Volume 287, number 1,2, 197–202 FEBS 10020 August 1991

ADONIS 001457939100731H

An upstream regulatory element of the NCAM promoter contains a binding site for homeodomains

Marie-Rose Hirsch[1], Isabelle Valarché[1], Hermine Deagostini-Bazin[1], Christine Pernelle[2], Alain Joliot[3] and Christo Goridis[1]

[1]*Centre d'Immunologie INSERM-CNRS de Marseille-Luminy, case 906, F-13288 Marseille Cedex 9, France,* [2]*Institut des Biotechnologies, Centre de Recherche de Vitry, Rhône-Poulenc Santé, B.P. 14, F-94403 Vitry sur Seine, France* and [3]*CNRS URA 1414, Département de Biologie, Ecole Normale Supérieure, 46, rue d'ULM, F-75005 Paris, France*

Received 5 June 1991

In the present study, we have analyzed an upstream regulatory element of the neural cell adhesion molecule (NCAM) promoter which is required for full promoter activity. It contains an ATTATTA motif that resembles the core recognition sequence of homeodomain (HD) proteins of the *Antennapedia* (*Antp*) and related types. Electrophoretic mobility shift (EMSA) and DNase I footprinting analyses revealed that the *Drosophila* HDs coded by the *Antp* and the *zerknüllt* (*zen*) genes bind this site in vitro. In contrast, the *engrailed* (*en*) protein did not produce a detectable footprint. The functional relevance of the ATTATTA motif was demonstrated by showing that a two-nucleotide exchange curtailed stimulation of an heterologous promotor. An oligonucleotide known to be recognized with high affinity by *Antp*-like HDs efficiently competed for endogenous factor binding. These results suggest that the NCAM gene may be a target for HD proteins.

Homeobox; NCAM; Promoter; N2a cell

1. INTRODUCTION

During embryonic development, adhesive interactions between cell surfaces mediated by cell–cell adhesion molecules (CAMs) participate in the control of cell recognition, sorting and migration and may also influence proliferation and differentiation. NCAMs represent a family of closely related proteins that arise from a single gene by differential processing of the pre-mRNA (for reviews, see [1–4]). NCAM shows a dynamic pattern of expression during embryonic development. After its initial appearance at the blastoderm stage, NCAM is expressed transiently on derivatives of all three germ layers often in morphogenetically active regions [4–6]. In the adult, NCAM expression is mainly restricted to nervous tissue [2–4]. On the basis of its adhesive function and its evocative expression patterns, NCAM is believed to be involved in specifying cell patterning and movement in the embryo. This conjecture is supported by experiments, in which nerve and muscle development has been perturbed by injection of anti-NCAM antibodies in vivo [7,8].

It can thus be anticipated that the control of NCAM expression is crucial for the regulation of embryonic development. To learn more about the transcriptional control of NCAM expression, we [9] and others [10,11] have begun to analyze the upstream region of the gene. These studies show that the NCAM gene is transcribed from a single promoter which does not contain a typical TATA box. We found that 750 bp of upstream sequence from the mouse gene contain sufficient information for high-level expression of a reporter gene in transiently transfected NCAM-expressing cells, whereas promoter activity is low in NCAM-negative cells. Much of the transcriptional activity can be ascribed to the NCAM proximal promoter region from −462 to −37 relative to the start site of translation. This activity is modulated by a positively acting element between −645 and −498 and a negatively acting element between −941 and −646. We mapped eight domains of interaction with nuclear proteins to the region analyzed, of which three, termed *a*, *b* and *c*, were localized to the −645 to −498 or *abc* element [9]. Footprint *b* contains a 5′ATTATTA3′ motif, which as shown here, is essential for factor binding and for transcriptional activation of a reporter gene. This motif resembles the central core shared by the sequences recognized by most if not all transcription factors which contain as DNA binding domain a homeodomain of the *Antp* or a related type [12–14]. This prompted us to investigate whether the ATTATTA sequence may be a target for this class of transcription factors, to which belong the products of major developmental control genes in *Drosophila* and probably also in the mouse [12,14–16].

Abbreviations: EMSA, electrophoretic mobility shift analysis; HD, homeodomain; NCAM, neural cell adhesion molecule.

Correspondence address: C. Goridis, CIML, Luminy case 906, F-13288 Marseille Cedex 9, France. Fax: (33) (91) 26 94 30.

