence of activator protein that presumably mimics the functions of the RSV LTR transcription initiation machinery (Fig. 4B,C). However, if a short (14) bp alternating CG tract is placed immediately upstream of the RSV transcription initiator site, the level of its constitutive (non-induced) transcription increases about 50–60 fold (Fig. 1), and upon further activation the level of transcription increases only up to 8–10 fold (data not shown).

What might be the molecular basis of this phenomena? The

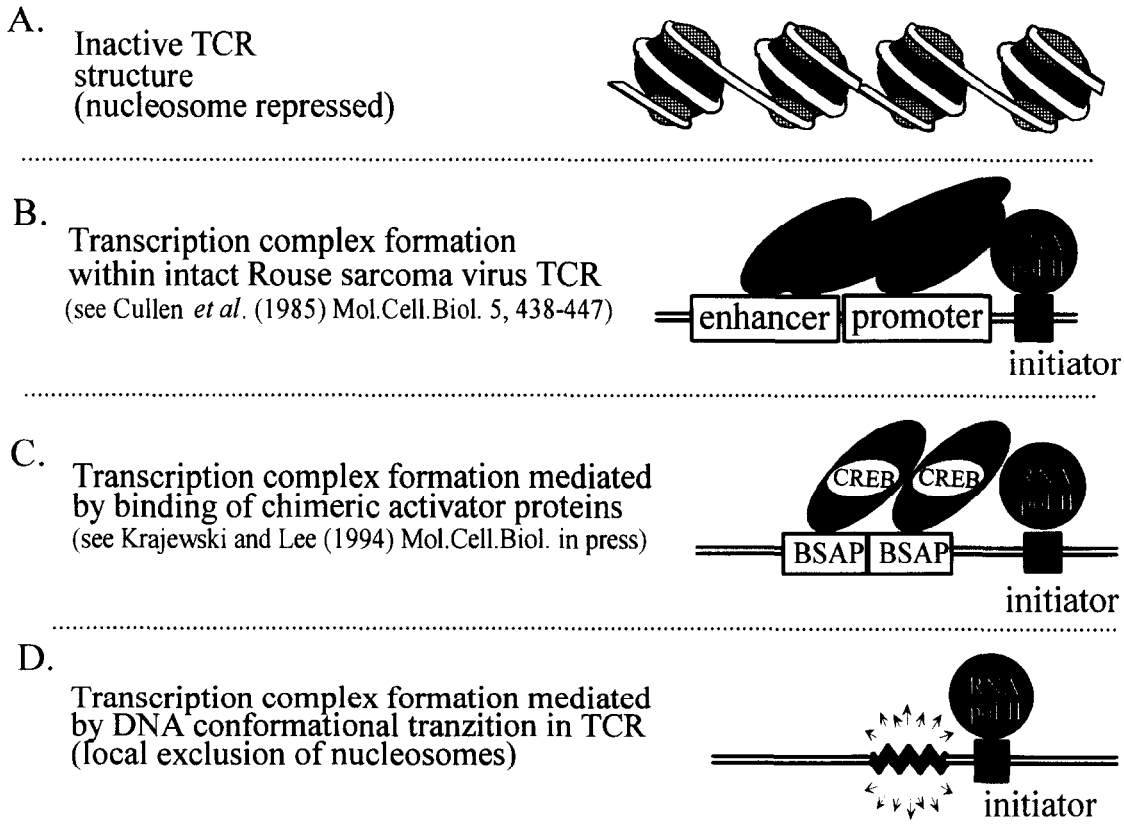

Fig. 4. Schematic representation of pre-initiation complex assembly within Rouse sarcoma virus LTR transcriptional control region clones.

alternating CG sequence is both a direct repeat and an inverted repeat, and its relatively short stretches (8–10 bp or more) can readily adopt a left-handed helix (Z-DNA) or a cruciform structure in response to superhelical stress [19–21]. Cruciforms and Z-DNA present an essentially different structures to the regular B form of DNA, and their association with core histones was concluded to be energetically disfavored (reviewed in [22,23]). In a number of studies these DNA isoforms were shown to be resistant to in vitro nucleosome packaging, resulting in an alternative phasing of the histone octamers along the DNA [22,24,25], that implies their possible role in an vivo refasing of nucleosomes. Although the bulk of DNA in eukaryotic chromatin is generally believed to be in a relaxed state [26], now there is a great body of evidence that local regions of the chromosomes, presumably active chromatin domains, could undergo continuous changes in superhelical stress [22,23,27]. Torsional tensions in chromatin could arise in vivo as a result of processing polymerases: the 'twin supercoil domain' model of Liu and Wang [28,29], as well as exist independently from ongoing transcription [30]. The in vivo existence of Z-DNA was successively demonstrated in a number of mammalian species [22,31,32] as well as in *Drosophila* [33] and *Xenopus laevis* cells [30].

