

# An insulinoma nuclear factor binding to GGGCCC motifs in human insulin gene

Louise Reibel, Corinne Besnard, Patrick Lores, Jacques Jami and Gerard Gacon\*  
Institut Cochin de Génétique Moléculaire, INSERM U. 257, 24 Rue du Fbg St Jacques, 75014 Paris, France

Received January 7, 1993; Revised and Accepted March 1, 1993

## ABSTRACT

**Cell specific expression of the insulin gene is achieved through transcriptional mechanisms operating on 5' flanking DNA elements. In the enhancer of rat I insulin gene, two elements, the Nir and Far boxes, located at positions –104 and –233 respectively and containing the same octameric motif are essential for B cell specific transcription activity. Homologous sequences are present in the human insulin gene. While studying the binding of nuclear proteins from insulinoma cells to the –258/+241 region of the human insulin gene, we observed a previously undetected protein binding site in the intron I region between nucleotides +160 and +175. The binding activity was present in insulin producing cells such as RIN and HIT insulinoma cells but not in fibroblasts or insulin negative fibroblast × RIN hybrid cells. DNase I footprinting and gel retardation/methylation interference experiments allowed us to define the core binding site of the intron binding factor as a GGGCCC hexamer. This factor is also capable to bind to a related sequence, contiguous to the Far-like element in rat and human insulin genes. The binding of the GGGCCC binding factor in this critical region of the insulin gene enhancer may participate in the regulation of insulin gene expression.**

## INTRODUCTION

Specific expression of the insulin gene in the B cells of pancreatic islets has been shown to be controlled by 5' flanking regulatory regions of DNA. Studies of rat insulin 1 gene have demonstrated the presence of cell specific enhancer and promoter elements (1, 2). In particular, the so called Nir and Far boxes, located at positions –104 and –233 from the transcription start site respectively, contain two identical octameric motifs which are essential for B cell specific transcription activity (3, 4). Proteic factors capable to bind to these motifs have been described (5–9). Sequences homologous to these motifs are also present in rat II (10–12) and human insulin genes. Recently, protein binding studies of the human insulin gene have shown that there are multiple binding sites for proteic factors with varying cell specificities between positions –278 and –77 (13).

Previous studies on human insulin gene have demonstrated that sequences spread from –258 to +241 of the transcription origin are sufficient to direct the expression of a linked reporter gene in rat insulinoma cells (1) and its extinction in insulinoma × fibroblast hybrid cells (14–15). This indicated that the –258/+241 fragment contains cis acting elements involved in both positive and negative control of insulin gene expression. We have analysed nuclear extracts from insulin producing (RIN and HIT insulinoma) and non producing (fibroblast and RIN × fibroblast hybrid) cells for proteic factors capable to bind to the –258/+241 region. During the course of this study, a previously undetected binding site for proteic factor(s) present in insulin producing cells was observed in the intron I region of the human insulin gene. The binding site was characterized by DNase I footprinting and gel retardation/methylation interference experiments. The core sequence of the binding site consists of the hexamer GGGCCC. A closely related sequence, adjacent to the Far-related sequence in rat and human insulin genes, is capable to bind the same factor.

## MATERIALS AND METHODS

### Cells

The insulin producing cells RIN 2A (RIN) derive from an X-Ray induced rat insulinoma (16). C11D fibroblast clone is a derivative of the L mouse cell line (17). The somatic cell hybrid Rp3L, generated by fusing RIN cells with C11D cells, is devoid of insulin production, as previously described (14, 15). HIT M 2. 2. 2 (HIT) is an insulin producing cell line derived from syrian hamster endocrine pancreas (2). All the cells were grown in Dulbecco's modified Eagle's medium supplemented with 8% fetal bovine serum.

### Preparation of nuclear extracts

Nuclear extracts were prepared according to Dignam et al. (18). Nuclear proteins precipitated by 50% (w/v) ammonium sulfate were resuspended in 20 mM Hepes pH 8, 25% (v/v) glycerol, 1.2 mM EDTA, 60 mM KCl, 2mM dithiothreitol, 1mM PMSF, dialysed against the same buffer, aliquoted and frozen at –80°C.

