The NssBF element, a sequence of the *Drosophila melanogaster* retrotransposon 1731 potentially implicated in transcriptional repression and replication

Jérôme Lacoste, Frédérique Fourcade-Peronnet*

UA CNRS 1135, Groupe de Génétique Cellulaire et Moléculaire, Université P. et M. Curie, 7, quai St-Bernard, 75005 Paris, France

Received 18 November 1994

Abstract The nuclear single-stranded DNA binding factor (NssBF) has been characterized as a nuclear protein that binds to a 26 nucleotides sequence in the long terminal repeat (LTR) of the *Drosophila melanogaster* 1731 retrotransposon. This sequence, called NssBF element, was analysed by gel retardation experiments using wild-type and mutated oligonucleotides. In vitro transcription experiments were performed and suggest that NssBF element binding protein(s) represses transcription through the 1731 promoter. Furthermore, computer assisted sequence comparisons put forward a possible role of this element and/or its associated DNA binding proteins in replication.

Key words: Retrotransposon 1731; *Drosophila melanogaster*; Single-stranded DNA binding protein; Repression; Replication

1. Introduction

Retrotransposons are mobile genetic elements similar to the provirus form of Vertebrates retroviruses (for reviews see [1,2]). Many mutations in *Drosophila melanogaster* are due to the genomic insertion of such transposable elements. In the past few years, it has been demonstrated that retrotransposition occurs via reverse transcription of retrotransposons RNA [3–5]. The existence of a link between transcription and retrotransposition levels was postulated [6]. This hypothesis led us to examine the transcriptional regulation of these elements. The expression of the *Drosophila melanogaster* retrotransposon 1731 was shown to be enhanced by heat-shock and repressed by 20-hydroxyecdysone, the steroid molting hormone of insects [7–9]. Functional analysis of the 1731 promoter, i.e. the 1731 long terminal repeats (LTR), has allowed to localize regulating sequences in the U3 part of the LTR (positions −168 to +1 with regard to the transcription initiation site) [8,9]. Furthermore, computer analysis have revealed several conserved boxes upstream the 1731 TATA-box that might act as regulatory signals. Gel shift experiments with *Drosophila melanogaster* Kc cell nuclear extracts and synthetic oligonucleotides covering this region have confirmed that multiple proteic factors do bind to this region [10–12].

Among these factors, the Nuclear single-stranded DNA Binding Factor (NssBF) has been characterized as a nuclear protein that specifically binds in vitro to single-stranded DNA between positions −99 and −74 of the LTR with regard to the transcription start of 1731 (+1), i.e. just upstream the TATA-box [10]. Numerous single-stranded DNA binding proteins that exhibit sequence specificity have been described in eukaryotic cells in the last few years. They have been shown to be implicated in several functions such as recombination, replication or transcription (see [10,13–16] and refs. therein). Due to the close proximity of the NssBF binding site and the TATA-box, we have been interested in the possible role of this protein in the regulation of 1731 transcription.

Contact points between NssBF and the 1731 LTR have been previously determined [10] (Fig. 1A). They are distributed into two distinct domains 6 nucleotides and 15 nucleotides long, characterized by an alternation of purines and pyrimidines, and separated by the five nucleotides GAACA. The whole sequence is called NssBF element. It is of interest to note the presence of a palindromic sequence, emphasized by horizontal arrows on Fig. 1A, which might induced the formation of an hairpin.

In the present work, we have analysed this element by gel shift experiments using mutated oligonucleotides and we have shown that in vitro transcription of the 1731 promoter is repressed by NssBF element binding protein(s). Computer analysis indicate similitudes between the NssBF element and chromosomal origins of replication.

2. Materials and methods

2.1. Cells and cell culture

The *Drosophila melanogaster* Kc cells were used throughout these experiments [17]. Cells were grown in Falcon flasks at 23°C in D22 medium supplemented with 5% fetal calf serum.

2.2. Preparation of nuclear proteins

Nuclear proteins were extracted from Kc cells and partially purified by heparin-agarose chromatography as previously described [10].

