

A *Drosophila* nuclear localisation signal included in an 18 amino acid fragment from the *serendipity* δ zinc finger protein

Stéphane Noselli and Alain Vincent

Centre de Recherche de Biochimie et Génétique Cellulaire, 118 route de Narbonne, 31062 Toulouse Cedex, France

Received 20 December 1990; revised version received 22 January 1991

Sequence analysis of the nuclear *Drosophila serendipity* δ Cys-2/His-2 finger protein indicated the presence of a short motif of positively charged amino acids, with homology to the SV40 large T and c-myc nuclear localisation signals. Using P-element mediated transformation we constructed transgenic *Drosophila* lines expressing β -galactosidase fusion proteins, containing (or not) an 18 residue segment of *sry* δ including this basic, PTKKRVK, motif. Histochemical detection of fusion proteins on dissected tissues showed that this segment of *sry* δ can act autonomously to drive the β -galactosidase in nuclei.

Nuclear localisation sequence; β -Galactosidase fusion; Finger protein; Transgenic fly; *Drosophila melanogaster*

1. INTRODUCTION

In eukaryotic cells, a subset of the proteins that are synthesized in the cytoplasm have to enter the nucleus for their function. Protein uptake into nuclei is a selective and specific two step process [1–3]. First, there is a specific targeting of nuclear proteins, to the nuclear pore complex (NPC). Second, there is an energy-dependent translocation of the proteins to the nucleus through the NPC. Nuclear proteins contain one or more short sequences, designated as nuclear localisation sequences (NLS), that are required for specific targeting to the nuclear pore complex (for review, see [4–6]). This specific targeting involves recognition by receptors, the NLS-binding proteins. A few have been described recently in yeast, rat and human cells, but their subcellular localisation, i.e. cytoplasmic or at the NPC, is not yet clear [7–10]. In past years, several NLS have been identified in various nuclear proteins of diverse origins, including viruses, yeast, *Xenopus* and human [4–6]. The best-characterised NLS is that of the SV40 large T antigen. This seven amino acid long NLS is rich in basic residues and can function autonomously when fused to cytoplasmic proteins in mammalian cells and *Xenopus* embryos [11,12]. This NLS was shown to be also functional in yeast, although not identified yet in any yeast protein [13]. While numerous genes coding for nuclear proteins have been characterised in *Drosophila*, no NLS has yet been precisely identified in any Dipteran.

The *Drosophila sry* δ zinc finger protein is a nuclear protein found in transcriptionally active cells [14]. It binds to specific DNA sequences in vitro, and to specific sites on polytene chromosomes of third instar larvae [15]. In this report, we used P-element mediated transformation to show that an 18 amino acid segment from the *sry* δ protein contains a nuclear localisation sequence able to direct the *E. coli* β -galactosidase into the nucleus in all *Drosophila* tissues examined. Included in this 18 amino acid stretch is the sequence PTKKRVK, similar to the SV40 large T and human c-myc NLS, suggesting that this motif could be a *Drosophila* nuclear localisation signal.

2. MATERIALS AND METHODS

2.1. Plasmid constructions

Nucleotides and amino acid numbers refer to sequences published in Vincent et al. [16].

2.1.1. pSDC construct

An EcoRI-Sau3A1-klenow DNA fragment containing part of the *sry* δ gene, starting 4138 bp upstream of the transcription start site and coding for the first ten N-terminal amino acid of the protein, was subcloned in pTZ18R [17] cut by EcoRI-SmaI, giving pE-Sau plasmid. Then, a XbaI (2695 bp upstream of the transcription start site of *sry* δ) Sall fragment from pE-Sau was fused in frame to the amino terminus of the *E. coli* β -galactosidase, in the P-element vector pSDL [14] cut by XbaI and Sall, to give the plasmid pSDC. pSDC encodes a chimaeric protein containing the *sry* δ residues 1–10 fused to the β -galactosidase (Fig. 1).