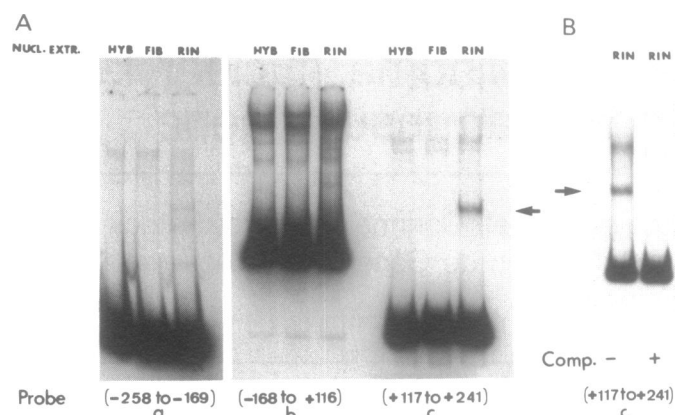
\* To whom correspondence should be addressed

### DNase I footprinting

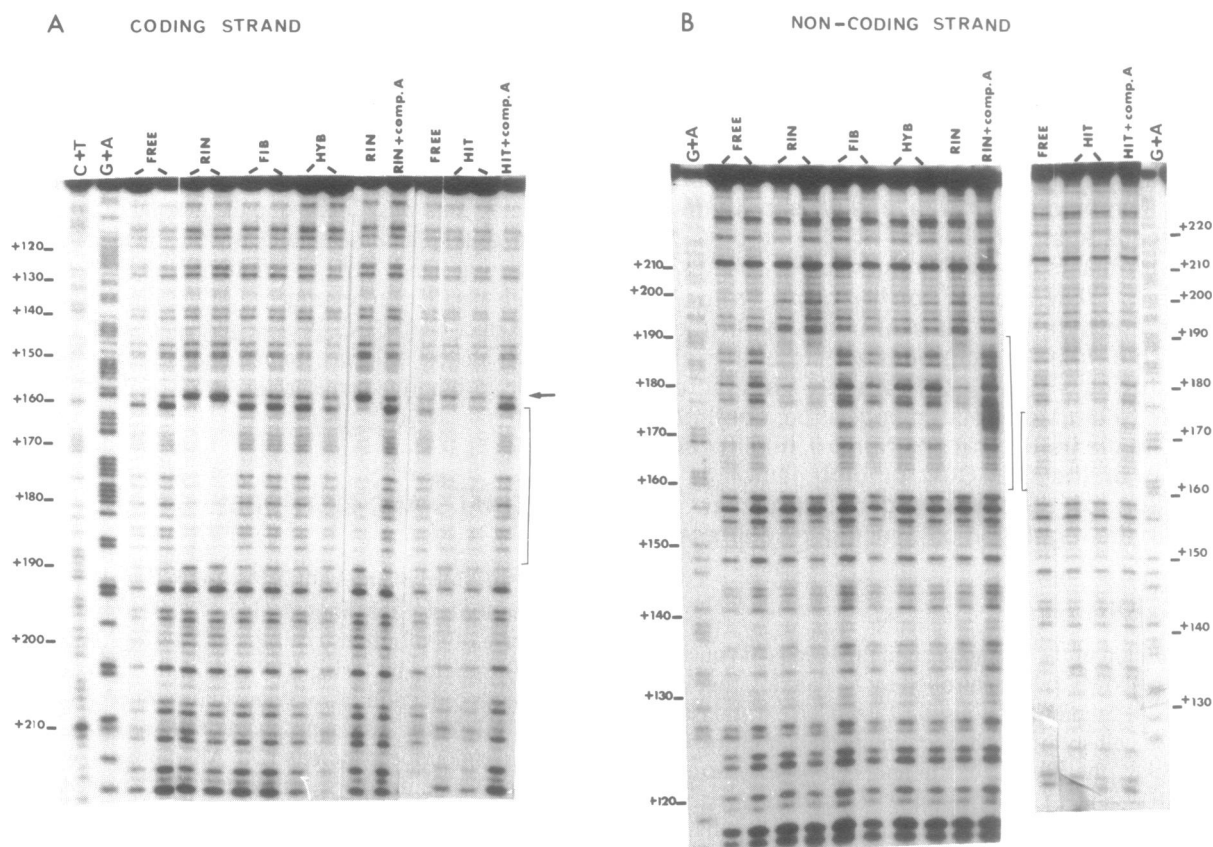
The fragment spanning nucleotides +94 to +241 was 3' end labeled at position +241 (coding strand) or +94 (non coding strand). The reaction was performed in a final volume of 20  $\mu$ l containing 20 mM Hepes (pH 8), 50 mM NaCl, 80 mM KCl, 0.1 mM EDTA, 3 mM dithiothreitol, 5 mM MgCl<sub>2</sub>, 4 mM spermidine, 30% (v/v) glycerol, 500 ng of poly(dI-dC), 1 ng of labeled probe and 10 to 30  $\mu$ g of nuclear extract. The mixture was preincubated for 30 min at 0°C; the concentration of CaCl<sub>2</sub> was then adjusted to 2.5 mM and the incubation was prolonged for one minute at 20°C. The amount of DNase I was adjusted according to the nuclear extracts to produce in one minute at 20°C a homogenous cleavage pattern. Digestion was stopped by the addition of 35  $\mu$ l of a solution containing 150  $\mu$ g/ml tRNA, 450  $\mu$ g/ml proteinase K, 0.06% (w/v) SDS, and 6 mM EDTA. Digests were incubated at 42°C for 30 minutes, then phenol and ether extracted, ethanol precipitated and run on a 7% polyacrylamide/7M urea sequencing gel.

