# A GC box in the bidirectional promoter is essential for expression of the human dihydrofolate reductase and mismatch repair protein 1 genes

Hiroyuki Fujii[a], Eiji Shinya[a,b] and Takashi Shimada[a,b]

[a]*Clinical Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA* and [b]*Department of Biochemistry and Molecular Biology, Nippon Medical School, 1-1-5 Sendagi, Bunkyo-ku, Tokyo 113, Japan*

The human dihydrofolate reductase and mismatch repair protein 1 genes are organized in a head-to-head configuration separated by an 88 base-pair segment and directed by a bidirectional promoter. In vivo transient assays of the site directed mutant promoters using firefly luciferase as a reporter showed that an AT-rich sequence, ACAAATA, in the GC-rich promoter sequence is not required for transcription. However, two out of four GC boxes were shown to function as bidirectional positive regulatory elements. Among them, a GC box at the midpoint of the region between the two initiation sites is essential for supporting minimal bidirectional activity.

Dihydrofolate reductase; Bidirectional promoter; Luciferase; Transcription control; GC box; TATA box

## 1. INTRODUCTION

We have recently isolated and characterized cDNA clones derived from transcripts initiated 89 base pairs upstream from the DHFR gene but transcribed from the opposite strand [1]. The divergently transcribed gene encodes a protein highly homologous to a bacterial DNA mismatch repair protein, MutS [2], and therefore, we have named it the human mismatch repair protein 1 (MRP1). A homologous gene also has been identified in the region upstream from the mouse DHFR gene [3].

We have previously shown that as small as a 165 bp DNA fragment from −111 to +54 relative to the DHFR initiation site has bidirectional promoter activity [4]. This sequence is characterized by richness in guanosine and cytosine, presence of four GC boxes, and lack of the typical TATA or CAAT boxes. Instead, this promoter contains an AT-rich sequence ACAAATA at 29 bp upstream from the DHFR start site [5]. Previous studies, using in vitro transcription assays, have shown that the multiple GC boxes and the sequence surrounding the initiation site play important roles in transcription of the DHFR gene [6–9]. However, the promoter elements for MRP1 expression have not been studied so far, primarily because MRP1 transcripts could not be detected by in vitro transcription assays [10]. We have recently demonstrated that an in vivo assay based on

transient expression of heterologous reporter genes can be used for detection of bidirectional activity of the 165 bp promoter sequence [4]. In this study, using the firefly luciferase gene as a reporter, we examined the role of the AT-rich sequence and the multiple GC boxes in the control of both DHFR and MRP1 expression.

## 2. MATERIALS AND METHODS

### 2.1. *Plasmid construction*

A basic plasmid $\alpha\beta8X$ was constructed by insertion of two polyadenylation signals from the human globin genes into the plasmid 8X [11], to minimize interference between sense and antisense transcripts derived from the bidirectional promoter in a circular plasmid. The 114 bp wild type or mutant bidirectional promoter sequences from −99 to +15 relative to the DHFR initiation site were chemically synthesized (BP). These fragments were inserted between the *Pst*I and *Hind*III sites of $\alpha\beta8X$, yielding BP/$\alpha\beta8X$. The DNA fragment containing the firefly luciferase coding sequence and the SV40 polyadenylation signal (Luc [12]) was inserted into either the *Pst*I and *Bam*HI sites (MPR1 side; BP-M-Luc) or the *Hind*III and *Xho*I sites (DHFR side; BP-D-Luc) (Fig. 1).

### 2.2. *In vivo transient assay*

Ten $\mu$g of luciferase plasmids and 1 $\mu$g of RSV-CAT as an internal control were co-transfected into $5 \times 10^6$ HeLa cells in 10 cm dishes by the CaPO$_4$ method, and cell lysates were prepared two days after transfection. Luciferase activity [12] was normalized for transfection efficiency by CAT activity [13].

