The GC Box as a Silencer

Jacek M. Jankowski and Gordon H. Dixon

Received November 23, 1987

KEY WORDS: gene expression; GC box; negative regulation.

A DNA control sequence $^\text{GGGCGGCGC}_\text{TATA}$, or the "GC" box, has been described in the promoter regions upstream of a number of eukaryotic genes transcribed by polymerase II (for review, see Dynan, W. S. and Tjian, R., Nature 316:774, 1985). The "GC" box can occur in single or multiple copies and is the binding site for a protein factor, Sp1, which activates initiation of transcription. We have observed in the rainbow trout protamine gene 3' to the TATA box, three "GC" boxes spaced at 80 bp intervals. The first is 5' to the cap site and possesses the ability to "silence" transcription from the protamine promoter in constructs linking this promoter to the bacterial chloramphenicol acetyl transferase (CAT) coding sequence following transfection to COS-1 cells. A model is proposed to account for the silencing of the protamine gene in all tissues except developing sperm cells.

INTRODUCTION

Rainbow trout (Salmo gairdnerii) protamines are small, strongly basic nuclear proteins (1). Their expression occurs in male germ cells during late stages of sperm development and they eventually replace the somatic histones completely with the result that sperm chromatin is profoundly restructured and condensed into a genetically inactive non-nucleosomal form (2). It seems evident, therefore, that control of this set of genes must be very precise so that protamine gene expression is permitted only in developing spermatocytes and spermatids.

Several unsuccessful attempts have been made to achieve expression of the cloned...
protamine gene *in vivo* by transfection of the cloned gene into mouse L cells (3). Substantial expression of protamine genes as a protamine polypeptides following transfection into cells in tissue culture would be predicted, however, if quantitatively significant, to lead to restructuring and repression of the host chromatin with resulting “killing” of the cells. In fact, it was observed after selection, using co-transfection with the HSV-1 thymidine kinase gene, that the only cells that survived appeared to be those in which the protamine gene was inactivated, often as a result of rearrangement or deletion (ibid.). Recent experiments by Peschon *et al.* (4) in which mouse protamine genes have been inserted into mouse embryos to produce mice transgenic for a protamine gene, have shown that this inserted gene is expressed only in male germ cells and thus its expression must be strongly suppressed in other tissues.

In order to study the regulatory signals that control this gene *in vivo*, we have replaced the potentially “toxic” protamine coding sequence with that of the prokaryotic chloramphenicol acetyltransferase (CAT) gene (5) (kindly sent to us by Dr B. Howard, NCI, Bethesda). The CAT gene product has been shown to be non-toxic (ibid.) and is readily assayed at low concentration. To avoid any possible rearrangements during the integration process, these constructs were inserted into an SV40 virus vector. As a recipient, COS-1 cells (6) (a gift from Dr D. I. Hoar, Department of Medical Biochemistry, The University of Calgary) were used so as to allow the plasmid constructs to be maintained in an episomal, unintegrated form.

**MATERIALS AND METHODS**

**Plasmids**

Plasmid pTkOcat (Fig. 2) contains a 1782 bp fragment of pBR322 (including the origin of replication and the ampicillin resistance gene) and the chloramphenicol acetyltransferase (CAT) gene (Hind III—BamHI fragment from plasmid pSV2cat [5] which also contains SV40 polyadenylation signals). The CAT gene is placed under the control of the herpes simplex-1 thymidine kinase (tk) gene promoter (BamHI-BglII fragment from the pX14T, a gift from Dr D. I. Hoar). Plasmid pTkOcat also contains the SV40 ori in Fig. 2, to allow replication in the COS-1 cells.

The plasmids pP5cat, pP7cat, pP9cat are derivatives of pTkOcat in which the tk promoter was replaced with truncated protamine promoters (fragments from the BglII site [compare Fig. 1 for the protamine sequence] down to the positions shown in Fig. 2).

The plasmids pTk7cat, pTk9cat are derivatives of pTkOcat in which the fragments illustrated (Fig. 3) of the protamine gene originally located between the TATA box and structural gene of protamine, are inserted between the tk promoter and the CAT gene.

