IDENTIFICATION OF DNA SEQUENCES SPECIFIC FOR 5'-FLANKING REGIONS OF GLUCOCORTICOID-REGULATED GENES USING COMPUTER ANALYSIS

T. I. Merkulova, V. V. Solov'ev,
N. A. Kolchanov, S. Yu. Plisov,
and R. I. Salganik*

UDC 612.6.05:575.113.12]-06:612.453.018

KEY WORDS: computer analysis; glucocorticoid-controlled genes; consensus

Although the structure of many genes of higher organisms has been studied in detail in recent years the mechanisms controlling their expression remain unexplained. Controllers of gene expression include, in particular, steroid hormones, complexes of which with specific receptor proteins interact with 5'-regions of DNA close to the genes controlled by these hormones. This interaction leads to a change in the intensity of transcription of these genes [3, 4, 6].

The group of genes controlled by one hormone must evidently have similar nucleotide sequences, identified by the given hormone-receptor complex, in the 5'-flanking regions, where the promotors are located. Sequences homologous with the TGTTCT hexanucleotide (consensus) have recently been discovered in regions of the human metallothioneine II A gene and in the long terminal repeats (LTR) of DNA of mouse mammary tumor provirus (MMTV), protected by glucocorticoid-receptor complexes (GlRC) against deoxyribonuclease I [3, 6, 7]. The authors cited suggested that the hexanucleotide they discovered participates in the recognition and binding of GlRC, which activate the expression of these genes [3, 7]. Other homologs of this hexanucleotide have been found in the 5'-flanking region of the glucocorticoid-controlled tyrosine aminotransferase gene [8].

However, in the work so far published [3, 7, 8] the number of glucocorticoid-controlled genes does not exceed three, and they have not been compared with genes not controlled by glucocorticoids. This state of affairs makes the above hypothesis significantly less convincing.

To identify specific consensuses, it was decided to investigate all glucocorticoid-controlled genes whose nucleotide sequences have been decoded up to the present and to compare them with genes not controlled by glucocorticoids, and the investigation described below was conducted for this purpose.

EXPERIMENTAL METHOD

Experiments were carried out by the method of context computer analysis described previously, by which it is possible to find similar nucleotide sequences of different lengths and degrees of homology in extensive regions of DNA or RNA [1]. The 5'-flanking regions of 11 genes controlled by glucocorticoids and of 14 genes not controlled by these hormones, ranging in length from 1 to 400 base pairs (b.p.), were investigated. The nucleotide sequences of the genes and their characteristics were taken from [5].

EXPERIMENTAL RESULTS

Hexanucleotides homologous to the TGTTCT consensus (with replacement of not more than one base) were found in regions from −400 to the point of initiation of transcription in all glucocorticoid-controlled genes studied, and they numbered from one to six copies (Table 1). They were found also in the corresponding regions of 10 of 14 genes not controlled by glucocorticoids. The number of these consensuses in them varied from two to four. The expected number of

*Corresponding Member, Academy of Sciences of the USSR.

Laboratory of Molecular Genetics and Laboratory of Population Genetics, Institute of Cytology and Genetics, Siberian Branch, Academy of Sciences of the USSR, Novosibirsk. Translated from Byulleten' Eksperimental'noi Biologii i Meditsiny, Vol. 101, No. 4, pp. 466-468, April, 1986. Original article submitted May 21, 1985.

 0007-4888/86/0004-0502$12.50

TABLE 1. Location of TGTTCT Consensus in 5'-Flanking Regions of Glucocorticoid-Controlled Genes

Genes	Nucleotide sequence	Distance to point or initiation of transcription, b.p.
Human ACTH-β-LPH	CGTTCT, AGTTCT, TGTGCT	—267, —232, —78
Mouse ACTH	AGTTCT, AGTTCT, TGTGCT	—261, —217, —173
Bovine ACTH	TGCTCT	—176*
Human growth hormone (GH)	TTTTCT	—228
Rat GH	TGGTCT	—253
Human metallothioneine (MT) II A	TGTCCT	—252
Mouse MT	AGTTCT, TGTTCC	—298, —232†
Rat tyrosine aminotransferase	TGTACT, TTTTCT, AGTTCT	—322, —246, —79
Rat tryptophan pyrrolase	TGTTCT, AGTTCT	—321, —101
Rabbit uteroglobin	TGGTCT, TCTTCT	—305, —59
LTR MMTV	TGTTGT, TATTCT, TGTTCT	—400, —269, —173
	TAGTCT, TGTTCT, TGTTCT	—112, —96, —81

Legend. *) 203 b.p. decoded, †) 300 b.p. decoded upward from transcription initiation point. In all other cases not less than 400 b.p. decoded above transcription initiation point.