### 2.3. *RNase protection assay*

Fifty $\mu$g of BP-M-Luc or BP-D-Luc and 10 $\mu$g of the internal control plasmid containing the RSV driven human growth hormone gene (RSV-GH) were co-introduced into $1.5 \times 10^7$ HeLa cells in 15 cm dishes and total RNA from transfected HeLa cells was analyzed by RNase protection analyses [1] using the ML or DL riboprobes plus the growth hormone exon 1 riboprobe. The ML riboprobe is derived from the 370 bp *Hin*fl (within the luciferase gene 250 bp downstream from the junction to the DHFR/MRP1 promoter)–*Hind*III fragment

*Correspondence address:* T. Shimada, Department of Biochemistry and Molecular Biology, Nippon Medical School, 1-1-5 Sendagi, Bunkyo-ku, Tokyo 113, Japan. Fax: (81) (3) 5685 3054.
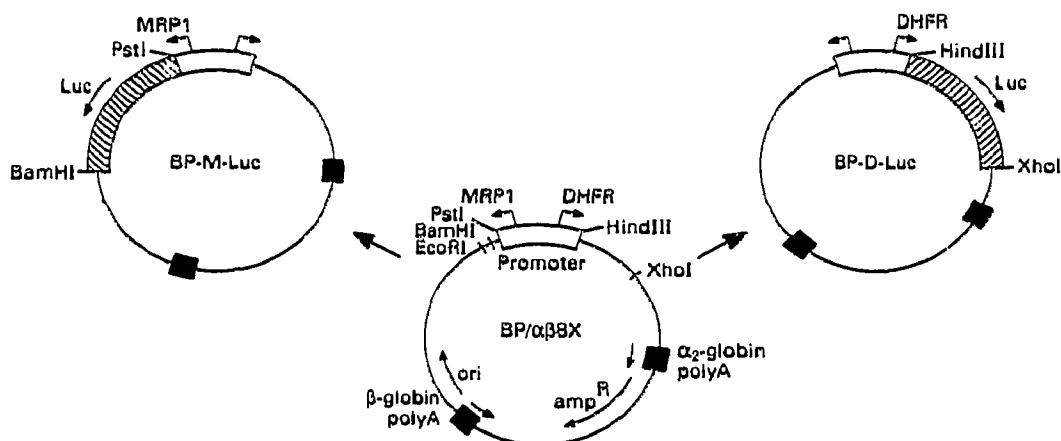
Fig. 1. Construction of firefly luciferase expression vectors for studying the DHFR/MRP1 bidirectional promoter.

of BPwt-M-Luc, while the DL riboprobe is derived from the 370 bp PstI–HinfI fragment of BPwt-D-Luc.

## 3. RESULTS

The 114 bp bidirectional promoter was chemically synthesized (BPwt), and the luciferase coding sequence was linked to either the DHFR side (BPwt-D-Luc) or the MRP1 side (BPwt-M-Luc) of the promoter (Fig. 1). These luciferase constructs and the internal control plasmid (RSV-CAT) were co-introduced into HeLa cells and the promoter activity was determined by monitoring the luciferase activity after 48 h (Fig. 2). BPwt-M-Luc gave about two-fold higher activity than BPwt-D-Luc, consistent with our previous results using the 165 bp promoter linked to the CAT gene [4].

In the first experiment, we mutated an AT-rich segment, ACAAATA, located 29 bp upstream from the DHFR initiation site. This is the only AT-rich stretch in the GC-rich promoter sequence. This sequence is not

a consensus match for the TATA box, but this position is usually occupied by the TATA box with respect to the DHFR initiation site. Such AT-rich segments have been also found in several GC-rich, TATA-less promoters [14–16], and some of these segments have been shown to be able to functionally substitute for the TATA box [17]. Accordingly we wanted to examine two possible functions of this element. The first possibility was that this was one of the variant forms of the TATA box and was involved in positioning of the DHFR initiation site. The second possibility was that this sequence was irrelevant to promoter function. In this case, however, it may be possible that lack of the TATA box was responsible for heterogenous initiation including bidirectional transcription.