### Gel mobility shift assay

Band shift assays using 3' end labeled DNA fragments or  $\gamma$ <sup>32</sup>P-ATP kinased oligonucleotides were performed as described (19, 20). 5  $\mu$ g of nuclear extract and 0.5 to 1 ng of probe were added



**Figure 1.** Binding of nuclear extracts to the -258/+241 region of the human insulin gene. The binding specificities of nuclear proteins were tested by gel shift assay. Probes used are defined at the bottom of the figure. Nuclear extracts were prepared from RIN2A cells (RIN), CL1D fibroblasts (FIB) and RIN $\times$ Fibroblast hybrid cells (HYB). A/ Nuclear extracts were added to <sup>32</sup>P labeled probes a, b and c and the resulting protein DNA complexes were separated by electrophoresis on a 6% polyacrylamide non denaturing gel. B/ The binding specificity of RIN-probe c complex was tested by adding a 40 fold excess of unlabeled probe. Arrows indicate the position of the complex formed.



**Figure 2.** Dnase footprint analysis of the binding of RIN and HIT factors to the +94/+241 fragment of human insulin gene. Fragment +94/+241 labeled at positions +241 (coding strand, panel A) or +94 (non coding strand, panel B) was used as probe in the DNA footprinting assay with nuclear extracts prepared from RIN2A cells (RIN), CL1D fibroblasts (FIB), RIN $\times$ Fibroblast hybrid cells (HYB) and HIT M2.2.2 cells (HIT) and added as indicated below. Panel A, from left to right: C+T and G+A sequencing ladders, no protein added (free), 16 and 24  $\mu$ g of RIN extract, 13 and 20  $\mu$ g of FIB extract, 16 and 32  $\mu$ g of HYB extract, 16  $\mu$ g of RIN extract, 16  $\mu$ g of RIN extract plus 20 ng oligo A (see table I), no protein added (free), 25 and 40  $\mu$ g of HIT extract, 25  $\mu$ g of HIT extract plus 20 ng oligo A. Panel B, from left to right: G+A sequencing ladder, no protein added (free), 16 and 24  $\mu$ g of RIN extract, 13 and 20  $\mu$ g of FIB extract, 16 and 32  $\mu$ g of HYB extract, 16  $\mu$ g of RIN extract, 16  $\mu$ g of RIN extract plus 20 ng of oligo A, no protein added (free), 10 and 20  $\mu$ g of HIT extract, 20  $\mu$ g of HIT extract plus 20 ng oligo A, and G+A sequencing ladder. Brackets indicate sequences protected against DNase I digestion; the arrow indicates the hypersensitive site produced by RIN extract.

in a final volume of 20  $\mu$ l containing 20 mM Hepes pH 8, 30% (v/v) glycerol, 0.1 mM EDTA, 50 mM NaCl, 5mM MgCl<sub>2</sub>, 4 mM spermidine, 2 mM dithiothreitol, 1  $\mu$ g salmon sperm DNA (when using DNA fragments) or 1 to 2 mg poly(dI-dC) (when using oligonucleotides) and incubated for 30 minutes at 0°C. Protein-DNA complexes were resolved on a low ionic strength native 6% polyacrylamide gel.