2. EXPERIMENTAL

2.1. *Plasmids*

The *abc* fragment used as probe in EMSA and DNase protection assays is the *Nde*I-*Sau*3A fragment from plasmid NS2 [9]. It was cloned in front of the Thy1 promoter in plasmid Thy1-CAT, in which the mouse Thy1.2 promoter region (position − 66 to + 403 according to the numbering of Ingraham and Evans [17]) is linked to the promoterless chloramphenicol acetyltransferase (CAT) gene from the pconaCAT vector [18] yielding Thy1abc-CAT. A mutant version of the *abc* element with the ATTATTA box changed to ACTATCA was derived by polymerase chain reaction of the whole plasmid [19] using as primers the lower strand of oligonucleotide *a* and the upper strand of oligonucleotide b0.

2.2. *Oligonucleotides*

Oligonucleotide *a*:

```
aattCTTTGAAAATCGAACCGAATCTAAAATTCT.....
....GAAACTTTTAGCTTGGCTTAGATTTTAAGAAAAAG
```

is derived from the footprint *a*, oligonucleotide *b*:

```
TTTTCCCCCTAATTATTAAAAACGTTCA....
.....GGGGATTAATAATTTTTGCAAGTTTAA
```

from the footprint *b* sequence [9]; the mutated oligonucleotide b0[9]; oligonucleotide LP [20]; a 26-mer oligonucleotide derived from the Hox1.3-protein binding site [13]; oligonucleotide en 2/2′ [21]; oligonucleotide rat E3 recognized by TARP [22].

2.3. *Expression and purification of polypeptides*

All peptides were expressed in *E. coli* strain BL21 after cloning in pET3a [23] (a gift of William Studier). The 60 amino acid HD of the *Antp* protein was purified to electrophoretic homogeneity as described [24]. Bacterial expression vectors containing the full-length coding sequences of the *Drosophila* HD proteins *zen* (pARzen) and *en* (pAREn) were kindly provided by Timothy Hoey [25]. They were expressed and *E. coli* extracts prepared essentially as described [26].

The methods used for EMSA, DNaseI footprinting and CAT assays have been described [9].

3. RESULTS

Previous experiments [9,10] have shown that the upstream region of the NCAM gene can be subdivided into a proximal promoter region, the integrity of which is required for transcriptional activity, and upstream regulatory elements that modulate this activity. The *abc* element centered around the three adjacent footprints *a*, *b* and *c* increased promoter efficiency around two-fold [9]. This element functions also in the context of a heterologous promoter, since it stimulated transcription from the Thy1.2 promoter five- and two-fold in L and N2a cells, respectively (Fig. 1). As shown below, a two-nucleotide exchange within the centrally located ATTATTA motif contained in footprint *b* (ATTATTA to ACTATCA) is sufficient to eliminate factor binding to this site. When the same mutation was introduced into the *abc* element, its ability to stimulate CAT gene transcription from the Thy1.2 promoter was lost.

When tested by EMSA, the *abc* element was found to form two specific complexes, termed C1 and C2, with N2a cell nuclear factors suggesting that it contains binding sites for at least two nuclear proteins [9]. We investigated the relation between the footprinted domains and the retarded bands by the EMSA experiments shown in Fig. 2. The ligated *a* and *b* oligonucleotides were efficient competitors for the formation of the C1 and C2 complexes and reproduced the pattern seen with the intact fragment when used as labeled probe. Since the other sequences seem dispensable, we used the *ab* sequence as labeled probe in subsequent experiments. In the presence of an oligonucleotide corresponding to footprint *b* formation of complex C2 was completely inhibited, while addition of the *a* sequence eliminated complex C1. Hence, C1 and C2 represent nuclear factor binding to footprints *a* and *b*, respectively. Oligonucleotide b0 bearing the ATTATTA box mutation was unable to compete for factor binding strengthening our conclusion that this motif is essential for recognition of the *b* site. We therefore tested whether binding sites for known factors that contain a central 5′ATTA3′ motif may compete for C2 formation. Two identified recognition sequences of HD proteins containing the motif (en 2/2′ [21] and Hox 1.3