We constructed two mutant promoters and checked their promoter activity by the luciferase assay (Fig. 2) and the transcription initiation site by the RNase protection assay (Fig. 3). In BP5, the AT-rich sequence was destroyed by substitution with G or C. This mutation
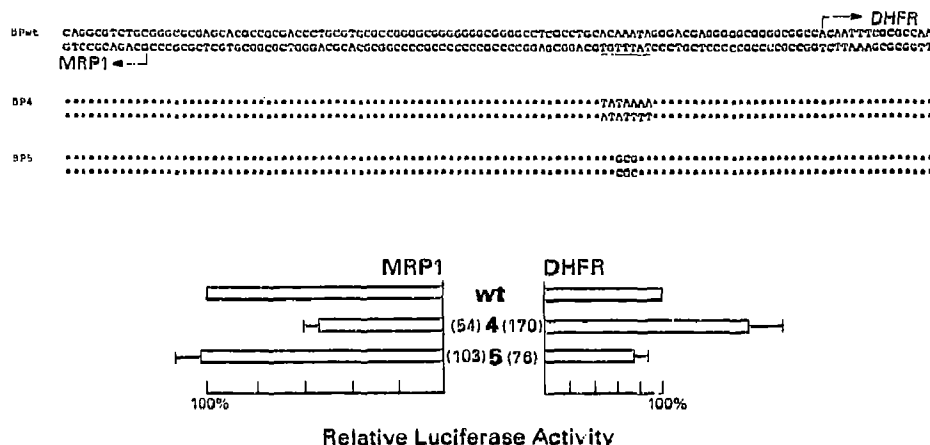




Relative Luciferase Activity

Fig. 2. Relative luciferase activity of the DHFR/MRP1 bidirectional promoters with mutation of the AT-rich sequence. The AT-rich sequence, ATAAACA, is underlined. Asterisks of the mutant promoters indicate the same nucleotide as in the wild type promoter (BPwt). The luciferase activity of each construct relative to BPwt-M-Luc or BPwt-D-Luc is shown in bar graph form, and its value (%) is indicated in parentheses. Data from at least two separate experiments with duplicate samples are averaged. Error bars represent the standard error of the mean.
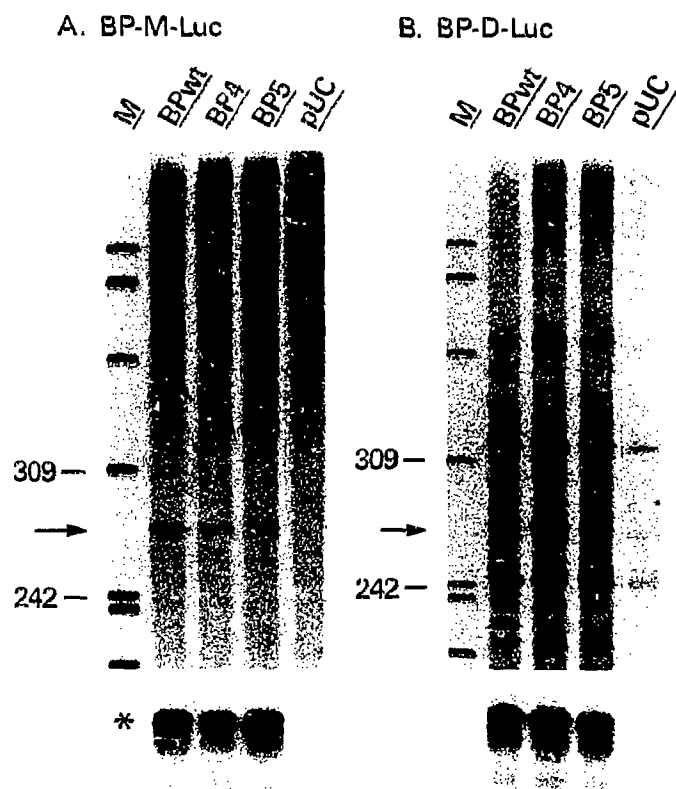
34

## A. BP-M-Luc

## B. BP-D-Luc



Fig. 3. RNase protection analysis of luciferase transcripts. M, end-labeled MspI digested pBR322 marker; pUC, control RNA extracted from HeLa cells transfected with pUC plasmid DNA. Arrows indicate the protected fragments derived from correctly initiated transcripts. Asterisks indicate the 68 bp fragment from growth hormone transcripts that served as an internal control.

did not result in any significant changes in promoter activity or the initiation site for transcription in either direction. Therefore, this particular AT-rich sequence is not functional in this promoter.

In BP4, the AT-rich sequence was substituted by the perfect consensus sequence of the TATA box, TA-TAAA. This substitution increased DHFR promoter activity (BP4-D-Luc) and decreased MRP promoter activity (BP4-M-Luc) without changing the initiation sites. This mutant promoter, however, still retained bidirectional activity. Therefore, bidirectionality of the DHFR/MRP1 promoter is not simply due to lack of the TATA box.