### Methylation interference analysis

100 ng of  $\gamma$ -<sup>32</sup>P-ATP kinased oligonucleotide annealed with the non-labeled complementary strand were partially methylated with dimethylsulfate as described (21), incubated with 80  $\mu$ g of nuclear extract under usual conditions for band shift assays. Following resolution by gel electrophoresis as above, free and bound DNA were eluted from the gel, alkali cleaved, and analysed on a 15% acrylamide/7M urea gel as described (20).

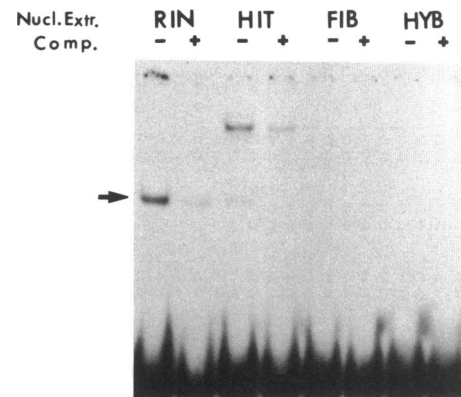
## RESULTS AND DISCUSSION

Gel retardation experiments were performed with three probes a, b and c spanning nucleotides -258 to -169, -168 to +116, and +117 to +241 respectively. Nuclear extracts derived from insulin-producing RIN cells and from fibroblasts and RIN $\times$ fibroblast hybrid cells which do not express the insulin gene, were employed. When incubated with these nuclear extracts, all three probes generated weak retarded bands. In particular, several complexes were formed by the binding of nuclear proteins from RIN extract to probes a and b. This may correspond to the multiple binding sites detected in the enhancer-promoter region of the human insulin gene by a recent study (13). A strong retarded band was generated by incubating probe c with RIN but not fibroblast or RIN $\times$ fibroblast hybrid nuclear proteins (figure 1A). This band was abolished in the presence of excess unlabeled probe (figure 1B), thus indicating the specific binding of a protein. The binding site was further investigated by DNase I footprinting and methylation interference experiments.

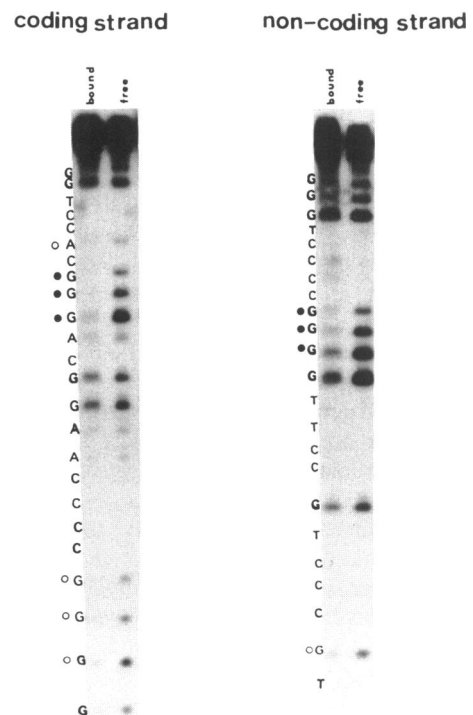
In DNase I footprinting, RIN extract generated on the coding strand a strong protection from +160 to +175, a weaker one from +176 to +189 and a hypersensitive site mapping at +159 (figure 2). On the non coding strand a protection spread from +159 to +175 and a weaker protection spanning nucleotides +176 to +190 were detected. In contrast, the DNase I nucleolytic patterns of the probes were not modified by nuclear extracts from fibroblasts or RIN $\times$ fibroblast hybrid cells. Protections generated by RIN extract were abolished by incubation with oligo A (see table 1), a double-stranded oligonucleotide spanning nucleotide +157 to +189 (figure 2).

