

Ubiquitous Nuclear Factors Bind Specifically to a 5'-Region Conserved in Carcinoembryonic Antigen-Related Genes

Fritz Rudert^{§*}, John Thompson and Wolfgang Zimmermann

Institute of Immunobiology, University of Freiburg, D-7800 Freiburg, Germany

[§] Laboratoire de Génétique Moléculaire du CNRS, Unité 184 de Biologie Moléculaire du Génie Génétique de L'INSERM, Faculté de Médecine, 11 Rue Humann, F-67085 Strasbourg CEDEX, France

Received April 28, 1992

Summary: We recently cloned members of the murine carcinoembryonic antigen (CEA) gene family, some of which are differentially expressed during placental development. By intra- and interspecies sequence comparisons, we identified an element in the putative promoter and/or 5'-nontranslated region which is conserved within all human and rodent CEA-related genes analyzed so far. Using gel retardation analysis and DNaseI hypersensitive site mapping, we now show that ubiquitously expressed nuclear factors specifically bind to the conserved region derived from the mouse gene *Cea-2 in vitro* and probably also *in vivo*. Another DNaseI hypersensitive site lies within or close to a simple sequence motif [(GGA)_n] located in the first intron of *Cea-2*. Such sequences have been reported to play a role in the regulation of certain genes. Therefore, this analysis has identified putative regulatory regions for *Cea-2* and possibly CEA-related genes in general. © 1992 Academic Press, Inc.

Carcinoembryonic antigen (CEA), a highly glycosylated protein, was originally described as an oncofetally expressed human tumor-specific antigen and was initially isolated from colon carcinomas (1,2). The later discovery of CEA in normal adult tissue and the existence of a large number of immunologically cross-reacting antigens has limited its potential use as a specific tumor marker (reviewed in 3). Through molecular cloning it has been shown during the last few years that CEA is a member of a large gene family comprising approximately 20 genes

*To whom correspondence should be addressed.

in humans clustered on the long arm of chromosome 19 (reviewed in 4). This family can be subdivided into the CEA subgroup consisting of genes encoding CEA and the classical cross-reacting antigens and the PSG subgroup which encode the pregnancy specific glycoproteins (PSG) (see 4). Recently, members of the rodent CEA gene family have been discovered (5-8,11). The CEA gene families of rodents and humans both belong to the immunoglobulin supergene family (9-11) and have been shown to have evolved after mammalian radiation (12). Due to the apparently rapid evolution, a great sequence divergence and also striking structural dissimilarities have been observed between rodent and human CEA-related genes (7,8), making it impossible to assign individual counterparts in the respective species (12). However, by comparing their patterns of expression we have recently identified analogues of human PSG genes in both rat and mouse (7,8). The identical expression pattern of the structurally dissimilar human and rodent PSG genes makes it very likely that they have similar or identical transcriptional regulation mechanisms. Indeed, sequence comparisons have allowed the identification of a conserved region in the regulatory regions of genes of the CEA families (8). In this study, using the mouse PSG gene analogue *Cea-2* (8), we demonstrate that ubiquitously expressed nuclear factors can bind to specific sequence elements of this conserved region *in vitro* and possibly also *in vivo*.

Materials and Methods

Preparation of nuclear extracts : Placentae of staged mice (8) or other tissues of adult mice were used as a starting material. All steps were carried out on ice. The usual procedure to prepare nuclear extracts was as described (13) with the following modifications: Tissues were removed, immediately washed in 0.14 M NaCl, 10 mM Tris-Cl pH 8.0 and cut into pieces. Buffer I was supplemented with 0.1 % Triton X-100. After dialysis against buffer III, precipitates were removed by centrifugation, the supernatants were frozen in liquid N₂ and afterwards directly used in gel retardation without further column chromatography. As an alternative method a hypotonic extraction procedure without detergent was used. Tissues were removed, transferred to PBS supplemented with 2 mM MgCl₂, 5 mM KCl and minced followed by homogenization in 10 ml/g tissue in the same buffer. The cell suspension was centrifuged for 5 min at 375xg and the pellet lysed in a hypotonic buffer containing 10 mM HEPES pH 8.0, 2 mM MgCl₂, 5mM KCl, 0.5 mM spermidine (Sigma), 0.15 mM spermine (Sigma), 0.5 mM phenyl-methyl-sulfonylfluoride (PMSF), 7 mM β-

