Novel DNA-binding protein from *Drosophila* embryos identified by binding site selection

Vesna Todorović[a], Snežana Kojić[a], Miloš Vujanac[b], Ana Savić[b], Dragana Stefanović[a,*]

[a]*Institute of Molecular Genetics and Genetic Engineering, Vojvode Stepe 283, P.O. Box 794, 11000 Belgrade, Yugoslavia*
[b]*Department of Biology, University of Belgrade, Belgrade, Yugoslavia*

Received 10 September 1996

\bstract The DNA sequence C/AGAGCGC/AGA, related to binding sites for GAF and Zeste transcription factors, was selected from a pool of degenerate PCR fragments for binding to the cytoplasmic protein of *Drosophila* preblastoderm embryos. Identical DNA binding activity was also detected in embryonic nuclei. Based on several criteria, such as size, intracellular distribution, sensitivity to ATP and protein kinase inhibitor 6-DMAP, kinetics during development and lack of cross-reaction with rabbit anti-GAF serum, protein recognizing selected sequence was shown to differ from either Zeste or GAF.

Key words: Embryonic protein; DNA binding site selection; *Drosophila melanogaster*

. Introduction

During preblastoderm development cells of *Drosophila* embryos do not divide, while their nuclei undergo a series of replications and divisions. Nuclear changes are executed by cell cycle driven machinery, which keeps embryonic chromatin in a state of extreme replicative activity and transcriptional silence [1,2]. To make an inventory of DNA binding activities that potentially influence chromatin behavior, we undertook cyclic selection and amplification of binding sites recognized by proteins of *Drosophila melanogaster* preblastoderm embryos. Proteins of interest were expected to capture appropriate DNA from a pool of PCR fragments containing totally degenerate core and to eventually appear as visible complexes after several rounds of selection and amplification of DNA isolated from different sections of the first mobility shift gel.

Here, we describe DNA binding activity specific for the C/AGAGCGC/AGA core sequence, determined after repeated cycles of selection and amplification of DNA isolated from a single protein-DNA complex. The selected sequence resembles binding sites for two known *Drosophila* transcription factors, but the selected protein clearly differs from them. It appears to be novel and predominantly required very early in development.

2. Materials and methods

2.1. Preparation of cytoplasmic and nuclear extracts

Drosophila embryos were collected 0–90 min or 0–17 h after egg laying, dechorionated with hypochloric acid and washed with water. Cytoplasmic extracts were prepared as in [3], except that ATP was omitted and 8 mM 6-dimethylaminopurine (6-DMAP) included in the homogenization buffer, where indicated. Nuclear extracts were prepared by a modification of Dignam's procedure [4].

2.2. Preparation of DNA fragments

Single-stranded oligonucleotide GAATTCGAGCTCGGATCC-$(N)_{16}$TCTAGACTGCAGAAGCTT was converted into a double-stranded form with *taq* polymerase and AAGCTTCTGCAGTCTAGA primer, labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. Double-stranded oligonucleotide was purified by native PAGE and used in the first round of selection, essentially as described in [5,6]. Double-stranded oligonucleotides, isolated from mobility shift gels at different stages of selection, were PCR amplified with GAATTCGAGCTCGGATCC and AAGCTTCTGCAGTCTAGA primers, end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase and purified by native PAGE. Selected PCR fragments were sequenced according to [5]. Sequencing was performed on both strands, using end-labeled PCR primers. The single-stranded oligonucleotide GAATTCGAGCTCGGATCCTTTTTTATTATTATTATTATTAGTCTAGACTGCAGAAGCTT was amplified with the same set of PCR primers as the degenerate oligonucleotide and purified by native PAGE. A double-stranded oligonucleotide with alternating A and T residues was obtained by renaturation of a single-stranded oligonucleotide that contained 15 ATT repeats with its complementary strand. A double-stranded oligonucleotide with alternating G and A residues, obtained by renaturation of AAGAGAAAGAGAGAAGAG and GGGCTCTTCTCTCTTTCTCTT single strands, was labeled with [α-^{32}P]dCTP, using *taq* polymerase and purified by native PAGE.