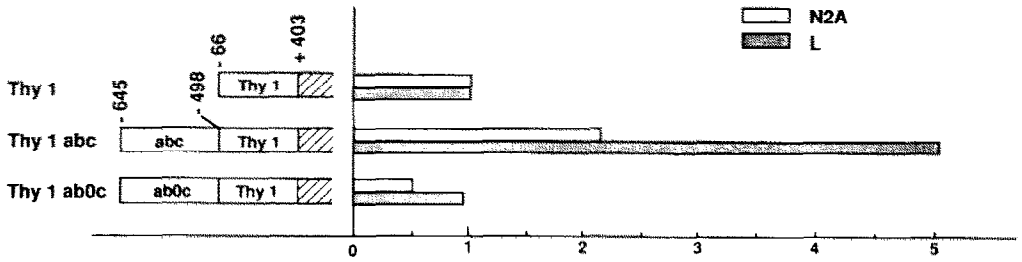

Fig. 1. Effect of wild-type and mutant *abc* elements on CAT expression from the Thy1.2 promoter. The CAT constructs used are schematically depicted at the left of the figure, the hatched boxes signify the CAT coding region. The wild-type and mutant versions of the element are indicated by abc and ab0c, respectively. Equimolar amounts of each plasmid were transiently expressed in N2a (open boxes) or L (black boxes) cells and CAT activities measured. After correction for variations in transfection efficiencies [9], they were normalized with respect to the activity of the proximal promoter without the upstream element in each cell type. The mean values are shown for two to six experiments. The changes due to the *abc* or *ab0c* elements varied not more than ± 17% except for the induction of the Thy1.2 promoter in L cells by the *abc* element where inductions between 2.7-fold and 8.5-fold were recorded.