The binding of RIN protein(s) to the protected sequences was confirmed by using oligo A as probe in gel shift assays. When oligo A probe was incubated with RIN extract, a single retarded band was generated indicating the formation of a protein-DNA complex. On the contrary, oligo A formed no complex with fibroblast and RIN $\times$ fibroblast hybrid nuclear extracts (figure 3). We also observed that nuclear extracts from the HIT insulinoma cell line generated three retarded bands one of which comigrated with and might therefore be analogous to the RIN complex. The relationship between the three retarded bands generated by HIT extracts has not been investigated; only two of them were strongly reduced by incubation with excess unlabeled oligonucleotide (figure 3) indicating that they correspond to specific DNA

binding. The presence in HIT extract of proteic factor(s) capable to bind to the same sequence as RIN factor(s) was confirmed by footprint experiments. On the coding strand, HIT extract generated a complete protection between nucleotides +161 and +175 and a weaker one spanning nucleotides +176 to +187, very similar to the pattern produced by RIN extract (figure 2A). However, at variance with the RIN induced footprint, no



**Figure 3.** Binding of nuclear extracts to oligonucleotide A. Oligo A corresponds to the +157/+189 sequence of human insulin gene (see table 1) protected by RIN extract in footprint experiments. The binding of nuclear extracts from RIN2A cells (RIN), HIT M2.2.2 cells (HIT), CL1D fibroblasts (FIB) and RIN $\times$  Fibroblast hybrid cells (HYB) to labeled oligo A was tested by gel shift assay in the presence (+) or absence (-) of a 40 fold molar excess of unlabeled oligoA.



**Figure 4.** Methylation interference analysis of the complex formed by RIN nuclear extract with oligo A. After partial methylation, the free and bound DNAs resolved on a band shift assay, were alkali cleaved and analysed on a denaturing gel. Residues interfering with protein binding are marked ●; ○ denotes partial interference.

**Table 1.** Positions and sequences of oligonucleotides used in this study

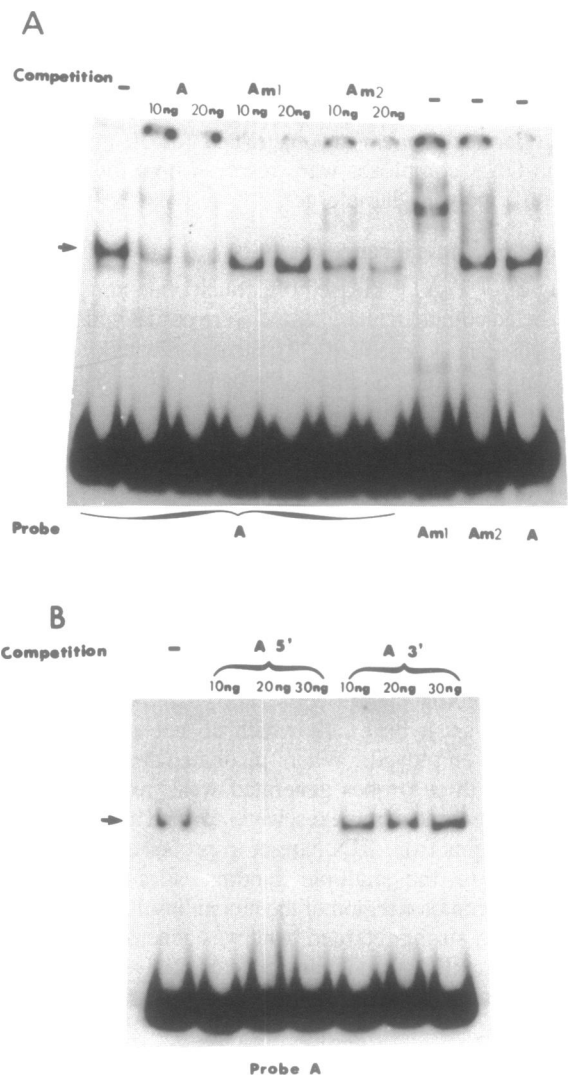
|     |      |   |      |
|-----|------|---|------|
| A   | +157 | G A G C C C A G G G G C C C C A A G G C A G G G C A C C T G G C C   | +189 |
|     |      | C T C G G G T C C C C G G G G T T C C G T C C C G T G G A C C G G   |      |
| A5' | +155 | G T G A G C C C A G G G G C C C C A A G G                           | +175 |
|     |      | C A C T C G G G T C C C C G G G G T T C C                           |      |
| A3' | +174 | G G C A G G G C A C C T G G C C T T                                 | +191 |
|     |      | C C G T C C C G T G G A C C G G A A                                 |      |
| Am1 | +157 | G A G C C C A G G G G C A A C A A G G C A G G G C A C C T G G C C   | +189 |
|     |      | C T C G G G T C C C C G T T G T T C C G T C C C G T G G A C C G G   |      |
| Am2 | +157 | G A G C C C A G G G G C C C C A A G G C A T T T G C A C C T G G C C | +189 |
|     |      | C T C G G G T C C C C G G G G T T C C G T A A C C G T G G A C C G G |      |
| B1  | -241 | T C T G G C C A C C G G G C C C T G G T T A A G A C T C T           | -212 |
|     |      | A G A C C G G T G G C C C G G G G A C C A A T T C T G A G A         |      |
| B2  | -235 | C T T T G G G C C C T G G   | -222 |
|     |      | G A A A C C C G G G G A C C   |      |
| C   |      | C T A A C T G G G C G G A G T T A T G C                             |      |
|     |      | G A T T G A C C C G C C T C A A T A C G                             |      |
| D   | -117 | C A G C T T C A G C C C C T C T G G C C A T                         | -96  |
|     |      | G T C G A A G T C G G G G A G A C C G G T A                         |      |
| E   | -241 | C A G G C C A T C T G G C C C C T                                   | -225 |
|     |      | G T C C G G T A G A C C G G G G A                                   |      |
| Em1 | -241 | C A G G C C A T C T G G A A A C T                                   | -225 |
|     |      | G T C C G G T A G A C C T T T T G A                                 |      |