Plasmids pTk6cat and pTk6Mcat (Fig. 4 and Fig. 5 respectively) are derivatives of the pTk7cat which contain altered GC boxes. The GC box in the plasmid pTk6cat was lengthened by single GC base pair. In the plasmid pTk6Mcat the GC box was replaced by another GC-rich sequence, GC...GC, of the same base composition but in which the sequence is randomized.
Cell Transfection

COS-1 cells were transfected as described (7) except that 1% dimethyl sulfoxide (DMSO) was included in the medium and 48 h later were harvested and chloramphenicol acetyltransferase activity was determined as described (8). Protein concentration was determined by the Bradford method (9). In order to measure the copy number of the pTk6cat plasmid in comparison to pTkOcat, cells (transfected with the appropriate plasmid) were subjected to the Hirt extraction protocol as described elsewhere (10) and the amount of DNA was measured by dot-blot-hybridization technique using Hybond nylon membrane (Amersham).

RESULTS AND DISCUSSION

The constructs, pP9cat, pP7cat and pP5cat were compared for their promoter strength in transfection experiments. As a reference we used a plasmid in which the

Fig. 1. Nucleotide Sequences of the Bgl II-Bam HI Restriction Fragment of Trout DNA Containing a Protamine Gene (States et al., 1982). The TATA and GC sequences are boxed and polyadenylation signals are underlined. Horizontal arrows indicate the cap sites of the protamine gene (20,21).
Fig. 2. Expression of CAT Activity in COS-1 Cells. (a) Plasmids. Plasmid pTk0cat contains a 1782 bp fragment of pBR322 (including the origin of replication and the ampicillin resistance gene) and the chloramphenicol acetyl transferase (CAT) gene (HindIII-BamHI fragment from plasmid pSV2cat (5) which also contains SV40 polyadenylation signals). The cat gene is placed under the control of the herpes simplex-1 thymidine kinase (tk) gene promoter (BamHI-BglII fragment from the pX1 plasmid). Plasmid pTk0cat also contains the SV40 origin of replication (SphI-HindIII fragment), which is denoted SV40 ori, in the figure to allow replication in the COS-1 cells. The plasmids pP5cat, pP7cat, pP9cat are derivatives of pTk0cat in which the tk promoter was replaced with truncated protamine promoters (fragments from the BglII site [compare Fig. 1 for the protamine sequence] down to the positions shown in the figure). The TATA box is underlined, the vertical arrow indicates the “cap” site and the GC box and CAAT sequences are boxed. (b) Cell Transfection. COS-1 cells were transfected as described (in Materials and Methods) and 48 h later were harvested and chloramphenicol acetyl transferase activity was determined as described (8). Protein concentration was determined by the Bradford method (9). Each data point presented here represents the mean of three experiments.

CAT gene was placed under the control of the HSV-1 thymidine kinase (tk) gene. In the constructs pP5cat, pP7cat and pP9cat, the control tk promoter is replaced by complete protamine promoters as follows: pP9cat (see Fig. 2) contains the fragment from pos. 1 to 429 (compare Fig. 1 for the protamine sequence), pP7cat contains the fragment from pos. 1 to 420 while pP5cat contains the fragment from pos. 1 to 397.

As can be seen in Fig. 2, when the “cap” site of the protamine gene was deleted (in plasmid pP5cat), CAT activity was barely detectable. These results are in good agreement with the data of Tokunaga et al. (1983) who used COS cells in transfection experiments and found the “cap” site region to be an important domain required for faithful and efficient initiation of transcription of the silk fibroin gene. On the other hand, when the cap site was preserved (in the pP7cat and pP9cat constructs), CAT activity was substantial (in comparison to pP5cat) but not as high as with the control tk promoter (pTk0cat). It might be argued that the low activity in pP5cat could also be due to the simultaneous deletion of the downstream CAAT box with the “cap” site. However, in plasmids pP7cat and pP9cat, in which the “cap” site is preserved, the presence or absence of the CAAT element has no effect so its influence on transcription must be neutral in this context. It has been our experience (12), consistent with that of
others (13), that the tk promoter is a fairly weak one compared to the natural protamine promoter when both were assayed in vitro under identical conditions (12). In contrast, with the series of CAT constructs (whose activity is illustrated in Fig. 2) it appeared that the tk promoter, whose activity is lower than that of the natural protamine promoter in vitro, was much more effective as a promoter when linked to the CAT gene and expressed in vivo in COS cells.