mercaptoethanol, 0.25 $\mu\text{g/ml}$ aprotinin (Sigma), 10 $\mu\text{g/ml}$ leupeptin (Sigma) with 15 strokes in a glass potter. Sucrose was slowly added to a final concentration of 0.3 M and the nuclei were centrifuged for 5 min at 900xg. The nuclei were washed with hypotonic buffer containing 0.3 M sucrose and pelleted by centrifugation for 5 min at 800xg. The nuclei were then resuspended in 5 to 7 volumes of the pellet in elution buffer (10 mM HEPES pH 8.0, 0.1 mM EDTA, 0.1 mM EGTA, 0.5 mM spermidine, 0.15 mM spermine, 420 mM NaCl, 10 % glycerol, 0.5 mM PMSF, 7 mM β -mercaptoethanol, 0.25 $\mu\text{g/ml}$ aprotinin, 10 $\mu\text{g/ml}$ leupeptin) and the nuclear proteins eluted for 30 min with occasional shaking. Insoluble material was removed by centrifugation at 23000xg for 20 min and the supernatant frozen in liquid N_2 .

Gel retardation assay : Binding reactions contained 40 mM KCl, 10 mM HEPES pH 7.6, 4 % Ficoll (Pharmacia), 1 mM EDTA, 1 mM DTT, 1 μg poly(dIdC) (Pharmacia), 0.5 mM ATP (as cold competitor for free [^{32}P]-ATP), 5 μg nuclear extract, 0.2 pmole [^{32}P]-ATP end-labeled double-stranded probe (specific activity 5 to 7.5 $\times 10^7$ dpm/ μg) and if required, a 250-, 300- or 1000-fold molar excess of the respective double-stranded unlabeled competitor in a reaction volume of 25 μl . Pipetting was done on ice followed by a 30 min incubation at room temperature. Three μl loading buffer were added and the complexes separated on 4 % nondenaturing polyacrylamide gels run in 0.5 x TBE at 12 V/cm. Gels were vacuum dried and autoradiographed using X-ray film.

DNaseI hypersensitive site mapping : All steps were carried out on ice. Placentae were removed and homogenized in 30 ml/g tissue buffer B (0.15 mM spermine, 0.5 mM spermidine, 15 mM Tris-Cl pH 7.5, 60 mM KCl, 15 mM NaCl, 2 mM EDTA, 0.5 mM EGTA, 0.5 M sucrose) with 10 strokes in a glass potter. Cells were centrifuged for 5 min at 1500xg, resuspended in 20 ml buffer C (buffer B with 0.5 % Triton X-100) and lysed with 10 strokes in a glass potter. Nuclei were pelleted by centrifugation for 5 min at 1500xg, washed in buffer D (buffer B with 0.35 M sucrose) and the pellet frozen in liquid N_2 . 5×10^8 nuclei were resuspended in 2.25 ml buffer E (buffer B without sucrose and with 0.2 mM EGTA, 1 mM PMSF) and aliquoted at 250 μl . These aliquots were digested with increasing amounts of DNaseI (Worthington) in a total volume of 500 μl buffer E for 15 min on ice. Reactions were started by addition of 20 μl MgCl_2 (100 mM) and stopped with 10 μl EDTA (0.5 M). The DNaseI concentration was increased by doubling the units, starting with 32 units (Fig. 3, lane 3). Genomic DNA was prepared, digested with BamHI and 6.5 μg per lane were separated on a 1 % agarose gel. The DNA was transferred to Immobilon-NTM and hybridized with a [^{32}P]-labeled 310 bp StuI/BamHI fragment from *Cea-2* comprising most of the second exon (8). Filters were washed in 0.1 x SSPE/0.1 % SDS at 65 $^\circ\text{C}$ and autoradiographed at -80 $^\circ\text{C}$ using an intensifying screen.