A, Am1 and Am2 are wild type and mutant versions respectively of the protected sequence in human insulin gene intron I. A5' and A3' correspond to the 5' and 3' parts of oligo A respectively. B1 and B2 are derived from human insulin gene sequence in the Far related region. C is derived from the 21 bp repeat of SV 40 and comprises an SP1 recognition site. D is a GC rich sequence from rat II gene enhancer region. E and Em1 are wild type and mutated versions respectively of rat I Far region. Mutated sequences are boxed. A continuous line overlies Far-related sequences in human and rat derived sequences. ● denotes residues interfering with protein binding from results reported in figure 4.

hypersensitive site at +159 was detected in this case. On the non coding strand, a footprint spanning nucleotides +159 to +175 was observed with a weak protection in the +176/+187 region (figure 2B).

The data presented above indicate that i) RIN extract contains a proteic factor capable to bind mainly in the +160/+175 region of the human insulin gene and ii) similar factors are also present in HIT extract but not in nuclear extracts from non insulin producing cells (fibroblasts and insulin negative RIN×fibroblast hybrid cells).

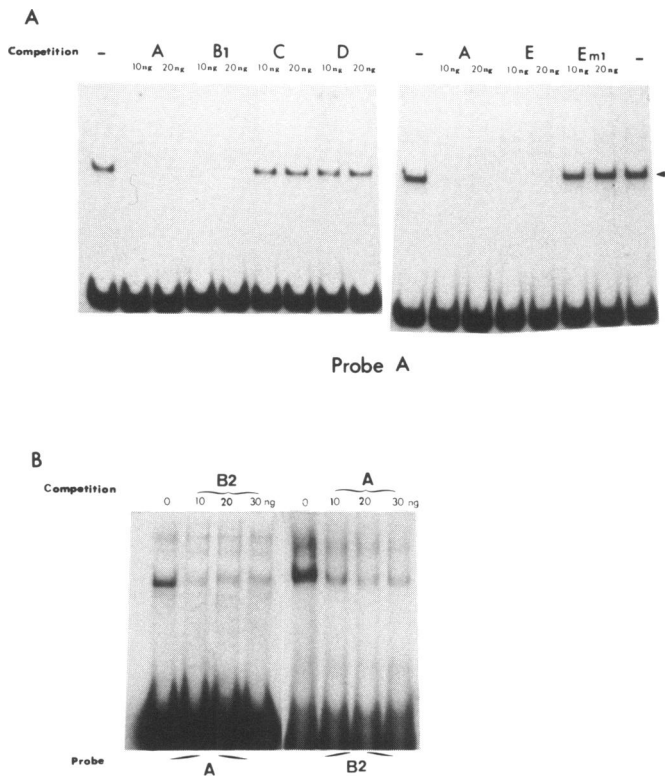
In order to delineate more precisely the protein DNA interactions, the contacts between RIN protein and the protected DNA sequence were analysed by methylation interference experiments. This revealed the binding of the RIN factor to two sequences i.e.: GGGCCC between nucleotides +165 and +170 and GGGCA between nucleotides +178 and +182 (figure 4). Oligonucleotides Am1 and Am2, which are mutated in either sequence (table 1) were used in gel shift assays as competitor or probe. Am1, mutated in the GGGCCC site, failed to bind to the RIN factor, whereas the binding of Am2, mutated in the GGGCA sequence, was not significantly impaired (figure 5A). These results suggested that the GGGCCC (+165/+170) sequence contributes predominantly to the formation of the