A possible explanation of this situation is that some characteristic of the protamine gene promoter may suppress CAT transcription in vivo. We had previously noticed that transcription from the protamine promoter in vitro was strongly stimulated when a physically separate tk promoter sequence was included in the reaction mixture (ibid.). A reasonable explanation of this observation would be that this extra tk promoter might bind, in a competitive fashion, an inhibitory factor or factors and thus release constraints on transcription from the stronger protamine promoter.

In the pP7cat and pP9cat constructs in the experiment shown in Fig. 2, two important elements are still present 3' to the TATA box. The first is a GC-rich hexamer, GGGCGG (in Fig. 1 it is shown as the complementary CCGCCC sequence). This sequence, which is seen in a number of gene promoters, is termed a GC box and binds at least one transcription factor, Sp1 (14). The second element is a CAAT box (in pP9cat) that can bind two transcription factors, the CAAT transcription factor (CTF) (ibid.) and/or CAAT binding protein (CBP) (17). It should be noted, however, that the CAAT sequence is normally present further upstream at-70 to-80 to the “cap” site.

As shown in Fig. 2, the transcription initiation or “cap” site of the protamine promoter in pP9cat is located between these two elements. It is possible, therefore, that one or both the GC box and CAAT binding factors, when bound to DNA, could cover enough flanking sequences (15) to block access to the “cap” site and thus inhibit the initiation of transcription. An analogous situation is seen when the large antigen T suppresses expression of the early genes in the SV40 genome by preventing progression of the polymerase (in domain T1) and/or initiation of transcription (in domain T2) [see ref. 16 and references therein].

In order to examine this hypothesis, starting with the pTkOcat construct (Fig. 3), we inserted the two sequence elements, GGGCGG and CAAT, between the tk promoter and CAT gene to determine whether, these elements when inserted 3' to the “cap” site, could cause a silencing of the expression of CAT activity under the control of the tk promoter. In this approach, both the precise distance and any possible secondary structures between the TATA box and the “cap” site of the tk promoter have been preserved. The inclusion of the complete tk promoter as a reference also allows for better assessment of the role of these additional regulatory elements. Thus, the region of the protamine gene from 392 to 429 (for the protamine sequence see Fig. 1) containing both the GC box and the CAAT sequence is inserted downstream of the cap site of the complete tk promoter in pTk9cat and the region from 392 to 420 containing only the GC box in pTk7cat. As can be seen (Fig. 3), in the presence of the tk promoter, which was shown to be weaker in vitro than the protamine promoter (12), the difference between the silencing effect of the GC box alone (pTk7cat) and GC and CAAT box together (pTk9cat) is more pronounced than in the constructs pP7cat and pP9cat containing only the protamine promoter (Fig. 2).
Fig. 3. Effect of the Insertion of a Protamine 5' Untranslated Region Insert Containing Additional GC and CAAT Boxes Downstream from the TATA Box on the Expressions of CAT Activity. Plasmids: The plasmids pTk7cat, pTk9cat are derivatives of pTk0cat (see Fig. 1) in which the illustrated fragments of the protamine gene originally located between the TATA box and structural gene of protamine, are inserted between the tk promoter and the CAT gene. For other details see Fig. 2. Data presented here represent results from seven experiments.

The GC box between the TATA box and “cap” site of the protamine gene corresponds only to the core, CCGCCC, of the consensus sequence (14) and it is, therefore, considered likely that its effect might be weaker than that of the full GC box (see insert in Fig. 4). We therefore added an additional C residue to the sequence (this would correspond to an extra G on the opposite strand) in pTk6cat and again tested its silencing effect on expression of the CAT gene (Fig. 4).

In Fig. 4, pTk6cat shows the predicted significant decrease in the expression of the CAT activity when the GC box, lengthened by an additional G, was inserted downstream from the TATA box. If instead a GC-rich hexanucleotide sequence of the same composition and length but a randomized sequence was inserted between the promoter and the CAT gene in pTk6Mcat (Fig. 5), chloramphenicol acetyltransferase activity is not suppressed as compared with the control (in this particular experiment the CAT activity in the case of pTk6Mcat is even higher than the control construct).