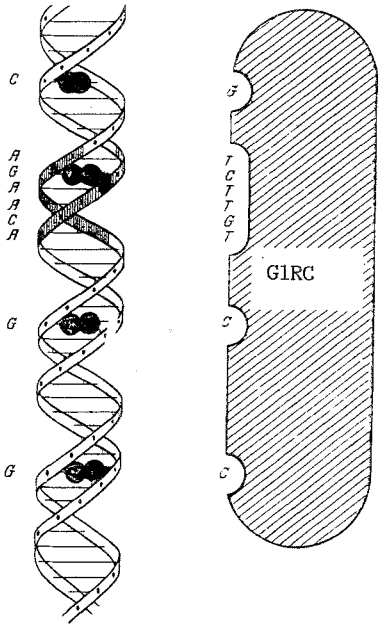

Fig. 1. Scheme showing hypothetical G1RC binding sites in the double helical DNA molecule.

these consensuses for random sequences 400 b.p. in length was about 0.1. The fact that many more of them were found than in the random case indicates the probable functional role of these consensuses. Meanwhile, it is quite evident from the results that they are not specific for glucocorticoid-controlled genes.

However, on analysis of the TGTTCT consensus and regions flanking it, in all 11 glucocorticoid-controlled genes, we found four cytosine residues regularly arranged. One of them (the 2nd or 3rd) definitely belongs to the TGTTCT consensus itself. It is important to note that three of these cytosine residues were located 8-10 b.p. apart, a distance which approximately corresponds to one turn of the DNA double helix. The 4th cytosine residue is located 6 b.p. from the 3rd. In some cases one of the four cytosine residues can be replaced by a thymine residue. In nine of the 10 genes not controlled by glucocorticoids, and containing a consensus in the 5'-flanking region, which we investigated, no such regularity could be found in the arrangement of the cytosine residues in the region of this consensus.

It is evident that the characteristic alternation of cytosine residues in the region of the TGTTCT consensus is specific only for glucocorticoid-controlled genes. It is very important to note that the position of three of these cytosine residues in the linear DNA molecule is such that they can be arranged along one side of the DNA double helix. A guanine residue, complementary to the 4th conservative cytosine residue also is found on this same side (Fig. 1).

Under these circumstances the TGTTCT consensus, together with the cytosine residues and guanosine residue surrounding it, forms a specific compact region for glucocorticoid-controlled genes in the DNA double helix which may be the recognition site for GlRC. This hypothesis is supported also by the fact that DNA regions protected by GlRC from DNase I in genes MT IIA and MMTV [3, 6, 7], coincide approximately in size with the structure 26-29 b.p. in length which we identified (Fig. 1).

In some glucocorticoid-controlled genes we also found a structure which differs from that described above only in that it is rotated through 180°. Since we know that DNA regions responsible for glucocorticoid control can function independently of their orientation [2], this suggests that these structures also form binding sites for GlRC.

It can thus be tentatively suggested that the specific binding sites for GlRC consist of a TCTTCT hexanucleotide nucleus, flanked by two cytosine and one guanine base, arranged along one side of the DNA double helix. It must be pointed out that the complementary nucleotide sequence may also serve as GlRC recognition sites, or a double-helical fragment formed by the $\begin{cases} \text{TGTTCT} \\ \text{ACAAGA} \end{cases}$ consensus and by one $\begin{cases} \text{G} \\ \text{C} \end{cases}$ and two $\begin{cases} \text{C} \\ \text{G} \end{cases}$ pairs flanking it.

LITERATURE CITED

1. N. A. Kolchanov, V. V. Solov'ev, and A. A. Zharkikh, Dokl. Akad. Nauk SSSR, 273, 741 (1983).
2. V. L. Chandler, B. A. Maler, and K. R. Yamamoto, Cell, 33, 489 (1983).
3. M. Karin, A. Haslinger, H. Holtgree, et al., Nature, 308, 513 (1984).
4. E. R. Mulwihill, J. P. LePennec, and P. Chambon, Cell, 24, 621 (1982).
5. Nucleotide Sequences 1984, Oxford (1984).
6. F. Payvar, D. DeFranko, G. L. Firestone, et al., Cell, 35, 381 (1983).
7. C. Scheidereit, S. Geisse, H. M. Westphal, and M. Beato, Nature, 304, 749 (1983).
8. T. Shinomiya, G. Scherer, W. Schmid, et al., Proc. Natl. Acad. Sci. USA, 81, 1346 (1984).