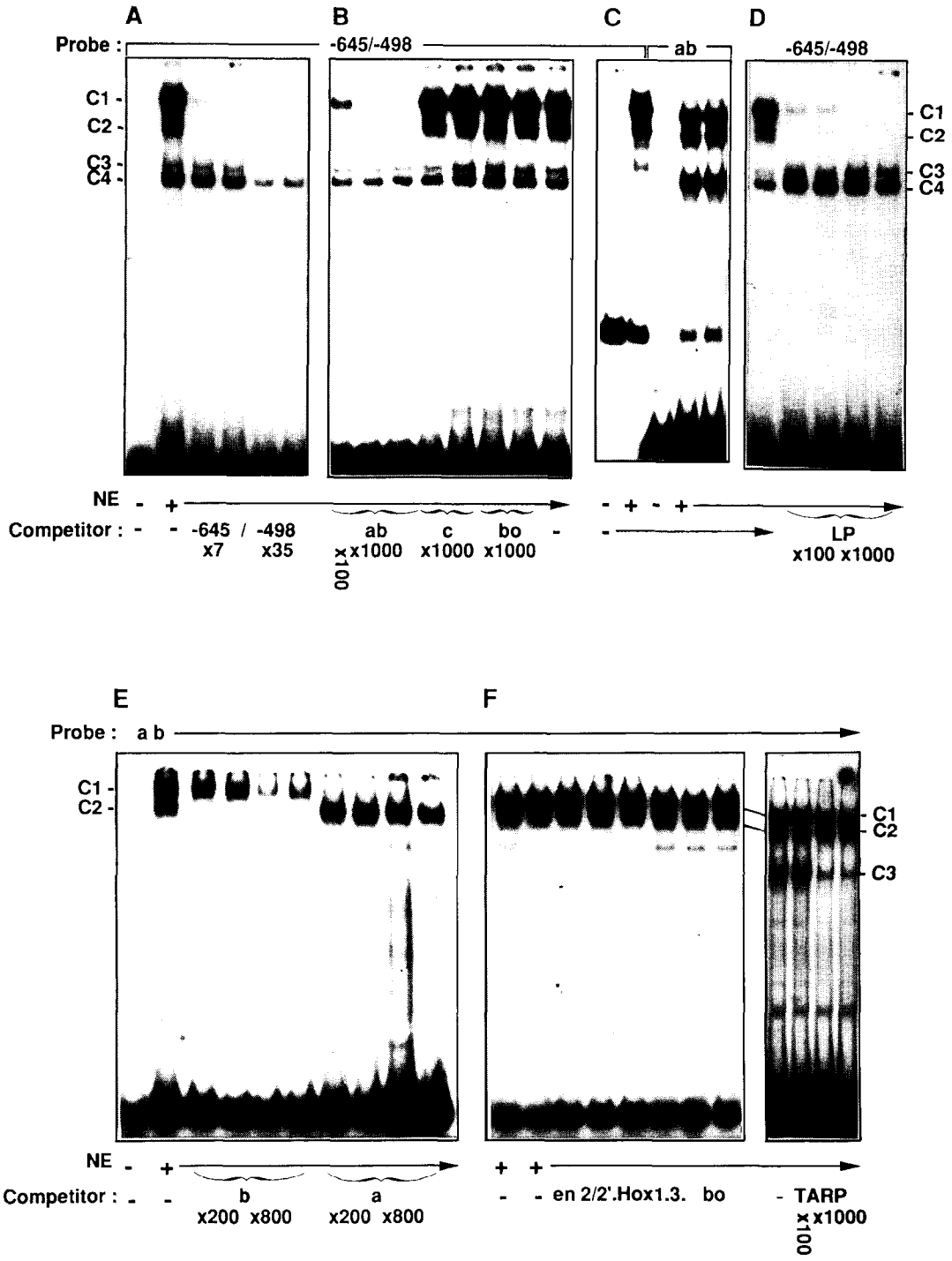

Fig. 2. EMSA analysis of N2a cell nuclear factors that bind to the *abc* element. The labeled probes used are the intact −645 to −498 fragment or the linked *a* and *b* oligonucleotides as indicated above (ab). The different competitor sequences added are indicated below with their respective molar excess with respect to the labeled probes.

[13]) did not affect factor binding. Neither did the A+T-rich TARP-recognition site [22] (Fig. 2F). By contrast, an oligonucleotide (LP) containing palindromic ATTA motifs, a high-affinity binding site for the *en* and *fushi tarazu* proteins [20], completely prevented formation of C2. Complex C1 was also affected suggesting that the factor that binds domain *a* recognizes also the LP sequence albeit with weaker affinity. Together, these results show that the N2a cell nuclear factors that bind the *abc* element interact with a specific subset of ATTA-containing sequences.

To test the possibility that HD proteins may

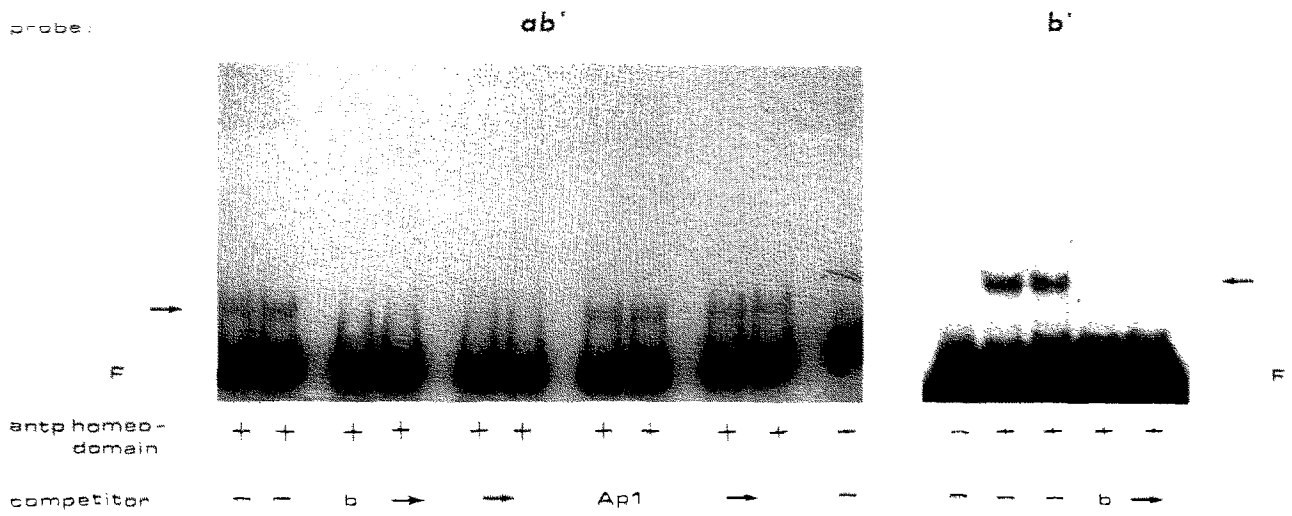

Fig. 3. An isolated *Antp* HD peptide forms a specific complex with the *b* sequence. The purified peptide (20 ng) was incubated with the labeled oligonucleotides *ab* (left panel) or *b* (right panel) (approximately 2.5 fmol) in the absence or presence of a 400- and 1200-fold molar excess of competitor sequences (oligonucleotide *b* or an AP1 binding site). The retarded bands are indicated by the arrow; they migrate close to the free probe (F) due to the small size of the 60 amino-acid protein.