**Figure 5.** A/ Binding of oligo A and mutants Am1 and Am2 to RIN extract. The complexes formed by RIN proteins with probes A, Am1 and Am2 were analysed by gel mobility shift assay. The competitor added in each case is indicated at the top of the figure. B/Inhibition of the binding of oligo A to RIN extract by oligo A5'. The complex formed by RIN proteins with probe A in the presence of various amounts of oligo A5' and A3' was analysed by gel mobility shift assay. The competitor added is indicated at the top of the figure.

RIN/oligo A complex detected in gel shift assay whereas the GGGCA (+178/+182) sequence might be accessory. This was confirmed by experiments using oligo A5' (comprising GGGCCC) and oligo A3' (comprising GGGCA) (table 1) which showed that the binding of oligo A to the RIN factor can be efficiently competed for by A5' but not A3' (figure 5B). Finally, taking into account the close similarity of the two sequences detected in methylation interference experiments (GGGCCC vs GGGCA(C)), one might expect that the same factor can bind to both of them but possibly with different affinities. This would be compatible with the strong and weak protections observed in footprint experiments in the +159/+175 and +176/+190 regions respectively.

We have detected the binding of a proteic factor from rat insulinoma cells to a GGGCCC core sequence located at position



**Figure 6.** A/ Competition of GC rich oligonucleotides for the binding of oligo A to RIN extract. The complex formed by RIN proteins with probe A was analysed by gel mobility shift assay. The competitor added in each case is indicated at the top of the figure. B/ Competitive binding of oligos A and B2 to RIN extract. The complexes formed by RIN factor with oligos A and B2 were analysed by gel mobility shift assay. Both oligos were used as competitors to the binding of each other, as indicated at the top of the figure.

+165 to +170 i. e in the intron I region of the human insulin gene. In order to assess the role of this intronic sequence, experiments are in progress to analyze the effects on gene expression of mutating this GGGCCC motif. However, the absence of obviously conserved GGGCCC-like sequence in introns of rat insulin genes led us to look for homologous sequences elsewhere in insulin genes. In particular, as shown below, we detected motifs related to GGGCCC (boxed) in the vicinity of the Far-related sequence (underlined) in rat, human and mouse insulin genes:

rat I (-243/-217) ATCAGGCCATCTGGGCCCTTGTTAATA  
 rat II (-233/-207) ATCAGGCCACCCAGGAGCCCTCTTAA  
 human (-241/-215) GTCTGGCCACCCGGGCCCTGGTTAAGA  
 mouse I (-243/-217) ATCAGGCCATCTGGTCCTTATTAAGA  
 mouse II (-243/-217) ATCAGGCCATCAGGGCCCTTGTTAAG

That these sequences may interact specifically with the GGGCCC binding factor, was demonstrated by gel retardation experiments: 1/Oligo E, a rat I gene derived oligonucleotide spanning nucleotides -241 to -225 and comprising the Far sequence plus the contiguous GGCCC motif (table 1) completely inhibited the binding to RIN extract of oligo A (figure 6A). By contrast, oligo Em1 in which GGCCC is mutated to GGAAA had no effect. 2/A similar inhibition was induced by oligo B1, spanning nucleotides -241 to -212 of human insulin gene and comprising

the GGGCCCC motif (table 1). 3/ Oligo B2, a human gene derived oligonucleotide (-235 to -222) comprising a GGGCCCC sequence but devoid of the 5' part of the Far-like sequence and mutated in its 3' part (table 1), was capable to bind to the same RIN factor as oligo A, as demonstrated by comigration of the complexes and cross competition between oligonucleotides A and B2 for complex formation (figure 6 B). 4/The binding of oligo A to the RIN factor, was not competed for by unrelated GC rich oligonucleotides (figure 6A) such as oligo C which contains an SP1 motif capable to bind to various nuclear extracts (not shown) or oligo D spanning nucleotides -117 to -96 of rat insulin II gene (table 1). Therefore, the GGGCCC binding factor appears to interact with the sequences immediately downstream the Far sequence in rat I gene and the Far homologous GCII motif in human insulin gene. Proteins binding the Nir/Far related elements have been described and characterized (5-9,13, 22). Recently, regulatory proteins binding to the Far linked A+T rich (FLAT) element located further downstream have also been characterized (23, 24). The GGGCCC binding protein detected here seems clearly distinct from proteins binding to Far related or FLAT elements since i) the Far sequence alone does not compete for the binding of oligo A to the GGGCCC binding factor (figure 6A) and ii) the binding site has no relatedness with the A+T rich element.