2.1.2. pSDC construct

A SphI site was created at nucleotide position 4783 of the *sry* δ sequence by site-directed mutagenesis [18], substituting Met-Pro to Lys-Ser residues 176 and 177. Then, a *sry* δ fragment bordered by the created SphI and an ApaLI site at position 4831 and coding for amino acids 182–197 was inserted into pSDC, between the *sry* δ and β -

Correspondence address: S. Noselli, Centre de Recherche de Biochimie et Génétique Cellulaire, 118 route de Narbonne, 31062 Toulouse Cedex, France

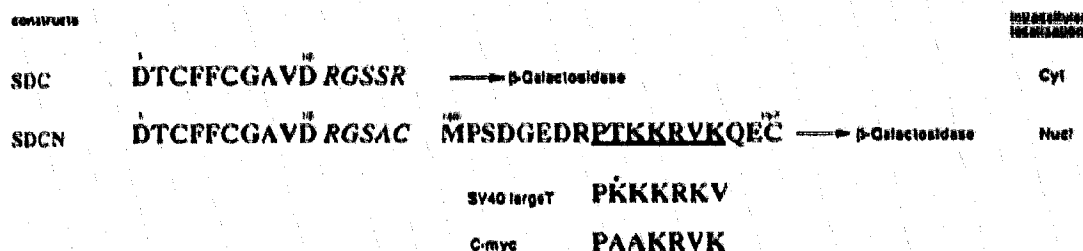


Fig. 1. *Sry* δ sequences fused to the β -galactosidase in SDC and SDCN constructs. Numbering refers to the *sry* δ protein sequence [15]. Italics represent additional residues encoded by polylinker sequences. Underlined *sry* δ amino acids in SDCN, point out homology with the SV40 T [11] and *c-myc* [22] nuclearisation signals. (*) points to a Lys residue in SV40 T NLS, that when mutated to Thr, does not affect the nuclear localisation [27].

galactosidase sequences using polylinkers. The resulting plasmid pSDCN codes for a tripartite protein containing the *sry* δ residues 1–10 and 180–197 fused to β -galactosidase (Fig. 1).

2.2. Transgenic *Drosophila* lines

SDC and SDCN plasmids were used to transform *ry*⁵⁰⁶ flies, according to standard protocols [19], except that DNA was injected in non-dechorionated embryos [20]. Transformant lines were made homozygous, and chromosomes carrying the individual P-element inserts determined by chromosomal linkage, using balancer stocks.

2.3. Western blot analysis

Ovaries dissected from young flies kept for two days on medium supplemented with fresh yeast at 25°C, were rapidly frozen in dry-ice. To prepare extracts, ovaries were homogenized in OV buffer (Hepes 25 mM, pH 7.5, NaCl 50 mM, EDTA 1 mM, EGTA 0.1 mM, DTT 1 mM, 100 μ g/ml PMSF, 2 μ g/ml aprotinin, 2 μ g/ml Leupeptin) before centrifugation at 4°C for 10 min. The supernatant was aliquoted and frozen at –80°C. Protein concentration was determined by Bradford assay (Bio-Rad). 100 μ g were subjected to 8% SDS-Polyacrylamide gel electrophoresis. Electrophoretic transfer, antibody reactions and signal detection were done as described in [14].

2.4. X-Gal staining

Histological detection of the β -galactosidase enzymatic activity on whole mount salivary glands and other tissues was done by X-Gal staining according to [21].

2.5. Photographs

Photographs were taken with Nikon optics and microscope, using Ilford FP4 film.

3. RESULTS AND DISCUSSION

We previously reported that a *sry* δ / β -galactosidase fusion protein containing most of the *sry* δ coding sequence, was localised in the nucleus [14]. More recently, other fusions were made, using different domains of the *sry* δ protein, which showed that the region between residues 1–195, i.e. excluding the finger domain, contained a nuclear targeting signal (Noselli, unpublished). The predicted *sry* δ protein sequence shows the presence of a short stretch of basic amino acids (amino acids 188–194) (Fig. 1) located just upstream from the finger domain, resembling the sequence of the SV40 large T antigen NLS [11]. Sequence analysis of the *sry* δ gene from *Drosophila pseudo-obscura* revealed that this short stretch is conserved in sequence, and at the same

position in the predicted *D. pseudo-obscura sry* δ protein (M. Crozatier, personal communication). In order to test whether this sequence could act as a nuclear localisation signal, we constructed two N-terminal fusion proteins with the *E. coli* β -galactosidase, differing only by an 18 residue long segment of *sry* δ (amino acids 180–197), and present in one fusion (SDCN) but not the other (SDC) (Fig. 1). Both SDC and SDCN fusion genes are placed under control of the *sry* δ promoter, previously characterised as active during oogenesis and in transcriptionally active cells throughout the fly life cycle [14]. SDC and SDCN transformed lines were ob-