Results and Discussion

The strong interspecies conservation of a region lying in the vicinity of or at the transcriptional start sites of all CEA-related genes identified so far (8) prompted us to investigate whether this region might have a role in the regulation of these genes. We have chosen gel retardation assays to analyze whether nuclear factors can bind to the conserved region *in vitro*. As the mouse PSG analogues seem to be exclusively expressed in placenta, we prepared nuclear extracts from staged mouse placentae to look for such putative factors in this organ. As shown in Figure 1A, a mobility shift of the labeled complex can be observed in a stage where the mouse PSG analogue *Cea-2* is expressed (day 18) as well as in an inactive stage [(day 11), 8]. The strongly labeled band seen at d18 and not at d11 presumably represents an ATP-binding or a phosphorylated protein, because it is not visible after including cold ATP in the binding reaction. The complexes indicated in brackets seem to differ qualitatively at the different stages (Fig. 1A). The uppermost complex formed with day 11 placental extract (marked by an arrow) is apparently not present when day 18 nuclear extract is used.

To analyze which part of the conserved region is involved in DNA binding we cut the whole region down to individual subregions using double-stranded oligonucleotides. Interestingly, only the flanking regions A and D gave a retarded protein/DNA complex (Fig. 1B). The formation of such complexes could be drastically reduced when an excess of the respective unlabeled DNA was added as a competitor to the binding reactions (Fig. 1C). In Figure 1D, a competition experiment using a high molar excess of either the specific competitors (A or D, respectively) or oligonucleotide B as a non-specific competitor is shown to rule out non-specific effects. Interestingly, it turned out that the two regions which bind nuclear factors are interchangeable to a certain extent in a competition assay, exhibiting cross-competition of each other (Fig. 1E). This makes it possible that related or even identical factors take part in complex formation with sequences present in probes A and D. In this context, it should be noted that regions A and D share an identical pentamer sequence which forms an inverted repeat (see Fig. 1 bottom). Whether such a motif represents a binding site for the putative regulatory factors remains to be shown.

To examine whether the expression of such factors is exclusively restricted to placenta or can also be found in other tissues, we tested nuclear extracts from various adult mouse tissues as well as from the human colon adenocarcinoma cell line Lovo and the promyelocytic cell