2.3. Oligodeoxynucleotides and gel retardation assays

Oligodeoxynucleotide Bc (taGTAATTTTGTATGAGAACATACATACATACACATGca) contains the NssBF element (the contact points between NssBF and single-stranded DNA are underlined) [10]. Oligodeoxynucleotide Bnc is the Bc complementary sequence. Oligodeoxynucleotide Bc1 (GTATGATACATACATACACAT) lacks the intermediate pentanucleotide GAACA and the NssBF element flanking sequences. Oligodeoxynucleotide Bc2 (GTATGAGAACATACATACATACACAT) lacks the flanking sequences whereas oligodeoxynucleotide Bc3(taGTAATTTTGTATGATACATACATACACATGca) lacks the intermediate pentanucleotide GAACA. In oligodeoxynucleotide Bc4 (taGTAATTTTGTATGAAGGTGTACATACATACACATGca), the pentanucleotide GAACA was changed for AGGTG. These oligonucleotides were purchased from Appligene. They were 5′ end-labeled with [γ-^{32}P]ATP using bacteriophage T4 polynucleotide kinase. Gel retardation assays were performed as previously described [10,12].

2.4. In vitro transcription assays

In vitro transcription was performed with an in vitro eukaryotic

*Corresponding author. Fax: (33) (1) 44 27 32 06.
E-mail: Frederique.Peronnet@snv.jussieu.fr

SSDI 0014-5793(94)01373-X

transcription kit, following the indications of the supplier (Stratagene). The active *Drosophila melanogaster* embryo extract was incubated with the Bc oligonucleotide in order to be depleted in NssBF element binding protein(s). Control experiments were carried out using the mutated oligonucleotide Bc3. The D3 plasmid, which contains the NssBF element and the 1731 core promoter (positions −111 to +89) upstream the chloramphenicol acetyl-transferase (CAT) gene and the SV40 polyadenylation site, was used as a template [8]. About 400 ng of template and 0 to 100 molar excess oligonucleotide were used for each reaction. A synthetic 30-mer antisense oligonucleotide of the CAT gene (positions +52 to +78) [18], labeled with [γ-^{32}P]ATP, was used to detect 1731 transcripts. The products of primer extension were resuspended in formamide-dye mix and subjected to electrophoresis on a 6% polyacrylamide gel containing 8 M urea. The DNA sequence used as a marker was generated using the above 30-mer as the primer in dideoxy sequencing reactions on the double-stranded D3 template previously denaturated in 0.2 N NaOH. The transcription reactions were performed 4 times, and data presented are representative of these experiments.

3. Results

3.1. The NssBF element

Nuclear extracts of *Drosophila* cells were partially purified by heparin-agarose chromatography. The NssBF containing fractions (NssBF is eluted at about 0.5 M KCl), [10], were used for gel retardation assays with oligonucleotides exhibiting the wild-type NssBF element (oligonucleotide Bc), no NssBF element (oligonucleotide Bnc) or mutated NssBF elements (oligonucleotides Bc1, Bc2, Bc3 and Bc4). The results are shown on Fig. 1B. Restriction of the oligonucleotide to the previously determined NssBF binding site (oligonucleotide Bc2) as well as deletion of the pentanucleotide GAACA (oligonucleotide Bc3) suppressed the binding. On the contrary, substitution of GAACA by AGGTG (oligonucleotide Bc4) did not affect the binding. Computer analysis of Bc with the Oligo 4.0 software defines a free energy value of −3.5 kcal/mol for the hairpin structure corresponding to the palindromic sequences shown in Fig. 1A.

3.2. Repression of 1731 in vitro transcription via the NssBF element

The role of the NssBF element and its related binding proteins in in vitro transcription through the 1731 promoter was analysed using a *Drosophila melanogaster* embryonic nuclear extract (Fig. 2). This extract was depleted for NssBF element binding proteins by incubation with an oligonucleotide containing this sequence (Bc) (400ng template/10–50–100 molar excess oligonucleotide) for 20 min at room temperature before in vitro transcription assay. This procedure increased 1731 specific transcripts suggesting that single-stranded DNA binding activities(y) specific to Bc represse(s) transcription through the 1731 promoter. As a control, the extract was preincubated with the mutated oligonucleotide Bc3. In this last case, no quantitative variation of 1731 transcripts was observed.