In the next series of experiments, we studied the role of the multiple GC boxes in the control of the bidirectional promoter. There are four GC boxes in the minimal DHFR/MRP1 promoter in the same orientation; two overlapping GC boxes in the middle of the intergenic region and two overlapping GC boxes in the region immediately upstream from the DHFR start site. In the BP1 mutant, the two overlapping GC boxes were converted to a single copy of the GC box in both regions (BP1). Both the luciferase assay and an RNase protection assay (data not shown) showed that the BP1 pro-

moter has full activity in either direction, suggesting that the overlapping structure of the GC boxes is not important for bidirectional activity, and two GC boxes (I or II and III or IV) function in this promoter.

The consensus sequences of the GC boxes were mutated by substitution with A or T (Fig. 4). Mutation of both GC boxes III/IV (BP7) resulted in complete loss of promoter activity in both directions. Mutation of the other pair of GC boxes I/II (BP8) led to about 50% decrease in activity in both orientations. BP12 with mutations of all four GC boxes showed no activity. In BP15, the GC boxes I/II were mutated and the GC box IV was deleted, and therefore, this mutant promoter contains only one GC box (III) in the middle. BP15 retained bidirectional promoter activity (Fig. 4).

These results suggest that the GC boxes are able to activate transcription initiation complexes on both sides. One middle GC box is essential for supporting bidirectional transcription.

## 4. DISCUSSION

Promoters of several mammalian genes have been shown to have bidirectional activity [14,15,18,19]. However, none of these promoters has been characterized in detail. We showed that the GC box plays an important role in bidirectional activity of the DHFR/MRP1 promoter. Since the GC box has been found in either orientation in many promoters, this element has been thought to regulate transcription in an orientation-independent manner [20]. Our results showed that the GC box is able to function as a bidirectional activator element. It has been reported that Sp1 molecules bind to the GC boxes and stimulate transcription of the mouse DHFR gene [6]. Two initiation complexes for DHFR and MRP1 transcription appear to share these Sp1 molecules.

A middle GC box was shown to be essential for transcription of both the DHFR and the MRP1 genes. Another GC box immediately upstream of the DHFR initiation site is not essential, but activates transcription of both genes. The difference in activation efficiency of each GC box might depend on the sequence surrounding the GC box or the distance between the GC box and the initiation site.

Because the GC box has been found in many unidirectional promoters, the GC box alone could not be sufficient for bidirectional activity. Transcripts of both DHFR and MRP1 genes start at specific sites. Our results confirmed that positioning of the initiation sites of these genes is regulated in a TATA-independent manner. Recently, Means and Farnham [9] suggested that transcription initiation from the mouse DHFR gene may be positioned by an initiator element, that specifies the initiation site within the element itself [21]. In contrast, Blake et al. [22] recently reported that start site utilization in the Chinese hamster DHFR gene is mainly
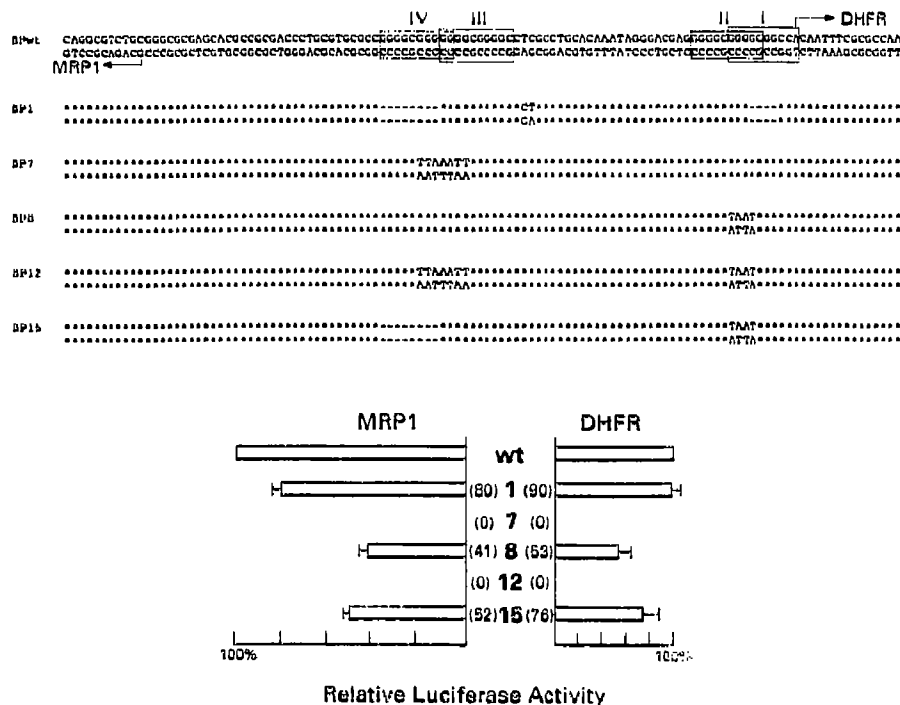
Fig. 4. Relative luciferase activity of the DHFR/MRP1 bidirectional promoters with mutation of the GC boxes. Consensus sequences of four GC boxes were boxed [23]. Experimental conditions are the same as described in Fig. 2.