2.3. DNA binding reactions

DNA binding was performed at 23°C in a total volume of 20 µl, with 10 µg of either nuclear or cytoplasmic extract. In the first round of selection, reaction mixtures were supplied with 2.5 ng of the labeled degenerate PCR fragment (containing approximately 10 copies of each PCR fragment with an identical 16 base core) and with poly-[(dI-dC)] in 2000–5000-fold weight excess. At later stages of selection, the amount of labeled DNA was decreased to 0.3–0.5 ng. In addition, reaction mixtures contained 20 mM HEPES, pH 7.9, 200 mM KCl, 5% glycerol, 8 mM $MgCl_2$, 0.05% Triton X-100 and 1 mM DTT (binding buffer). The mixtures were made 10 mM with respect to ATP, or 8 mM with respect to 6-DMAP, or supplied with unlabeled specific competitors in 80–100-fold molar excess, where indicated. Gel mobility shift assays were performed on 4% native polyacrylamide gels in Tris-borate buffer.

2.4. Glycerol gradient sedimentation

Cytoplasmic and nuclear extracts were clarified by centrifugation at $100\,000\times g$ and $15\,000\times g$, respectively, without noticeable loss of DNA binding activity. Corresponding supernatants were applied on 20–50% glycerol gradients prepared in binding buffer, with or without 8 mM 6-DMAP. Gradients were centrifuged for 20 h in a SW 50.1 Beckman rotor at 49 000 rpm and 4°C, fractionated from the bottom, and assayed for DNA binding. Molecular weight markers used in glycerol gradient sedimentation were: catalase from bovine liver (240 kDa), phosphorylase b (97 kDa), bovine serum albumin (67 kDa) and ovalbumin (43 kDa).

2.5. Western blots

Nuclear and cytoplasmic proteins were separated in 10% SDS-polyacrylamide gels and transferred to nitrocellulose membranes. Western blotting was performed with 2000-fold diluted rabbit serum raised against purified GAF and visualized using anti-rabbit IgG alkaline phosphatase conjugate with BCIP and NBT colored substrates.

*Corresponding author. Fax: (381) (11) 492-397.