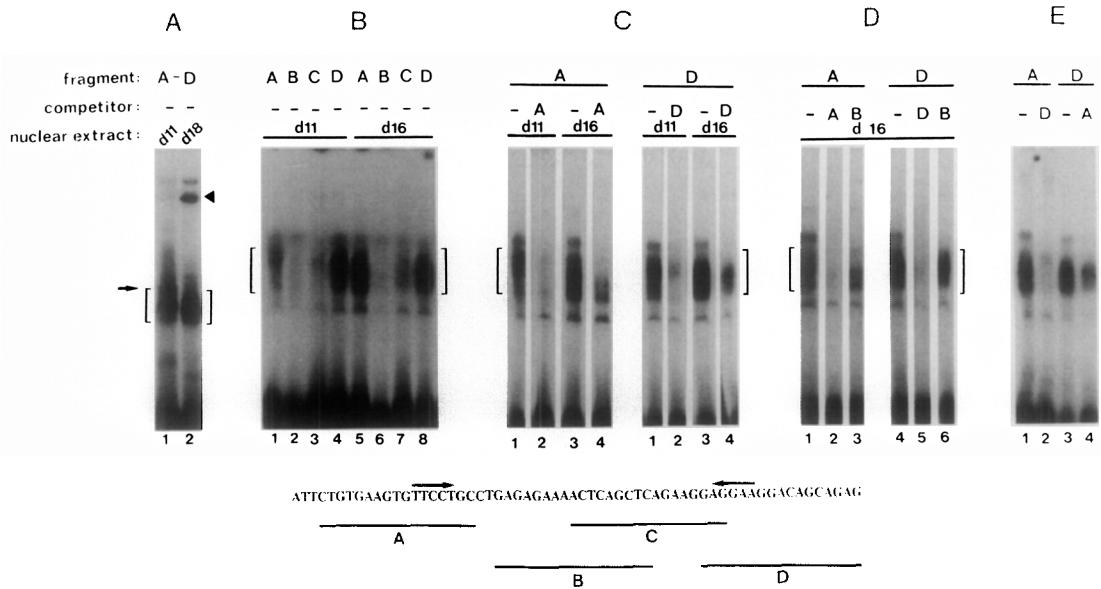


Fig. 1. Nuclear factors specifically bind to sequences in the conserved region of *Cea-2*. **A**, the conserved region generated by using [³²P]-labeled PCR primers A and D (see bottom) was used in the gel retardation assay. The arrow indicates the additional complex formed in the presence of d11 placenta nuclear extract. The filled arrow head marks a non-specific complex which is absent when cold ATP is included in the binding reaction. **B**, individual double-stranded oligonucleotides derived from the conserved region (A-D) were employed in the gel retardation assay to narrow down the binding region. **C**, specific competition of complex formation is shown using a 250x molar excess of the indicated unlabeled oligonucleotides. **D**, the binding specificity of regions A and D is analyzed using a high molar excess (1000x) of either the specific competitors A or D, respectively or of oligonucleotide B as a nonspecific competitor. **E**, the ability of oligonucleotides A and D to reciprocally compete each other has been tested (300x). Specific complexes are indicated by brackets. The stages of the mouse placental nuclear extracts are indicated above the lanes. In **E**, nuclear extract from day 16 mouse placenta was used. At the bottom of the figure the sequence of the conserved region of *Cea-2* (8) is shown together with the position of the oligonucleotides which have been used in gel retardation. The arrows above the sequence indicate an inverted repeat motif.

line HL60. We detected shifts in the mobility of oligonucleotide D with all the extracts tested (Fig. 2), which argues that the factors are ubiquitously expressed. However, slight differences in the mobilities of such complexes may be due to the binding of related but not identical factors or the absence of certain factors in some extracts. Moreover, the presence of a nuclear factor capable of binding to this region at least in HL60 cells (and possibly also Lovo) indicates that these

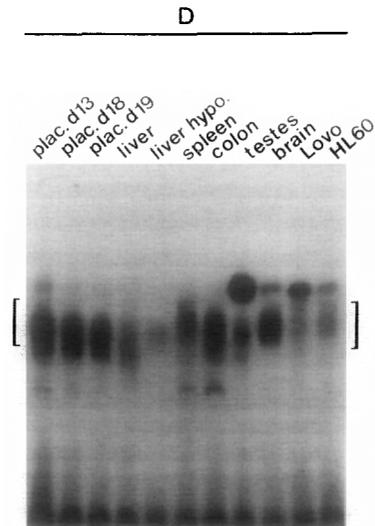


Fig. 2. Tissue distribution of the nuclear factors. A panel of nuclear extracts from different mouse tissues as well as two human cell lines was tested with probe D (see Fig. 1, bottom). Liver hypo. refers to a hypotonic extraction procedure for the nuclear proteins (see Materials and Methods). The specific complexes are bracketed.

proteins are evolutionarily conserved. This would not be surprising for a putative transcription factor as numerous examples exist that such proteins are under highly selective pressure and are therefore very often strongly conserved throughout the animal kingdom (14). Indeed, chinese hamster ovary cells which were stably transfected with a cosmid clone containing the complete human CEA gene correctly expressed CEA (15) confirming that rodent *trans*-acting factors interact with the regulatory sequences of the CEA gene.

The location of the conserved region in the middle of the 5'-nontranslated region of *Cea-2* also makes the involvement in posttranscriptional regulation possible, as has been shown for the TAR region of HIV1 mRNA (16). Although we found no mobility changes when single-stranded versions of the probes A to D were used (data not shown), we cannot firmly rule out a contribution of this region at the posttranscriptional level.