We have seen similar results in experiments in which the prokaryotic lacZ gene was inserted into the plasmid constructs as a reference in which beta-galactosidase activity was expressed and measured for normalizing the copy number of the transfected plasmid (data not shown). These experiments further support the view that a GC box in this position in a gene can act as an effective silencer of its expression.
The GC Box as a Silencer

Fig. 4. The Effect of the Lengthening of the GC Box by a Single G-C Base Pair on the Expression of CAT Activity Measured as Described in Fig. 1. Plasmid pTk7cat contains an insert GACCACCGCCCGCTCTAAACAIIIIAICC between the Tk promoter and the CAT box of pTk0cat while pTk6cat has the lengthened CG box sequence GACCACCGCCCG (underlined). These plasmids were transfected and the CAT activity measured as described in Materials and Methods. Data presented here represent results from seven experiments. In order to measure the copy number of the pTk6cat plasmid in comparison to pTk0cat, cells (transfected with appropriate plasmid) were subjected to the Hirt extraction protocol as described elsewhere (10) and the amount of DNA was measured by dot-blot-hybridization technique using Hybond nylon membrane (Amersham). Copy numbers for pTk0cat are $5.5 \times 10^5$ copies per cell and for pTk6cat $5.2 \times 10^5$ copies per cell.

The observation of negative or positive regulation by the Spl factor, depending on its position of binding in relation to the TATA box, is consistent with the mechanism of action of the T antigen in SV40 (for review see ref. 17) or the action of lambda repressor which can also either shut off or activate different genes (cro and cI respectively) (for review see ref. 18). Our observation of down regulation by the GC box is also consistent with findings that this sequence reduces the basal level of rat metallothionein gene expression (H. R. Herschman, Department of Biological Chemistry, UCLA, personal communication).

One would, however, ask, if this silencing effect is effective enough to keep the protamine genes quiescent in all but germ cells, why should there be any expression of the CAT activity at all in COS-1 cells? To answer this question, it should be realized that in these cells, plasmids which contain an SV40 origin of replication can be amplified up to about 500,000 copies per cell. Thus the number of copies of the gene (on this episome) could easily titrate out all available potentially-inhibitory transcription factors such as Spl and thus allow the CAT gene to be expressed.
Fig. 5. The Effect of Insertion of the Random GC-Rich Sequence on Expression of the CAT Activity. Plasmid pTk6Mcat contains a randomized CG sequence (of the same base composition) GACCA GCGCGGC and its CAT activity after transfection is compared with pTk6cat containing the sequence GACCAAGCGCCCG. For other details see the legend in Fig. 3.

By inspection of the protamine gene sequence (Fig. 1), it may be seen that two other GC boxes are present in the coding region downstream from the TATA box (compare also Fig. 6a) in addition to the one between the TATA box and the “cap” site (pos. 398–403 in Fig. 1). The other two are within the coding sequence for protamine.

a.

![Diagram of mRNA and GC boxes](image)

b.

![Diagram of TATA box and GC boxes in nucleosome](image)

c.

![Diagram of possible interactions with Sp1 factors](image)

Fig. 6. (a) The arrangement of the GC boxes in the trout protamine gene. (b) A model for their possible nucleosome location in a nucleosome. (c) Possible interactions with Sp1 factors.
The GC Box as a Silencer

(pos. 478–483 and 562–567 in Fig. 1). We note that the distance between the first (discussed above) and the second is equal to the distance between the second and the third and is close to 80 bp. As shown in Fig. 6b, this spacing could imply that all three GC boxes might be located in a domain occupied by a single nucleosome and thus up to three Sp1 or Sp1-like factors could potentially interact with this nucleosome, possibly in a co-operative manner. In this interaction other factors such as CTF (compare 15) could also be involved. A stable structure of this kind could very effectively silence the protamine gene and thus protect it from premature expression in actively transcribing somatic cells. A prediction of the model would be that only when the level of the Sp1 or Sp1-like factor declined markedly, or Sp1 was inactivated in some way, as is likely when the expression of the majority of the somatic genes is attenuated during spermatogenesis (Levy-Wilson and Dixon, 1977) and does not need the normal positive regulation of Sp1, could the protamine genes become activated and their products, the sperm protamines, restructure and completely repress gene expression in mature spermatozoa.

ACKNOWLEDGEMENTS

We thank Drs K. Iatrou, J. McGhee and G. A. Schultz of this Department for critical reading of the manuscript and providing helpful suggestions. We gratefully acknowledge support by Medical Research Council of Canada (operating grant to GHD) and Alberta Heritage Foundation for Medical Research (fellowship to JMJ). We also wish to thank Eva Walcyk and Manfred Herfort for skilled technical assistance and Ann M. Camponi for expert typing of the manuscript.

REFERENCES