PII S0014-5793(96)01078-2

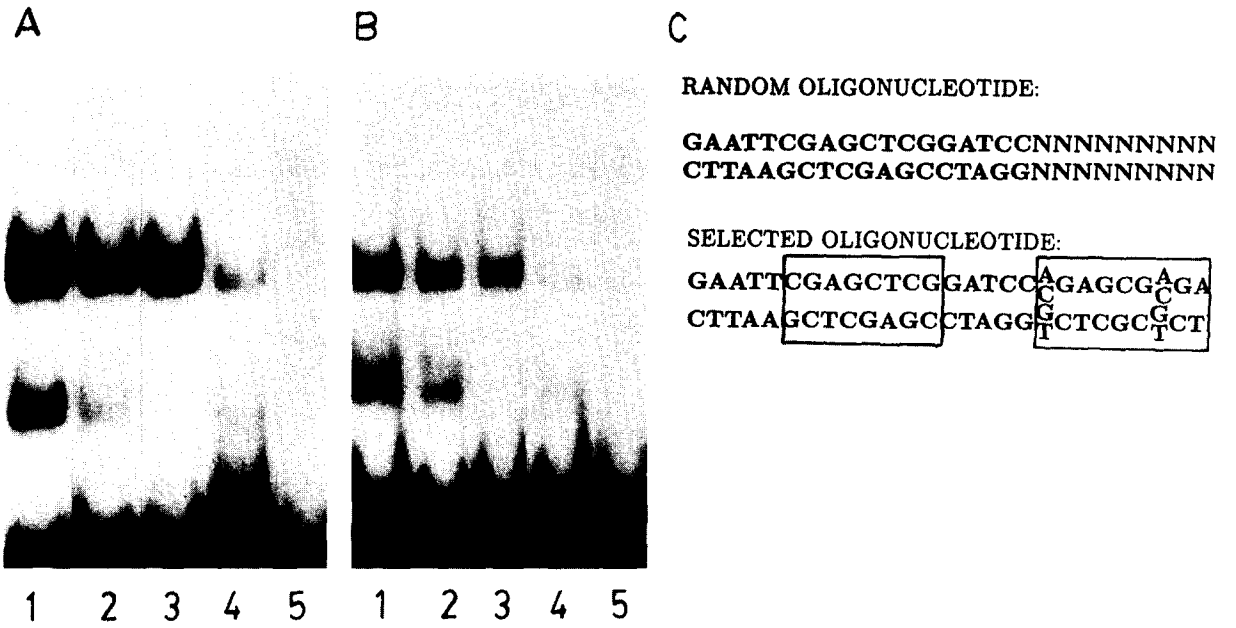

Fig. 1. DNA binding of cytoplasmic (A) and nuclear (B) embryonic proteins, without specific competitors (A and B, lane 1) or with: A+T rich PCR fragment (A and B, lane 2), A+T rich ds oligonucleotide (A and B, lane 3) and selected DNA (A and B, lane 4). Lane 5 in (A) and (B) is the reaction mixture without proteins. C: DNA binding motifs within the 5′-primer and core regions.

3. Results and discussion

Initial binding experiments were performed with a PCR fragment containing a 16 base totally degenerate core and cytoplasmic extract isolated from *Drosophila* preblastoderm embryos. Reaction mixtures were supplied with unspecific competitor in 2000–5000-fold weight excess and with ATP. Several protein-DNA complexes were detected in the first round of selection (not shown) and two of them could tolerate up to 5000-fold weight excess of unspecific competitor (Fig. 1A, lane 1). Those two complexes could recognize binding sites contained within primer regions or composed from parts of the primer and core regions, but not the individual core sequences, which were greatly under-represented in the degenerate DNA pool. Possible exceptions were short, monotonous sequences, such as AT tracts, relatively abundant in the degenerate pool. Since we already knew that embryonic extracts contained a protein recognizing A+T rich DNA (not shown), we applied two competitors: a double-stranded oligonucleotide composed of alternating A and T residues and a PCR fragment with primer sites identical to the degenerate oligonucleotide and with an A+T rich core sequence. Both competitors affected only one complex (Fig. 1A, lanes 1–3). The other was chosen for further selection. Its DNA was isolated, amplified and tested as competitor in a binding reaction performed with cytoplasmic extract and degenerate oligonucleotide. It displaced both bands (Fig. 1A, lane 4). After two more rounds of selection, nuclear proteins were tested for binding to selected DNA and basically the same pattern as with cytoplasmic extracts was obtained (Fig. 1B).