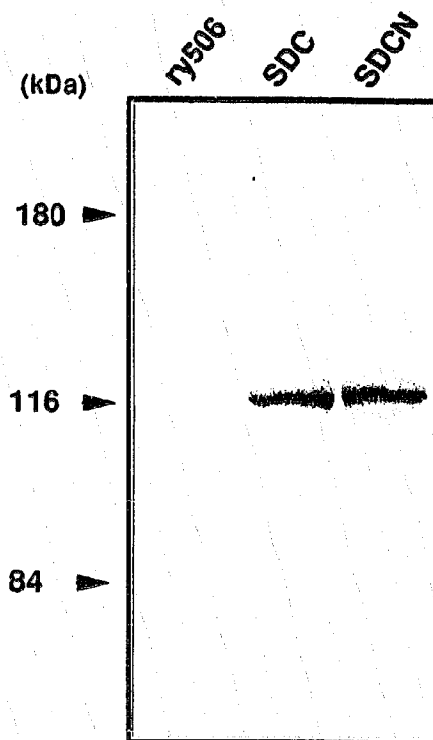


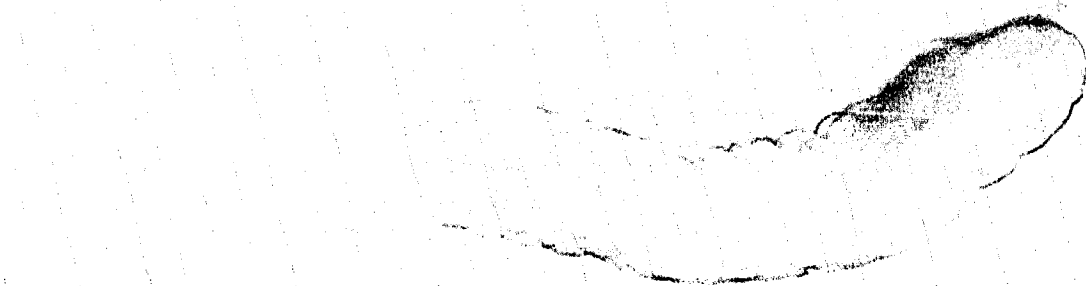
Fig. 2. Western blot analysis of protein extracts from ovaries of *ry*⁵⁰⁶, SDC and SDCN *Drosophila* lines, probed with a monoclonal anti- β -galactosidase antibody. The left column indicates the mobility of known M_r protein standards.

tained by injection into *ry⁵⁰⁶* embryos, and homozygous lines were established as described in section 2.

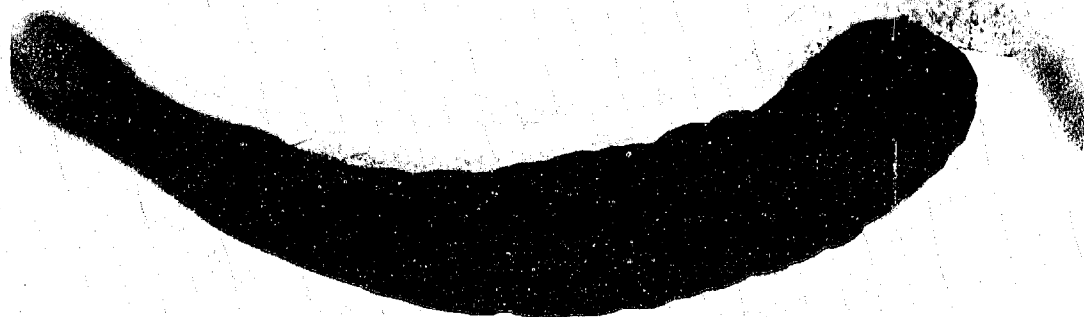
Production of the fusion proteins was verified by Western blot analysis of protein extracts from dissected ovaries, using an anti- β -galactosidase monoclonal antibody (Fig. 2). Ovaries were chosen because the wild-type *sry* δ protein accumulates in mature oocytes [14]. The levels of SDC and SDCN protein accumulation are comparable. The SDC and SDCN proteins show a relative molecular mass of about 120 kDa, SDCN migrating slightly slower (Fig. 2). This correlates well with the expected sizes for these proteins, of 120.5 kDa and 123 kDa, respectively.