Therefore, the results we presented could be conceivably explained if we accept that under conditions of unrestrained superhelical stress, the potential regions of Z-DNA could exclude a nucleosome from the surrounding region, thus positioning it outside the RSV transcription initiation site (Fig. 4D). The nucleosome exclusion provides free access for protein factors and transcription components to DNA that alone might be adequate for the direct assembling of the pre-initiation complex even in the absence of upstream *cis*-activating TCR elements. The Rouse sarcoma virus LTR transcription control region was shown to provide high levels of activity in an unusually wide variety of cells [7], which means that the structure of its transcription initiator site is sufficiently 'universal' to direct the correct formation of pre-initiation complexes. Thus, the results we obtain imply that the upstream functional domains in the intact Rouse sarcoma virus LTR may be responsible not only for the precise settling of the RNA pol II complexes in the initiator region, but for alleviation of the steric constraints of the chromatin environment as well. In this respect, it is most intriguing that intact Rouse sarcoma virus LTR itself contains a 9 bp alternating purine–pyrimidine tract that extends from position −135 to −144 [8], that was proposed by Nordheim and Rich [37] to be an important part of the RSV LTR enhancer, although it was also shown that there is no direct dependence between the presence of potential Z-DNA forming sequences and the level of enhancer activity in avian retroviral LTRs [8,35].

Although the model we propose for the functioning of the avian retroviral TCR somewhat resembles the one proposed previously for SV40 early TCR [36], we have presented the first evidence of the striking effect of potential Z-DNA forming sequences on eukaryotic TCR activity in vivo. However, the general significance of the proposed mechanism and its relationship to the enhancer–promoter interactions in general still remains to be tested.

References

[1] Drapkin, R., Merino, A. and Reinburg, D. (1993) Curr. Opin. Cell Biol. 5, 469–476.
[2] Kornberg, R.D. and Lorch, Y. (1992) Annu. Rev. Cell Biol. 8, 563–587.
[3] Gross, D.S., Adams, C.C., Lee, S. and Stentz, B. (1993) EMBO J. 12, 3931–3945.
[4] Batson, S.C., Rimsky, S., Sundseth, R. and Hansen, U. (1993) Nucleic Acids Res. 21, 3459–3468.
[5] Workman, J.L. and Buchman, A.R. (1993) Trends Biochem. Sci. 18, 90–95.
[6] Hayes, J.J. and Wolffe, A.P. (1992) BioEssays 14, 1–7.
[7] Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. and Howard, B.H. (1982) Proc. Natl. Acad. Sci. USA 79, 6777–6781.
[8] Cullen, B.R., Raymond, K. and Grace, J. (1985) Mol. Cell. Biol. 5, 438–447.
[9] Garner, M.M. and Felsenfeld, G. (1987) J. Mol. Biol. 196, 581–590.
[10] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY.
[11] Tsukada, T., Fink, J.S., Mandel, G. and Goodman, R.H. (1987) J. Biol. Chem. 262, 8743–8747.
[12] Camerini-Otero, R.D. and Felsenfeld, G. (1977) Nucleic Acids Res. 4, 1159–1181.
[13] Hansen, J.C., Ausio, J., Stanik, V.H. and Van Holde, K.E. (1989) Biochemistry 28, 9129 9136.
[14] Krajewski, W.A., Panin, V.M. and Razin, S.V. (1993) J. Biomol. Struct. Dyn. 10, 1011–1022.
[15] Krajewski, W.A. and Luchnik, A.N. (1991) Mol. Gen. Genet. 230, 442–448.
[16] Andrews, N.C. and Faller, D.V. (1991) Nucleic Acids Res. 19, 2499.
[17] Adams, B., Dorfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurer-Fogy, I. and Busslinger, M. (1992) Genes Dev. 6, 1589–1607.
[18] Krajewski, W.A. and Lee, K.A.W. (1994) Mol. Cell. Biol. (in press).
[19] McLean, M.J., Lee, J.W. and Wells, R.D.J. (1988) Biol. Chem. 263, 7378–7385.
[20] Wells, R.D. and Harvey, S.C. (1987) Unusual DNA Structures. Springer-Verlag, New York.
[21] Wells, R.D. (1988) J. Biol. Chem. 263, 1095–1098.
[22] Esposito, F. and Sinden, R.R. (1988) Oxf. Surv. Eukaryot. Genes 5, 1–50.
[23] Freeman, L.A. and Garrard, W.T. (1992) Crit. Rev. Eukaryot Gene Expr. 2, 165–209.
[24] Garner, M.M. and Felsenfeld, G. (1987) J. Mol. Biol. 196, 581–590.
[25] Nobile, C., Nickol, J. and Martin, R.G. (1986) Mol. Cell. Biol. 6 2916–2922.
[26] Sinden, R.R., Carlson, J.O. and Pettijohn, D.E. (1980) Cell 21, 773–783.
[27] Saavedra, R.A. (1990) BioEssays 12, 125–128.
[28] Liu, L.F. and Wang, J.C. (1987) Proc. Natl. Acad. Sci. USA 84, 7024–7027.
[29] Osborne, B.I. and Guarente, L. (1988) Genes Dev. 2, 766–772.
[30] Leonard, M.W. and Patient, R.K. (1991) Mol. Cell. Biol. 11, 6128–6138.
[31] Wittig, B., Dorbic, T. and Rich, A. (1991) Proc. Natl. Acad. Sci. USA 88, 2259–2263.
[32] Wittig, B., Wolfl, S., Dorbic, T., Vahrson, W. and Rich, A. (1992) EMBO J. 11, 4653–4663.
[33] Lancillotti, F., Lopez, M.C., Arias, P. and Alonso, C. (1987) Proc. Natl. Acad. Sci. USA 84, 1560–1564.
[34] Nordheim, A. and Rich, A. (1983) Nature 303, 674–678.
[35] Cullen, B.R., Raymond, K. and Ju, G. (1985) J. Virol. 53, 515–521.
[36] Gruskin, E.A. and Rich, A. (1993) Biochemistry 32, 2167–2176.