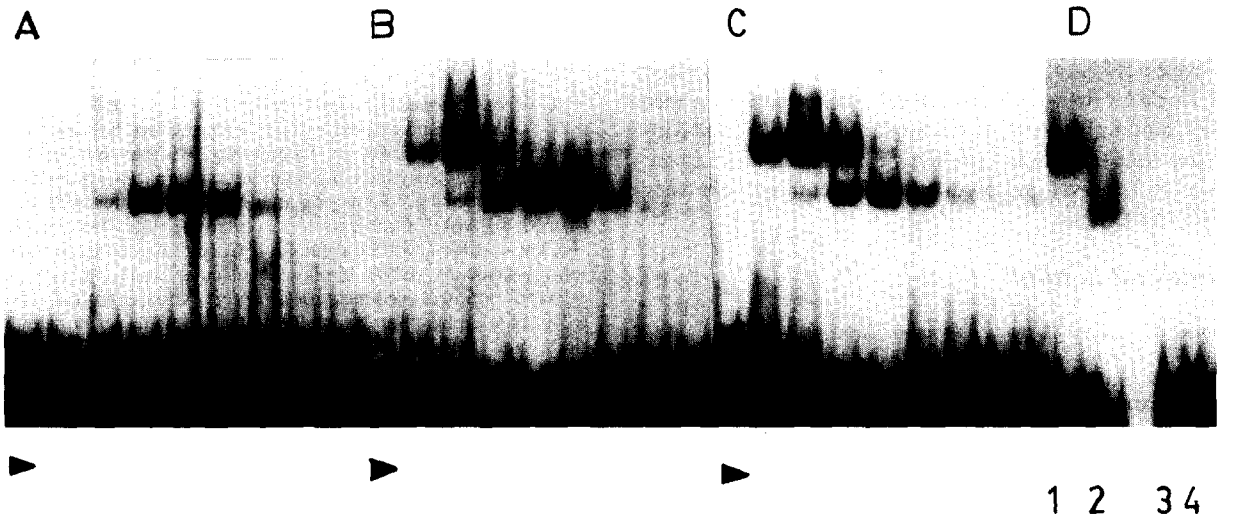

Fig. 2. Glycerol gradient sedimentation of cytoplasmic proteins, with (A) or without ATP (B, C) and with 6-DMAP (C). D: DNA binding of the 240 kDa (lane 1) and 67–100 kDa (lane 2) proteins; specific competition with selected DNA (lanes 3 and 4).

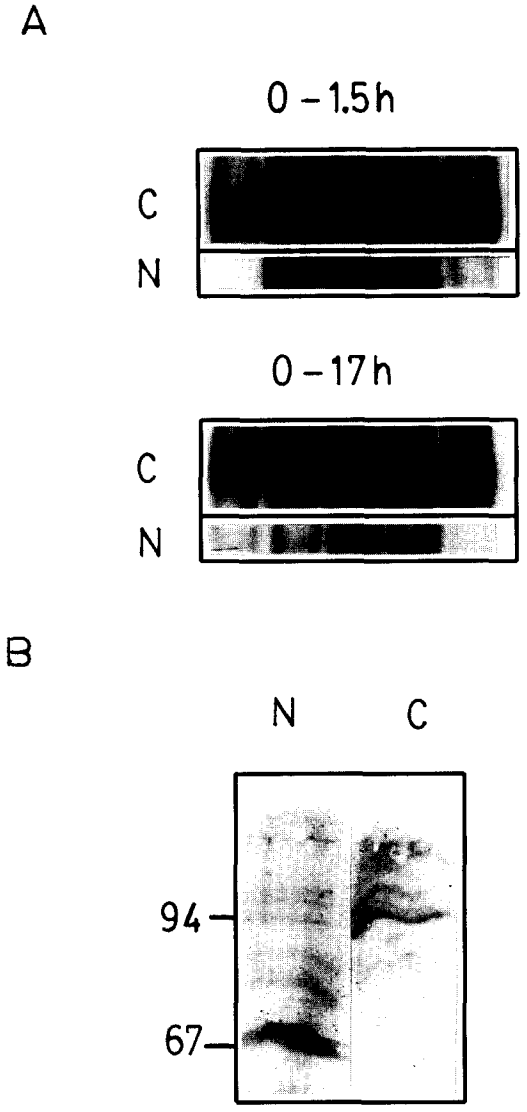

Fig. 3. A: 67–100 kDa DNA binding activity in extracts derived from embryos at different stages of development (C, cytoplasm; N, nuclei). B: Western blotting with anti-GAF rabbit serum.

As judged by competition experiments, one of the proteins detected exhibited dual binding specificity: for the selected PCR fragment and for A+T rich DNA. In that respect it was similar to several divergent *Drosophila* homeo box proteins shown to recognize both the G+C rich and the A+T rich class of related binding sites [7,8]. The other protein was specific for selected DNA only. Its binding site was determined by sequencing of DNA isolated after the third round of selection and confirmed after each of three additional rounds. The core sequence C/AGAGCGC/AGA was obtained in each sequencing reaction. Its position at the 5′-primer/core border and the occurrence of a closely matching motif within the 5′-flanking region (Fig. 1C) suggested the participation of the primer site in binding and possible recognition of the whole site by a dimeric binding unit.

Based on computer assisted database search, two *Drosophila* proteins were found with binding sites similar to the selected core sequence. Those proteins were transcription factors Zeste and GAF, which recognized C/TGAGC/T and C/AGAGAGAGC, respective canonical binding sites. To determine possible similarity of the selected protein to either Zeste or GAF, additional experiments were undertaken. In the binding reactions that follow, protein exhibiting dual DNA binding specificity was selectively removed by A+T rich DNA.