Retention of the enzymatic activity by the SDC and SDCN β -galactosidase fusion proteins allows a histochemical detection of these proteins by X-Gal staining of dissected tissues. Coloration of salivary glands of third instar larvae from SDC and SDCN lines is shown in Fig. 3. In an SDC transformant line X-Gal staining is cytoplasmic, while it is strictly nuclear in salivary glands and fat body in an SDCN transformant. The control *ry⁵⁰⁶* line does not show any staining. Expression of the SDC and SDCN proteins in the fat body is very weak as compared to that in giant salivary glands cells, making cytoplasmic staining in fat body of SDC lines, at best barely detectable (Fig. 3). In addition to

ry506



SDC



SDCN

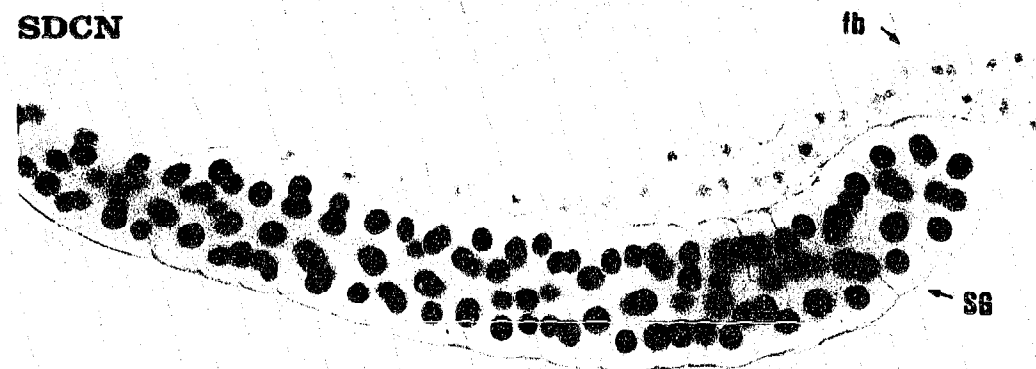


Fig. 3. Whole mount X-Gal staining of third instar larvae salivary glands of *ry506*, SDC and SDCN lines. fb = fat body; SG = salivary gland.

the salivary glands of the third instar stage, we looked at the subcellular localisation of SDC and SDCN fusions in embryos, first and second instar larvae and in adults. In all tissues where staining was detected, and apart from mature oocytes where both proteins, like the wild-type *sry* δ protein, accumulate in the ooplasm [14], the SDC protein was cytoplasmic and the SDCN nuclear (data not shown).

These results show that the extra segment of *sry* δ protein present in SDCN compared to SDC is capable of targeting the *E. coli* β -galactosidase to the nucleus. This 18 amino acid segment includes the short basic motif PTKKRVK showing strong homology to the SV40 large T and human *c-myc* NLS [11,22] (Fig. 1). It suggests that this motif could represent a *Drosophila* NLS.

So far, NLS have not been identified in *Drosophila*. One NLS is probably present in the 125 N-terminal domain of the P-element transposase but has not been delineated further [23,24]. There may be a nuclear targeting signal in the carboxy-terminal part of the *Antp* homeodomain [25], but again this signal has not been precisely mapped. In *sry* δ , the NLS is located immediately upstream to the DNA binding domain, a situation also suggested for the human zinc finger ZFX protein [26].

Our results do not exclude the existence of multiple NLS in the *sry* δ protein, as shown for other nuclear proteins [4]. Indeed, the nuclear zinc finger protein *sry* β , whose DNA binding domain is closely related to that of *sry* δ [15], does not contain a region homologous to the 18 amino acid segment of the *sry* δ protein studied here. A *sry* β / β -galactosidase fusion protein is also nuclear [15], suggesting that *sry* β and *sry* δ , although probably resulting from a gene duplication event, may have developed separate strategies for their active uptake by nuclei.

Homology to the SV40 and *c-myc* NLS makes the PTKKRVK sequence a putative *Drosophila* NLS. The evolutionary conservation of this sequence suggests in turn that at least some mechanisms and protein components involved in selective transport of proteins across the nuclear membrane could be conserved between *Drosophila* and higher vertebrates.

Identification of a NLS in the *sry* δ protein will allow its use for nuclear targeting of recombinant markers proteins in *Drosophila*, and opens the field of *Drosophila* genetics to study the mechanisms of active protein nuclear uptake.