Future studies will have to define the physiologically significant target site(s) of the GGGCCC binding factor and its potential role in insulin gene regulation.

## ACKNOWLEDGEMENTS

Our grateful thanks to T.Edlund for providing us with HIT cells. We also thank M.Raymondjean for fruitful discussions and critical reading of the manuscript. This work was supported in part by the Association pour la Recherche sur le Cancer, the Ligue Nationale contre le Cancer and the Fondation pour la Recherche Médicale.

## REFERENCES

- Walker M. D., Edlund T., Boulet A. M. and Rutter W.J. (1983) Nature 306, 557-561.
- Edlund T., Walker M.D., Barr P.J. and Rutter W.J. (1985) Science 230, 912-913.
- Karlsson O., Edlund T., Moss J. B., Rutter W. J. and Walker M.D. (1987) Proc. Natl. Acad. Sci. USA 84, 8819-8829.
- Karlsson O., Walker M. D., Rutter W. J. and Edlund T. (1989) Mol. Cell Biol. 9, 823-827.
- Ohlsson H. and Edlund T. (1986) Cell 45, 35-44.
- Ohlsson H., Karlsson O. and Edlund T. (1988) Proc. Natl. Acad. Sci. USA 85, 4228-4231.
- Nelson C., Shen L. P., Meister A., Fodor E. and Rutter W. J. (1990) Genes and Dev. 4, 1035-1043.
- Walker M. D., Pork C. W., Rosen A. and Aronheim A. (1990). Nucl. Acids Res. 18, 1159-1166.
- Shibasaki Y., Sakura H., Takaku F. and Kasuga M. (1990). Biochem. Biophys. Res. Commun. 170, 314- 321.
- Whelan J., Poon D., Weil P.A. and Stein R. (1989) Mol. Cell. Biol. 9, 3253-3259.
- Whelan J., Cordle S.R., Henderson E., Weil P. A. and Stein R. (1990) Mol. Cell Biol. 10, 1564-1572.
- Crowe D. T. and Tsai M. J. (1989) Mol. Cell Biol. 9, 1784-1789.
- Boam D. S. M., Clark A. R. and Docherty K. (1990) J. Biol. Chem. 265, 8285-8296.
- Besnard C., Lores P., Daegelen D., and Jami J. (1989) Exp. Cell Res. 185, 101-108.

15. Besnard C., Monthieux E., Lores P., Jami J. and Daegelen D. (1991) *J. Cell Physiol.* 146, 349–355.
16. Gazdar A. F., Chick W. L., Oie H. K., Sims H. L., King D. L., Weir G. C. and Lauris V. (1980) *Proc. Natl. Acad. Sci.* 77, 3519–3523.
17. Kit S., Dubbs D. R., Piekarski L. J. and Hsu T. C. (1963) *Exp. Cell Res.* 31, 297–312.
18. Dignam J. D., Lebowitz R. M. and Roeder R. G. (1983) *Nuc. Acids Res.* 9, 1351–1357.
19. Garner M. M. and Revzin A. (1981) *Nuc. Acids Res.* 9, 1351–1357.
20. Raymondjean M., Cereghini S. and Yaniv M. (1988) *Proc. Natl. Acad. Sci. USA* 85, 757–761.
21. Siebenlist U.L. and Gilbert W. (1980) *Proc. Natl. Acad. Sci. USA* 77, 122–126.
22. German M.S., Moss L.G., Wang J., and Rutter W.J. (1992) *Mol. Cell. Biol.*, 12, 1777–1788.
23. Karlsson O., Thor S., Norberg T., Ohlsson H. and Edlund T. (1990) *Nature* 344, 879–882.
24. German M. S., Wang J., Chadwick R. B., and Rutter W. J. (1992) *Genes & Dev.* 6, 2165–2176.