regulated by upstream GC boxes. Thus, transcription initiation from the TATA-less promoter of the DHFR gene seems to be regulated by interactions of upstream GC boxes and the initiator element. Although the mechanism of positioning of MRP1 transcripts has not been studied, another positioning element might be required for accurate and efficient expression of the MRP1 gene, and existence of a GC box and two positioning elements on both sides might be sufficient for minimal bidirectional activity.

## REFERENCES

[1] Fujii, H. and Shimada, T. (1989) J. Biol. Chem. 264, 10057–10064.

[2] Priebe, S.D., Hadi, S.M., Greenberg, B. and Lacks, S.A. (1988) J. Bacteriol. 170, 190–196.

[3] Linton, J.P., Yen, J.-Y.J., Selby, E., Chen, Z., Chinsky, J.M., Liu, K., Kellems, R.E. and Crouse, G.F. (1989) Mol. Cell. Biol. 9, 3058–3072.

[4] Shimada, T., Fujii, H. and Henry, L. (1989) J. Biol. Chem. 264, 20171–20174.

[5] Chen, M.-J., Shimada, T., Moulton, A.D., Cline, A., Humphries, K., Maizel, J. and Nienhuis, A.W. (1984) J. Biol. Chem. 259, 3933–3943.

[6] Dynan, W.S., Sazer, S., Tjian, R. and Schimke, R.T. (1986) Nature 319, 246–248.

[7] Swick, A.G., Blake, M.C., Kahn, J.W. and Azizkhan, J.C. (1989) Nucleic Acids Res. 17, 9291–9304.

[8] Blake, M.C. and Azizkhan, J.C. (1989) Mol. Cell. Biol. 9, 4994–5002.

[9] Means, A.L. and Farnham, P.J. (1990) Mol. Cell. Biol. 10, 653–661.

[10] Farnham, P.J. and Schimke, R.T. (1986) Mol. Cell. Biol. 6, 2392–2401.

[11] Liu, J.M., Fujii, H., Green, S.W., Komatsu, N., Young, N.S. and Shimada, T. (1991) Virology 182, 361–364.

[12] DeWet, J.R., Wood, K.V., DeLuck, M., Helinski, D.R. and Subramani, S. (1987) Mol. Cell. Biol. 7, 725–737.

[13] Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) Mol. Cell. Biol. 2, 1044–1051.

[14] Burbelo, P.D., Martin, G.R. and Yamada, Y. (1988) Proc. Natl. Acad. Sci. USA 85, 9679–9682.

[15] Soininen, R., Huotari, M., Hostikka, S.L., Prockop, D.J. and Tryggvason, K. (1988) J. Biol. Chem. 263, 17217–17220.

[16] Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkvens, Th.M., Van der Voorn, L., Van Ormondt, H. and Van der Eb, A.J. (1985) EMBO J. 4, 437–443.

[17] Singer, V.L., Wobbe, C.R. and Struhl, K. (1990) Genes Dev. 4, 636–645.

[18] Efrat, S. and Hanahan, D. (1987) Mol. Cell. Biol. 7, 192–198.

[19] Grichnik, J.M., French, B.A. and Schwartz, R.J. (1988) Mol. Cell. Biol. 8, 4587–4597.

[20] Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) Trends Biochem. Sci. 11, 20–23.

[21] Smale, S.T. and Baltimore, D. (1989) Cell 57, 103–113.

[22] Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J.W. and Azizkhan, J.C. (1990) Mol. Cell. Biol. 10, 6632–6641.

[23] Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) Cell 50, 863–872.