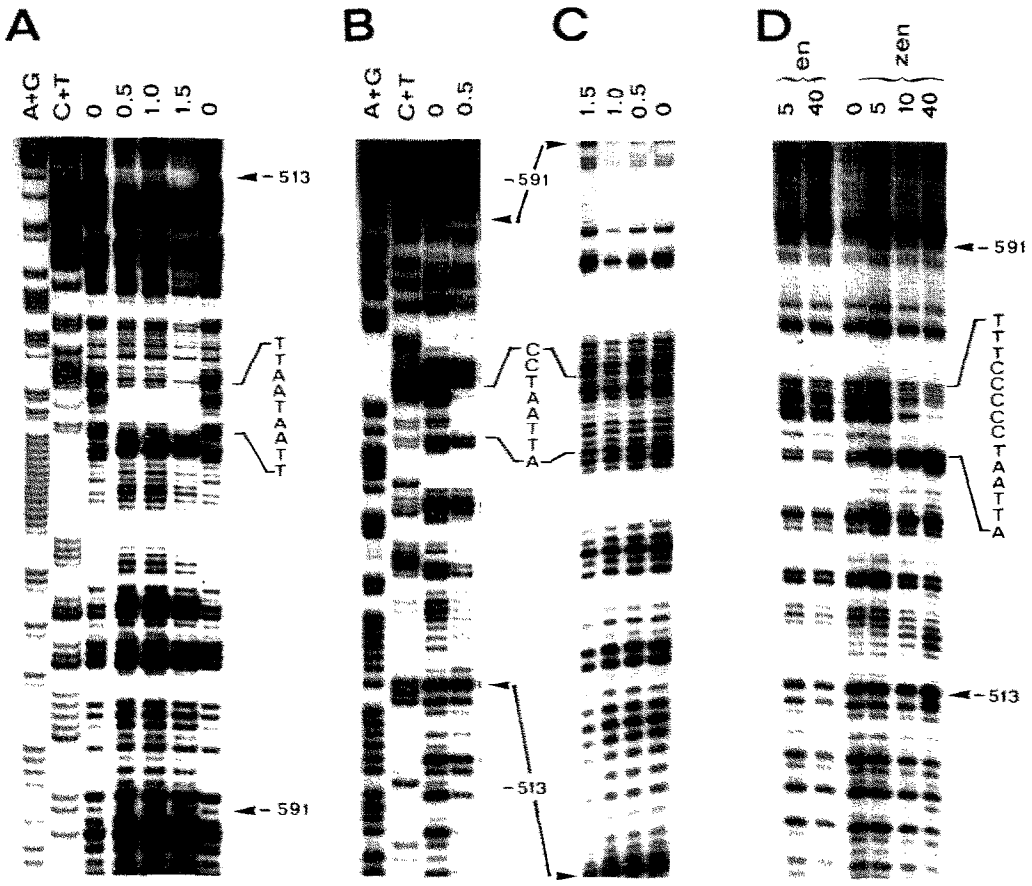

Fig. 4. DNaseI footprint patterns produced by the *Antp* HD peptide and full-length *zen* and *en* proteins. (A) and (B). The end-labeled non-coding (A) and coding (B) strands of the −645 to −498 fragment (20 fmol) were incubated in the absence of protein (0) or in the presence of increasing concentrations of the purified *Antp* HD (0.5 to 1.5 μg as indicated above). A Maxam-Gilbert sequence of the fragment was run alongside (A + G, C + T). (C) No detectable footprint was produced with the fragment bearing the ACTATCA mutation (D) Different amounts (5 to 40 μg of extract protein, 0 signifies incubation in the absence of extract) of bacterial extracts containing the *en* or *zen* proteins were incubated with the end-labeled coding strand.