Based on these *in vitro* studies we looked for *in vivo* protein/DNA interactions at the conserved region of *Cea-2*. To achieve this, we digested chromatin from d11 placental nuclei (a stage when no *Cea-2* mRNA can be detected) with increasing amounts of DNaseI to look for DNaseI hypersensitive sites in the putative regulatory regions of the *Cea-2* gene which would be indicative of such interactions. This analysis indicates the presence of a hypersensitive site within or near

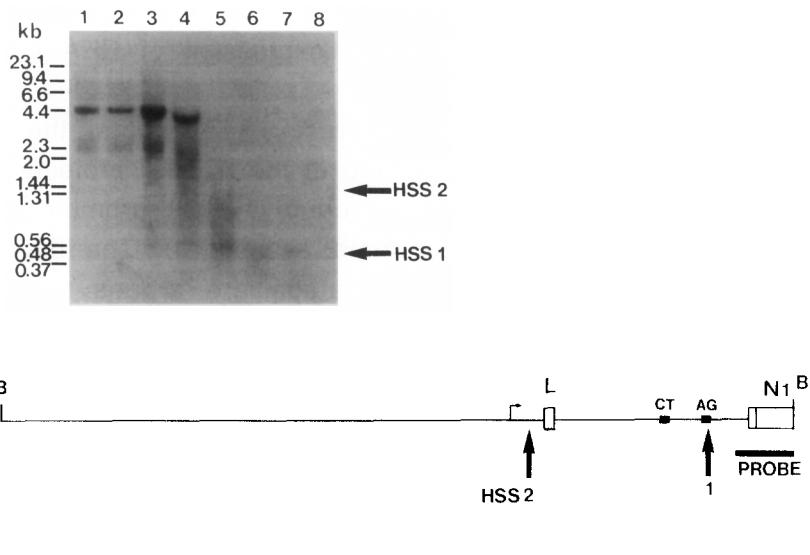


Fig. 3. Mapping of DNaseI hypersensitive sites in the putative control regions of *Cea-2*. Nuclei, isolated from day 11 mouse placenta were digested with increasing amounts of DNaseI (lanes 3-8). No DNaseI was added in lanes 1 and 2. The DNA was extracted from the nuclei and digested with BamHI to completion. After transfer of the DNA the membrane was hybridized with a genomic probe derived from *Cea-2*. In the lower part of the Figure, a physical map of the genomic BamHI fragment of *Cea-2* which is detected by the probe is shown. The positions of DNaseI hypersensitive sites (HSS) are indicated with arrows. The thin horizontal arrow indicates the transcriptional start of *Cea-2*. The filled boxes represent polypyrimidine [(CCTT)_n] or polypurine [(AGG)_n, unpublished] simple sequences, respectively. Open boxes represent exons. L, leader exon; N, N-terminal exon; B, BamHI.

to the conserved region (HSS2 in Fig. 3) which correlates well with the *in vitro* data (see Fig. 1). It remains to be shown whether this *in vivo* interaction reflects a repressor activity present at a stage when *Cea-2* is not transcribed or the inactive state of this gene is caused by the absence of additional activating factors required at this site. In addition, another hypersensitive site (HSS1 in Fig. 3) has been found in the first intron of *Cea-2* which apparently coincides with a (GGA)_n simple sequence (unpublished). Simple sequences have been reported for a number of other rodent (7,8,11) and human (17,18) PSG genes. Similar copies of such sequences are also found in other types of genes and have been implicated in a variety of biological functions, such as gene regulation (19) or as hot spots for recombination (20). Whether these sequences are involved in the regulation of the mouse PSG analogue *Cea-2* and possibly other PSG genes remains to be elucidated. The

present study, however, already gives some insight into putative control regions for the regulation of expression of CEA-related genes for future functional analysis.

It will be of particular interest to elucidate the regulatory potential of the conserved region and of the nuclear proteins which bind to it, especially in view of the lack of typical promoter sequences (TATA-box, CAAT-box) in CEA-related genes (8). The high sequence conservation of the region investigated here argues for a common regulatory element such as a minimal promoter or an "initiator" (21). However, a potential involvement in spatiotemporal transcriptional regulation via tissue-specific factors is not excluded. The latter possibility is strengthened by the fact that a region of the human CEA gene comprising 314 nucleotides upstream of the transcriptional start and therefore including the whole conserved region analyzed here has been shown to convey cell-specific regulation in transient transfection experiments (22). Such behavior, however, was not found for the respective, highly similar region of the non-specific cross-reacting antigen (NCA) gene, a closely related member of the human CEA-subgroup. In this context, it might be interesting to note that a number of nucleotide sequence differences are observed between *CEA* and *NCA* in this conserved region. Future functional studies on the promoter activity should shed some light on the regulation of the restricted spatiotemporal expression of PSG genes in rodents and human.