Protein of interest was present both in cytoplasm and in nuclei, but its cytoplasmic store (a rough estimation gives 10^6 copies per embryo, with more than nine tenths in the cytoplasm) greatly exceeded the amount present in nuclei. With cytoplasmic extracts prepared as in [3], separated by glycerol gradient sedimentation and assayed in the presence of ATP, it migrated as a single peak of 67–100 kDa (Fig. 2A). Without ATP, an additional form, of around 240 kDa, was detectable (Fig. 2B). With extracts prepared by a modified procedure, modifications being the exclusion of ATP from the homogenization buffer and the addition of the protein kinase inhibitor 6-DMAP, DNA binding activity assayed in the presence of ATP also migrated as a single peak (not shown). When assayed without ATP, two forms appeared again (Fig. 2C). A comparison of glycerol gradient profiles of extracts prepared without (Fig. 2B) and with 6-DMAP (Fig. 2C) and assayed without ATP revealed a redistribution which occurred between the two forms. In the presence of protein kinase inhibitor, the amount of the 240 kDa form increased, while the 67–100 kDa form decreased. This indicates that, directly or indirectly, phosphorylation regulated the level of each form. In addition, ATP affected the binding or integrity of the 240 kDa form in a manner independent of 6-DMAP. Regardless of the presence of ATP or protein kinase inhibitor, only the 67–100 kDa form was detected in nuclei (not shown). It should be stressed that, like the 67–100 kDa form, the 240 kDa form specifically recognized selected DNA (Fig. 2D).

Zeste and GAF proteins are predominantly located in embryonic nuclei [9–12]. The respective molecular weights of their monomers are 97 kDa and 67 kDa [9,10]. Both multimerize and recognize repeated binding sites [13–16]. The extent of multimerization of Zeste is extremely high. Multimerization is promoted by the C-terminus, correlated with biological activity and required for efficient DNA binding [13,14]. Zeste protein, isolated from fly nuclei by three different methods, forms aggregates, which predominantly sediment in the $100\,000 \times g$ pellet [13]. As shown by gel filtration, even in the supernatant, wild-type protein is present in forms ranging in apparent molecular weight from several million kDa down to the monomer [13]. Due to extreme multimerization, only the protein expressed in vitro or synthesized in *E. coli* fused with β-galactosidase can be assayed for DNA binding activity by gel mobility shift [14,17].

A comparison of glycerol gradient profiles of cytoplasmic and nuclear extracts isolated from approximately the same number of embryos, collected for either 0–90 min or 0–17 h and assayed for DNA binding, demonstrated no significant change in the overall cytoplasmic 67–100 kDa DNA binding activity and a moderate decrease in nuclear binding (Fig. 3A). Since the number of nuclei significantly increased during the first 17 h of development, the binding activity per single nucleus decreased much more than that overall. In contrast, a large increase in GAF and Zeste binding to DNA elements within the *Ultrabitorax* promoter was demonstrated during the first 16 h of development and was required for increased *Ubh* transcription in nuclear embryonic extracts [9].

The size of the selected protein and its DNA binding kinetics during development indicated that it differed from either Zeste or GAF. However, GAF was demonstrated to exist in several forms (produced by multiple, by alternative splicing generated and developmentally regulated mRNAs), and to recognize a wide variety of binding sites, almost all containing a GAGAGAG consensus sequence [11,18]. Contrary to that, the selected protein did not recognize a 21-mer containing alternating G and A residues and was not recognized by rabbit antiserum raised against purified GAF (kindly provided by C. Wu) in gel mobility supershift assays (not shown). The same antibody recognized in Western blotting experiments characteristic GAF forms, which included a 67 kDa protein and a series of higher molecular weight polypeptides (Fig. 3B). The protein was detected in the nuclei, but not in the cytoplasm of preblastoderm embryos. Immunohistochemical staining of GAF, restricted to nuclei, was previously observed in preblastoderm embryos and in embryos at the syncytial blastoderm stage, prior to cellularization [11,12].