3.3. Sequence similarities between known nucleic signals and the NssBF element

Similarities between the NssBF element and nucleic signal databases (GNOMIC, TFD) were examined using the BISANCE programs [19]. It reveals similitudes with *Drosophila melanogaster* topoisomerase II cleavage sites [20], autonomous replicating sequences (ARS) of Yeast and *Drosophila* [21–23] and procaryotic origins of replication [24,25] (Fig. 3A). DNA topoisomerase II mediated-cleavage sites in this region of the

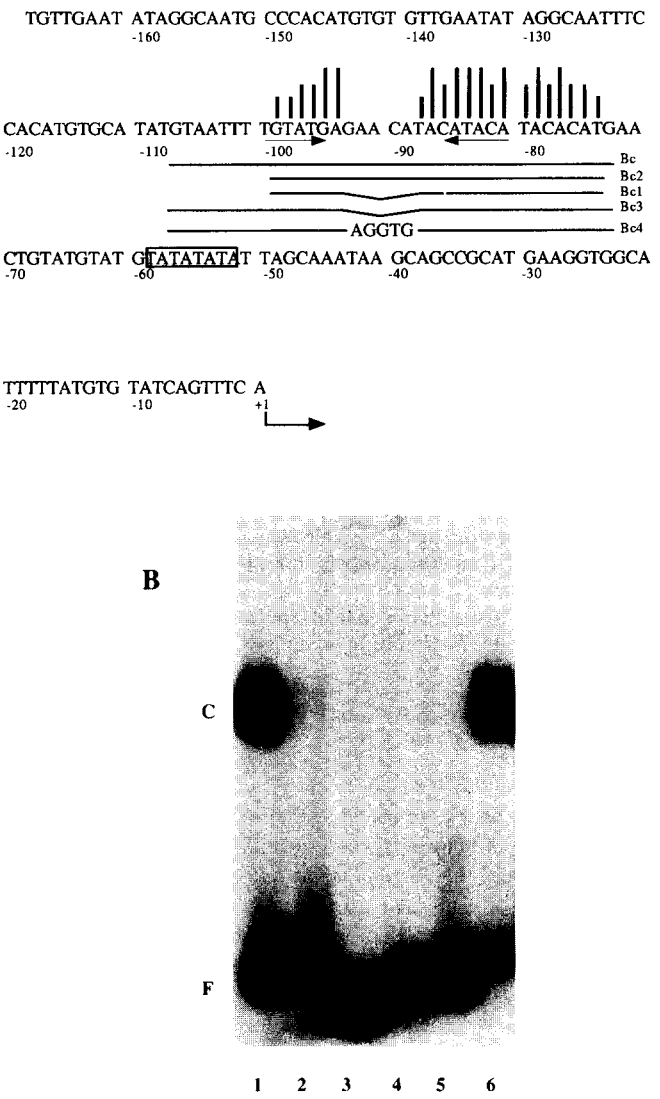

Fig. 1. Characterization of the NssBF element. (A) Nucleotide sequence of the 1731 5′ LTR U3 region [7,10]. Numbers refer to the nucleotide position relative to the transcription initiation site (+1). The TATA-box is framed. Vertical lines represent the points of contact between NssBF and Bc as described in [10]. Horizontal arrows emphasize a palindromic sequence. Sequence of the Bc and mutated oligodeoxynucleotides used in gel shift assays are shown. (B) Gel retardation assay examining the 0.5 M KCl heparin-agarose fraction with the wild-type Bc probe (line 1), the complementary probe Bnc (line 2), the shortened probe Bc2 (line 4), and the mutated probes Bc1 (line 3), Bc3 (line 5) and Bc4 (line 6). C: NssBF/DNA complexes; F: free probes.

1731 long terminal repeat was indeed previously reported by Nahon et al. [26]. Moreover, we found similarities with a consensus sequence proposed by Dobbs et al. for a potential initiation regions of eukaryotic chromosomal DNA [27] (Fig. 3B).

4. Discussion

The U3 part of retrotransposons LTR contains signals controlling the expression of both *gag* and *pol* genes. Among these

signals, the NssBF element occurs at position −99 with regard to the transcription start site of the 1731 retrotransposon (+1). Data concerning the primary structure and the role of this element are reported in the present paper.