Acknowledgments

We would like to thank L. Shemshedini for critical reading of the manuscript. This work was supported by a grant from the Dr. Mildred Scheel Stiftung für Krebsforschung.

References

1. Gold, P., and Freedman, S.O. (1965) *J. Exp. Med.* **121**, 439-462.
2. Gold, P., and Freedman, S.O. (1965) *J. Exp. Med.* **122**, 467-481.
3. Thompson, J., and Zimmermann, W. (1988) *Tumor Biol.* **9**, 63-83.
4. Thompson, J., Grunert, F., and Zimmermann, W. (1991) *J. Clin. Lab. Anal.* **5**, 344-366.
5. Turbide, C., Rojas, M., Stanners, C., and Beauchemin, N. (1991) *J. Biol. Chem.* **266**, 309-315.
6. Lin, S.H., and Guidotti, G. (1989) *J. Biol. Chem.* **264**, 14408-14414.

7. Rebstock, S., Lucas, K., Thompson, J., and Zimmermann, W. (1990) *J. Biol. Chem.* **265**, 7872-7879.
8. Rudert, F., Saunders, A.M., Rebstock, S., Thompson, J., and Zimmermann, W. (1992) *Mam. Genome*, in press.
9. Paxton, R., Mooser, G., Pande, H., Lee, T.D., and Shively, J.E. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 920-924.
10. Williams, A.F. (1987) *Immunol. Today* **8**, 298-303.
11. Kodolja, V., Lucas, K., Barnert, S., von Kleist, S., Thompson, J.A., and Zimmermann, W. (1989) *J. Biol. Chem.* **264**, 6909-6912.
12. Rudert, F., Zimmermann, W., and Thompson, J.A. (1989) *J. Mol. Evol.* **29**, 126-134.
13. Heberlein, U., and Tjian, R. (1988) *Nature* **331**, 410-415.
14. Amero, S.A., Kretsinger, R.H., Moncrief, N.D., Yamamoto, K.R., and Pearson, W.R. (1992) *Mol. Endocrinol.* **6**, 3-7.
15. Hefta, L.J.F., Schrewe, H., Thompson, J.A., Oikawa, S., Nakazato, H., and Shively, J.E. (1990) *Cancer Res.* **50**, 2397-2403.
16. Edery, I., Petryshyn, R., and Sonenberg, N. (1989) *Cell* **56**, 303-312.
17. Thompson, J.A., Mauch, E.M., Chen, F.S., Hinoda, Y., Schrewe, H., Berling, B., Barnert, B., von Kleist, S., Shively, J.E., and Zimmermann, W. (1989) *Biochem. Biophys. Res. Commun.* **158**, 996-1004.
18. Thompson, J., Koumari, R., Wagner, K., Barnert, S., Schleussner, C., Schrewe, H., Zimmermann, W., Müller, G., Schempp, W., Zanietta, D., Ammaturo, D., and Hardman, N. (1990) *Biochem. Biophys. Res. Commun.* **167**, 848-859.
19. Hamada, H., Seidman, M., Howard, B.H., and Gorman, C.M. (1984) *Mol. Cell. Biol.* **4**, 2622-2630.
20. Slightom, J.L., Blechl, A.E., and Smithies, O. (1980) *Cell* **21**, 627-638.
21. Smale, S.T., and Baltimore, D. (1989) *Cell* **57**, 103-113.
22. Schrewe, H., Thompson, J., Bona, M., Hefta, L.J.F., Maruya, A., Hassauer, M., Shively, J.E., von Kleist, S., and Zimmermann, W. (1990) *Mol. Cell. Biol.* **10**, 2738-2748.