Taken together our results demonstrate that DNA binding activity, specifically recognizing the C/AGAGCGC/AGA sequence, belongs to a protein which differs from both Zeste and GAF. That protein is present in embryonic extracts isolated from transcriptionally silent embryos (collected even 0–15 min after egg laying, not shown), which indicates its maternal origin. The protein is stored in the cytoplasm of preblastoderm embryos and it appears in two DNA binding forms, one being exclusively cytoplasmic (240 kDa) and the other (67–100 kDa) both cytoplasmic and nuclear. ATP and protein kinase inhibitors regulate the level of both forms in a so far unknown manner. Since the nuclear binding activity decreases during development, the protein may perform a function predominantly required very early in development. Its proper function, as well as its relation to the other protein, recognizing both the selected sequence and A+T rich DNA, remains to be determined.

From the beginning, our selection procedure was directed by that part of the DNA binding site which was unintentionally included in the primer region. Because of that, all protein-DNA complexes containing core regions only were lost. However, even without participation of the primer sites in binding, separation of individual protein-DNA complexes too early in selection could result in losses. For simultaneous analysis of many different DNA binding proteins, we intend to immunoprecipitate protein-DNA complexes formed with degenerate oligonucleotide with rabbit serum raised against total proteins from *Drosophila* preblastoderm embryos and to amplify immunoprecipitated DNA. After several rounds of amplification of immunoprecipitated DNA, the DNA pool would contain only binding sites recognized by *Drosophila* embryonic proteins and would be suitable for sequencing after final separation of protein-complexed DNA by long mobility shift gel.

Acknowledgements: We are grateful to Carl Wu and Toshio Tsukiyama for providing rabbit anti-GAF serum and to Ivan Brukner, Mensur Dlakić, Sabera Ruždijić and Selma Kanazir for help and advice.

References

[1] O'Farrell, P.H., Edgar, B.A., Lakich, D. and Lehner, C.F. (1989) Science 246, 635–640.
[2] McKnight, S.L. and Miller, O.L. Jr. (1976) Cell 8, 305–3193.
[3] Crevel, G. and Cotterill, S. (1991) EMBO J. 10, 4361–4369.
[4] Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Nucleic Acids Res. 11, 1475–1480.
[5] Blackwell, T.K. and Weintraub, H. (1990) Science 250, 1104–1110.
[6] Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N. and Weintraub, H. (1990) Science 250, 1149–1151.
[7] Hoey, T. and Levine, M. (1988) Nature 332, 858–861.
[8] Hoey, T., Warrior, R., Manak, J. and Levine, M. (1988) Mol. Cell. Biol. 8, 4598–4607.
[9] Biggin, M.D. and Tjian, R. (1988) Cell 53, 699–711.
[10] Biggin, M.D., Bickel, S., Benson, M., Pirrotta, V. and Tjian, R. (1988) Cell 53, 713–722.
[11] Soeller, W.C., Euk Oh, C. and Kornberg, T.B. (1993) Mol. Cell. Biol. 13, 7961–7970.
[12] Raff, J.W., Kellum, R. and Alberts, B. (1994) EMBO J. 13, 5977–5983.
[13] Bickel, S. and Pirrota, V. (1990) EMBO J. 9, 2959–2967.
[14] Chen, J.D. and Pirrota, V. (1993) EMBO J. 12, 2075–2083.
[15] Tsukiama, T., Becker, P.B. and Wu, C. (1994) Nature 367, 525–531.
[16] Zollman, S., Godt, D., Prive, G.G., Couderc, J.-L. and Laski, F.A. (1994) Proc. Natl. Acad. Sci. USA 91, 10717–10721.
[17] Mansukhani, A., Crickmore, A., Sherwood, P.W. and Goldberg, M.L. (1988) Mol. Cell. Biol. 8, 615–623.
[18] Granok, H., Leibovitch, B.A., Shaffer, C.D. and Elgin, S.C.R. (1995) Curr. Biol. 5, 238–241.