Gel retardation assays using shortened or mutated oligonucleotides show that the NssBF binding site previously determined by chemical footprint is not sufficient by itself to ensure binding of NssBF: the presence of surrounding sequences is also required. Moreover, it allows to define the consensus sequence GTATGARRRYRTACATACATACACAT for the NssBF element where R is A or G and Y is C or T. Lastly, it suggests that the distance between the two binding domains more than the sequence is required to maintain the binding. The mechanism by which the protein binds to DNA might be connected to the presence of an hairpin structure in this region.

In vitro transcription experiments suggest a down-regulation of the 1731 promoter via single-stranded DNA binding proteins that exhibit specific affinity for the NssBF element. Furthermore, sequence comparisons lead to speculate about a possible role of the element in the initiation of replication. Sequences which may provide signals for initiation of replication are indeed characterised by the presence of a DNA unwinding element (DUE) and a cluster of modular elements as, for instance, SAR (Scaffold Associated Sequence) and/or ARS (Autonomous Replicating Sequence) and/or DNA topoisomerase II binding site [27]. Such modular elements were found in the NssBF element and its surrounding sequences. In addition, the

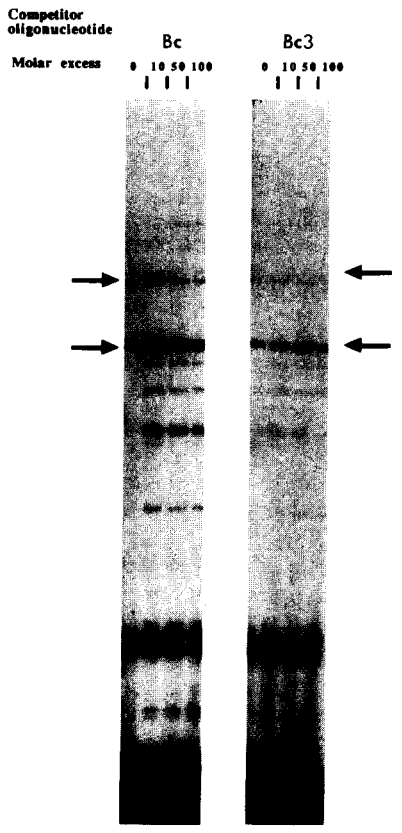

Fig. 2. In vitro transcription experiments: In vitro transcription reactions were performed as described in section 2. The arrows point out the products of primer extension corresponding to the major 1731 transcripts. 0 to 100 molar excess of oligodeoxynucleotides Bc or Bc3 were added to the reaction mixture as indicated.

A

Drosophila topoisomerase II cleavage site consensus sequence [20] GTNWAYATTNATNNG

1731 -91 ACATACATACATACA

1731 -95 GAGAACATACATACA

Yeast and *Drosophila* ARS consensus sequence [21,22,23] WAAAYATAAAW

1731 -94 AGAACATACAT

1731 -92 AACATACATAC

1731 -82 CATACACATGA

1731 -103 (complementary strand) AAAACATACTC

1731 -110 (complementary strand) ATACATTAAAA

Prokaryotic origins of replication

E. coli [24] TATACACA

1731 -82 CATACACA

ori c region of the *B. subtilis* chromosome [25] TTATMACA

1731 -83 ACATAACA

B

Consensus sequence for potential initiation region in Eukaryotic chromosomes [27] WAWTTDDWWWDMWGWMMAWTT

1731 -105 AATTTTGTATGAGAACATACA

1731 -106 TAATTTTGTATGAGAACATAC

Fig. 3. Similarities between the 1731 promoter and replication sequence elements. (A) *Drosophila* topoisomerase II cleavage site, Yeast and *Drosophila* ARS, Prokaryotic origins of replication. (B) Potential initiation region in eukaryotic chromosomes. Numbers indicate the position of the alignment in the 1731 promoter. Y = C + T; M = A + C; W = A + T; D = G + A + T; N = G + A + T + C.

U3 part of the 1731 LTR exhibits a high A+T content (63% against 54% in the whole element); the relative lability of the double helix in this region might then facilitate the access of the replication machinery. We propose that NssBF and others NssBF element binding proteins stimulate unwinding of DNA.