recognize the essential ATTATTA motif within site *b*, we used the isolated HD from the *Antp* protein and the full-length *zen* and *en* proteins in binding assays. Binding of the purified *Antp* HD was indeed detected by EMSA and could be competed for by the *b* sequence but not by an unrelated factor binding site (Fig. 3). When used in DNaseI protection experiments in amounts (500 ng) comparable to the ones (400 ng) shown to be required for complete protection of an authentic binding site [27], the *Antp* HD protected the CCTAATTA motif on the transcribed and the TTAATAATT sequence on the non-transcribed strand within element *b* (Fig. 4A and B). As expected from its binding specificity, the *Antp* HD peptide did not produce a footprint on the fragment bearing the two C-T exchanges of oligonucleotide b0 (Fig. 4C). The *zen* protein produced a footprint which included the sequence protected by the isolated *Antp* domain. By contrast, comparable amounts of the *en* protein did not produce a detectable footprint. We tried five-fold higher amounts with equally negative results (not shown). The bacterially produced protein was shown to be functional by its ability to form a retarded band when tested with an oligonucleotide corresponding to an authentic *en* binding site (results not shown). Hence, the HDs belonging to two classes, *Antp* and *zen*, are capable of interacting with the *b* element, whereas the *en* protein, which contains an HD of a third type [14], does not bind with an affinity high enough to produce a detectable footprint.

4. DISCUSSION

The HD is an evolutionary conserved 60 amino acid DNA binding domain which has been found in proteins from nearly every eukaryote investigated. Based on sequence homologies, HDs fall into several groups that are conserved across species. *Drosophila* HDs similar to the *Antp* prototype are present in the products of several classes of developmental control genes. In the mouse, the *Hox* genes encode closely related HDs. The *Hox* proteins are supposed to play major roles in specifying anterior–posterior polarity and cell fate in the embryo [12,14–16]. Several HDs with more divergent sequences also have counterparts in flies and mice and display embryonic expression patterns suggesting a role in controlling morphogenetic processes [28,29]. HDs of the *Antp* and similar types including the *en* domain bind with high affinity to a DNA sequence containing matches to either TCAATTAAAT or $(TAA)_n$ [12,13,20].

The natural recognition sequences of these proteins have been found in the promoter regions of genes encoding other transcription factors. However, HD proteins must also influence the expression of other genes which can affect cellular phenotype more directly. Despite some progress towards identifying candidate genes [30,31], these ultimate effector genes have remained largely unknown. Among the prime candidates are the genes that code for CAMs such as NCAM, which are believed to play key roles in specifying cell assembly and movement in the embryo. We have mapped a factor binding site in the NCAM distal promoter region, termed site *b* [9], which contains a potential HD binding site. The EMSA and DNase footprinting experiments shown here demonstrate that this site binds the *Drosophila Antp* and *zen* HDs in vitro suggesting that NCAM expression might be modulated by the products of related genes in the mouse. The two-fold stimulation of NCAM promoter activity afforded by the *abc* upstream element (up to 5-fold when linked to the Thy1.2 promoter) may seem modest. An only two-fold change in transcriptional efficiency may, however, have important functional consequences as comparable changes in NCAM expression have been shown to result in large functional effects [32,33].

The *Ant* and *zen* HDs share 59% amino acid identity [14] and protected the same sequence within site *b*. The *en* HD is somewhat more divergent but recognizes the same sequences [25] and shares with the other two a glutamine in helix 3 which seems to determine binding site specificity [34]. It came thus as a surprise that in contrast to *zen*, a similarly prepared *en* protein did not protect the same site. Apparently, the sequences which flank the ATTA box in site *b* are incompatible with efficient binding of the *en* protein.

Although these results demonstrate that the *Antp* and at least one other *Drosophila* HD bind to site *b* in the NCAM distal promoter region in vitro, we do not know whether the endogenous factor which recognizes this site in N2a cells contains a HD as its DNA binding site. Experiments showing that the LP sequence, which can be assumed to be recognized by all *Antp*-like HDs, very efficiently competed for endogenous factor binding suggest that this may be the case. Together, these results make the NCAM gene an interesting candidate target gene for HD proteins.

Acknowledgements: We thank D. Campèse and E. Termine for oligonucleotide synthesis. We are grateful to T. Hoey (Dept. of Biochemistry, University of California, Berkeley) for the full length *en* and *zen* expression vectors. Other plasmids were kindly provided by A. Israël (Institut Pasteur, Paris), N. Arai (DNAX Research Institute Inc., Palo Alto) and W. Studier (Brookhaven National Lab., Upton, Long Island). We thank A. Prochiantz for stimulating discussions and J.-F. Brunet and P. Golstein for critical reading of the manuscript. This work was supported by institutional grants from INSERM and CNRS and by specific grants from Association pour la Recherche contre le Cancer and Association Française de Lutte contre les Myopathies.