Acknowledgements: We are grateful to D. Cribbs and M. Crozatier for critically reviewing the manuscript. This work was supported by grants from CNRS and Association pour la Recherche sur le Cancer (ARC), Grant 6457. S.N. was supported by a fellowship from the Ministère de la Recherche et de la Technologie.

REFERENCES

- [1] Newmeyer, D.D. and Forbes, D.J. (1988) *Cell* 52, 641-653.
- [2] Richardson, W.D., Mills, A.D., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1988) *Cell* 52, 655-664.
- [3] Hall, M.N., Craik, C. and Hiaroka, Y. (1990) *Proc. Natl. Acad. Sci. USA* 87, 6954-6958.
- [4] Newport, J.W. and Forbes, D.J. (1987) *Annu. Rev. Biochem.* 56, 535-565.
- [5] Gerace, L. and Burke, B. (1988) *Annu. Rev. Cell Biol.* 4, 335-374.
- [6] Wagner, P., Kunz, J., Koller, A. and Hall, M.N. (1990) *FEBS Lett.* 275, 1-5.
- [7] Lee, W.C. and Mélése, T. (1989) *Proc. Natl. Acad. Sci. USA* 86, 8808-8812.
- [8] Silver, P., Sadler, I. and Osborne, M.A. (1989) *J. Cell. Biol.* 109, 983-989.
- [9] Benditt, J.O., Meyer, C., Fasold, H., Barnard, F.C. and Riedel, N. (1989) *Proc. Natl. Acad. Sci. USA* 86, 9327-9331.
- [10] Li, R. and Thomas, J.O. (1989) *J. Cell. Biol.* 109, 2623-2632.
- [11] Kalderon, D., Roberts, B.L., Richardson, W.D. and Smith, A.E. (1984) *Cell* 39, 499-509.
- [12] Goldfarb, T.S., Gariépy, J., Schoolnick, G. and Kornberg, R.D. (1986) *Nature* 322, 641-644.
- [13] Benton, B.M., Eng, W.K., Dunn, J.J., Studier, W., Sternglanz, R. and Fisher, P.A. (1990) *Mol. Cell. Biol.* 10, 353-360.
- [14] Payre, F., Yanicostas, C. and Vincent, A. (1989) *Dev. Biol.* 136, 469-480.
- [15] Payre, F., Noselli, S., Lefrère, V. and Vincent, A. (1990) *Development* 110, 141-149.
- [16] Vincent, A., Colot, H.V. and Rosbash, M. (1985) *J. Mol. Biol.* 186, 149-166.
- [17] Mead, D.A., Szczesna-Skorupa, E. and Kemper, B. (1986) *Protein Eng.* 1, 67-74.
- [18] Kunkel, T.M. (1985) *Proc. Natl. Acad. Sci. USA* 82, 488-492.
- [19] Rubin, G.M. and Spradling, A.C. (1982) *Science* 218, 348-353.
- [20] Robertson, H.M., Preston, C.R., Phillips, R.W., Johnson-Schlitz, Q., Benz, W.K. and Engels, W.R. (1988) *Genetics* 118, 461-470.
- [21] Glaser, R.L., Wolfner, M.F. and Lis, J.T. (1986) *EMBO J.* 5, 747-754.
- [22] Dang, C.V. and Lee, W.M.F. (1988) *Mol. Cell. Biol.* 8, 4048-4054.
- [23] Bier, E., Vaessin, H., Sheperd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Grell, E., Jan, L.Y. and Jan, Y.N. (1989) *Genes Dev.* 3, 1273-1287.
- [24] Bellen, H., O'Kane, C., Wilson, C., Grossniklaus, U., Pearson, R.K. and Gehring, W. (1989) *Genes Dev.* 3, 1288-1300.
- [25] Gibson, G., Schier, A., LeMotte, P. and Gehring, W.J. (1990) *Cell* 62, 1087-1103.
- [26] Schneider-Gadicke, A., Beer-Romero, P., Brown, L.G., Mardon, G., Luoh, S.W. and Page, D.C. (1989) *Nature* 342, 708-711.
- [27] Smith, A.E., Kalderon, D., Roberts, D.L., Colledge, W.H., Edge, M., Gillet, P., Markham, A., Paucha, E. and Richardson, W.D. (1985) *Proc. R. Soc. Lond. Ser. B* 221, 43-53.