The question arises of the functional relationship between repression of transcription and replication. The conspicuous repression of 1731 transcription via NssBF element binding proteins might only reflect a competition between the transcription and the replication machineries. Initiation of DNA replication occurs at multiple specific sites in the genome [28]. Nevertheless, considering the existence of a replication origin in a transposon LTR, it appears that each new insertion of the element will introduce a new origin of replication in the genome. Moreover, the maintenance of extrachromosomic DNA circles corresponding to retrotransposition intermediates might be assure by these replication origins. Proteins that bind these replication origins might then be implicated in the regulation of retrotransposition frequency. We thus project to construct appropriate vectors to assess replication efficiency of this sequence.

Acknowledgements: The authors wish to thank Dr. Martin Best-Belpomme, Dr. Alain Debec and Dr. Yves Maurin for helpful discussions and critical reading of the manuscript. This work was supported by grants for Centre National de la Recherche Scientifique.

References

[1] Echalier, G. (1989) Adv. Virus Res. 36, 33–105.
[2] Grandbastien, M.A. (1992) Trends Genet. 8, 103–108.
[3] Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985) Cell 40, 491–500.
[4] Heidmann, O. and Heidmann, T. (1991) Cell 64, 159–170.
[5] Yoshioka, K., Kanda, H., Akiba, H., Enoki, M. and Shiba, T. (1991) Gene 103, 179–184.
[6] Strand, D.J. and McDonald, J.F. (1985) Nucleic Acids Res. 13, 4401–4411.
[7] Peronnet, F., Becker, J.L., Becker, J., d'Auriol, L., Galibert, F. and Best-Belpomme, M. (1986) Nucleic Acids Res. 14, 9017–9033.
[8] Ziarczyk, P., Fourcade-Peronnet, F., Simonart, S., Maisonhaute, C. and Best-Belpomme, M. (1989) Nucleic Acids Res. 16, 6113–6125.
[9] Ziarczyk, P. and Best-Belpomme, M. (1991) Nucleic Acids Res. 19, 5689–5693.
[10] Fourcade-Peronnet, F., Codani-Simonart, S. and Best-Belpomme, M. (1992) J. Virol. 66, 1682–1687.
[11] Codani-Simonart, S. (1993) Thèse, Université P. et M. Curie, Paris, France.
[12] Codani-Simonart, S., Best-Belpomme, M. and Fourcade-Peronnet, F. (1993) FEBS Lett. 3, 177–182.
[13] Chase, J.W. and Williams, K.R. (1986) Annu. Rev. Biochem. 55, 103–136.
[14] Kamada, S. and Miwa, T. (1992) Gene 119, 229–236.
[15] Marton, R.F., Thömmes, P. and Cotterill, S. (1994) FEBS Lett. 342, 139–144.
[16] Blackwell, L.J. and Borowiec, J.A. (1994) Mol. Cell. Biol. 14, 3993–4001.
[17] Echalier, G. and Ohanessian, A. (1969) C.R. Acad. Sci. 268, 1771–1773.
[18] Close, T.J. and Rodriguez, R. (1982) Gene 20, 305–316.
[19] Dessen, P., Fondrat, C., Valencien, C. and Mugnier, C. (1990) Cabios 6, 355–356.
[20] Sander, M. and Hsieh, T. (1985) Nucleic Acids Res. 13, 1057–1072.
[21] Montiel, J.F., Norbury, C.J., Tuite, M.F., Dobson, M.J., Mills, J.S., Kingsman, A.J. and Kingsman, S.M. (1984) Nucleic Acids Res. 12, 1049–1068.
[22] Marunouchi, T. and Hosoya, H. (1984) Mol. Gen. Genet. 196, 258–265.
[23] Broach, J.R., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A. and Hicks, J.B. (1983) Cold Spring Harbour Symp. Quant. Biol. 47, 1165–1173.
[24] Matsui, M., Oka, A., Takanami, M., Yasuda, S. and Hirota, Y. (1985) J. Mol. Biol 184, 529–533.
[25] Moriya, S., Ogasawara, N. and Yoshikawa, H. (1985) Nucleic Acids Res. 13, 2251–2265.
[26] Nahon, E., Best-Belpomme, M. and Saucier, J.M. (1993) Eur. J. Biochem. 218, 95–102.
[27] Dobbs, D.L., Shaiu, W.L. and Benbow, R.M. (1994) Nucleic Acids Res. 22, 2479–2489.
[28] DePamphilis, M.L. (1993) Annu. Rev. Biochem. 62, 29–63.