REFERENCES

[1] Edelman, G.M. (1988) Biochemistry 27, 3533–3543.
[2] Nybroe, O., Linnemann, D. and Bock, E. (1988) Neurochem. Int. 12, 251–262.
[3] Walsh, F.S. and Dickson, G. (1989) BioEssays 11, 83–88.
[4] Linnemann, D. and Bock, E. (1989) Dev. Neurosci. 11, 149–173.

[5] Thiéry, J.P., Duband, J.L., Rutishauser, U. and Edelman, G.M. (1982) Proc. Natl. Acad. Sci. USA 79, 6737–6741.
[6] Crossin, K.L., Chuong, C.M. and Edelman, G.M. (1985) Proc. Natl. Acad. Sci. USA 82, 6942–6946.
[7] Thanos, S., Bonhoeffer, F. and Rutishauser, U. (1984) Proc. Natl. Acad. Sci. USA 81, 1906–1910.
[8] Landmesser, L., Dahm, L., Schultz, K. and Rutishauser, U. (1988) Develop. Biol. 130, 645–670.
[9] Hirsch, M.-R., Gaugler, L., Deagostini-Bazin, H., Bally-Cuif, L. and Goridis, C. (1990) Mol. Cell. Biol. 10, 1959–1968.
[10] Barton, H.C., Mann, D.A. and Walsh, F.S. (1990) Biochem. J. 268, 161–168.
[11] Chen, A., Reyes A. and Akeson R. (1990) Mol. Cell. Biol. 10, 3314–3324.
[12] Levine, M. and Hoey, T. (1988) Cell 55, 537–540.
[13] Odenwald, W.F., Garbern, J., Arnheiter, H., Tournier-Lasserve, E. and Lazzarini, R.A. (1989) Genes Devel. 3, 158–172.
[14] Scott, M.P., Tamkun, J.W. and Hartzell III, G.W. (1989) Biochim. Biophys. Acta 989, 25–48.
[15] Akam, M. (1989) Cell 57, 347–349.
[16] Kessel, M. and Gruss, R. (1990) Science 249, 374–379.
[17] Ingraham, H.A. and Evans, G.A. (1986) Mol. Cell. Biol. 6, 2923–2931.
[18] Kimura, A., Israël, A., Le Bail, O. and Kourilsky, P. (1986) Cell 44, 261–272.
[19] Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G. and Galas, D.J. (1989) Nucleic Acids Res. 17, 6545–6551.
[20] Desplan, C., Theis, J. and O'Farrell, P.H. (1988) Cell 54, 1081–1090.
[21] Kassis, J.A., Desplan, C., Wright, D.K. and O'Farrel, P. (1989) Mol. Cell. Biol. 9, 4304–4311.
[22] Horlick, R.A., Hobson, G.M., Patterson, J.H., Mitchell, M.T. and Benfield, B.A. (1990) Mol. Cell. Biol. 10, 4826–4836.
[23] Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Methods Enzymol. 185, 61–89.
[24] Joliot, A., Pernelle, C., Deagostini-Bazin, H. and Prochiantz, A. (1991) Proc. Natl. Acad. Sci. USA 88, 1864–1868.
[25] Hoey, T. and Levine, M. (1988) Nature 332, 858–861.
[26] Hoey, T., Warrior, R., Manak, J. and Levine, M. (1988) Mol. Cell. Biol. 8, 4598–4607.
[27] Müller, M., Affolter, M., Leupin, W., Otting, G., Wüthrich, K. and Gehring, W.J. (1988) EMBO J. 7, 4299–4304.
[28] Bastian, H. and Gruss, P. (1990) EMBO J. 9, 1839–1852.
[29] Davis, C.A., and Joyner, A.L. (1988) Genes Devel. 2, 1736–1744.
[30] Gould, A.P., Brookman, J.J., Strutt, D.I. and White, R.A.H. (1990) Nature 348, 308–312.
[31] Immerglück, K., Lawrence, P.A. and Bienz, M. (1990) Cell 62, 261–268.
[32] Hoffman, S. and Edelman, G.M. (1983) Proc. Natl. Acad. Sci. USA 80, 5762–5766.
[33] Doherty, P., Fruns, M., Seaton, P., Dickson, G., Barton, C.H., Sears, T.A. and Walsh, F.S. (1990) Nature 343, 464–466.
[34] Treisman, J., Gönczy, P., Vashishta, M., Harris, E. and Desplan, C. (1989) Cell 59, 553–562.