ARTICLE

A novel protein-DNA interaction involved with the CpG dinucleotide at -30 upstream is linked to the DNA methylation mediated transcription silencing of the *MAGE-A1* gene

Jie ZHANG, Jian YU, Jun GU, Bao Mei GAO, Ying Jun ZHAO, Peng WANG, Hong Yu ZHANG, Jing De ZHU*

The State-key Laboratory for Oncogenes and Related Genes, Shanghai Cancer Institute, Shanghai Jiao Tong University, LN 2200/25, Xietu Road, Shanghai 200032, China.

ABSTRACT

To understand the DNA-methylation mediated gene silencing mechanisms, we analyzed in cell culture of the promoter function of the *MAGE-A1* gene, which is frequently demethylated and over-expressed in human hepatocellular carcinoma. We have established the correlation of the DNA methylation of the promoter CpG island with expression status of this gene in a panel of the established liver cancer cell lines. The crucial CpG dinucleotide(s) within the minimal promoter subjected to the control mediated by DNA methylation with profound biological functions was also delineated. Furthermore, a novel sequence-specific DNA-protein interaction at the -30 CpG dinucleotide upstream of the gene was found having a vital part to play in the DNA methylation mediated transcription silencing of the *MAGE-A1* gene. Our results would not only provide new insights into the DNA methylation mediated mechanisms over transcription of the *MAGE-A1* gene, but also pave the way for further defining the cross-talk among DNA methylation, histone modification and chromatin remodeling in detail.

Keywords: MAGE-A1, DNA methylation, transcriptional regulation.

INTRODUCTION

Addition of a methyl group at the fifth carbon of the cytosine in the CpG dinucleotide is the sole covalent modification of DNA in vertebrate genome and recognized as an important epigenetic signature [1]. The vital importance of the normal DNA methylation pattern to the well being of vertebrates has not only been suggested by the embryonic lethality of the knock-out mice where each of three methyl transferase genes was individually affected, but also by its aberrant profile in various human diseases, cancer in particular [1, 2]. In addition to the global level of reduction of methylation, changes in the promoter CpG islands have been widely detected in all the tumors so far tested. Their hypermethylated state has been experimentally linked to the transcriptional inactivation of the tumor suppressor genes, including the genes encoding for p16[INK4b], RASSF1A and a dozen of others [2]. Hence, the increased level of methylation at the promote CpG islands

has been regarded as an alterative to genetic defects [3] for functional loss of the tumor suppressor genes. Despite of lacking the mechanistic details and sequence of events for the DNA methylation mediated transcriptional silencing, the concept has been well established that hypermethylation of promoter CpG island reflects the long-term transcription silencing status of the genes in cell.

However, the reverse process where the local demethylation in parallel with the transcriptional reactivation of the otherwise methylated and silenced genes in the normal tissues, had not been reported in human cancers until recently. From our latest endeavor to methylation profile the promoter CpG island of two dozens of genes in human hepatocellular carcinoma (HCC), one of the most threatening cancerous diseases in China, we found that the *MAGE-A1* gene was heavily methylated in the normal liver tissues, but demethylated in 75% of tumor tissues [4]. Coincidently, its expression (at the serum protein level) was evident in more than 70% live cancer patients, while undetectable in the normal counterparts in an independent study [5].

The *MAGE-A1* gene is the founding member of a large

*Correspondence: Jing De ZHU
Tel: (86) 21 64224285, Fax: (86) 21 64224285
E-mail: zhujingde@yahoo.com or zhujingde@21cn.com

gene family encoding various testis/cancer antigens,which share a common domain, the MAGE homology domain. It was first identified on the surface of melanoma cells, recognized by cytotoxic lymphocytes and constituted by a HLA-A1-specificity-restricted nonapeptide [6]. Although the tumor specific expression of these proteins brought about the great interests in their potentials in the field of the anti-cancer immunotherapy, the biological function of this protein family remains largely unknown until recently [7]. Almost behind the same rationale, the promoter of the *MAGE-A1* gene has been used as a model system for the mechanistic studies of the transcriptional process per se as well as in the context of tumor biological behaviors [8-11]. To unveil the underlying mechanisms for the DNA methylation mediated control of transcription in HCC [4], we dissected in detail the *MAGE-A1* promoter in the context of its methylation status in cell and found that a novel DNA-protein interaction with the CpG at -30 upstream was sensitive to the changes in the methylation status and likely plays the vital role in the transcriptional control of the *MAGE-A1* gene in liver cancer cell lines. Both implication of our finding and merits of the molecular approaches used in this study were discussed.

MATERIAL AND METHODS

Report constructs

To assess the effects of the CpG methylation over the *MAGE-A1* promoter, a CpG deficient luciferase reporter gene construct based upon pGL3-luc (Promega, USA) was made by PCR cloning the CpG free luciferase gene from commercially available reporter pMOD-Luc::Sh (Invivogen USA) with a pair of primers (Tab. 1). The PCR fragment was digested with Hind III and Xba I and cloned at the same sites of pGL3-Basic (Promega, USA), to create the pGL3-MOD.

A 1.1 Kb PCR fragment containing the promoter region of the *MAGE-A1* gene (-793 to +263) from the HEK-293 cellular DNA was PCR cloned into pGL3-MOD. The 5'-deletion mutant reporter constructs (-369 ~ +263 and -85 ~ +263) was made by restriction digestion and the deletion mutants (-56 and -38 ~ +263) was made by PCR cloning with the primer pairs listed in Tab. 1.

Each of five CpG sites that are within the Hpa II (CCGG, -81, -60 and -52) or Hha I site (GCGC, -30 and +129) was mutated by use of a bridging-based two round PCR method [12-14]. Special precautions were made to avoid any accidental changes in the putative cis-elements that may reside in the sequence involved, according to the bioinformatics analysis for the cis-elements recognized by the known transacting factor (http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi).

Cell culture, and transient transfection/reporter gene assays [12-14]

SMMC-7721 (human hepatocellular carcinoma cell line, No. TChu13 in Shanghai Cell Bank), BEL-7402 (human hepatocellular carcinoma cell line, Catalogue No. TChu68 in Shanghai Cell Bank), HEK-293 (human embryonic kidney cell line, ATCC No. CRL-1573, HepG2 (human hepatocellular carcinoma cell line, ATCC No. HB-

8065), and Hep3B (human hepatocellular carcinoma cell line, ATCC No. HB-8064) were grown at 37ºC with 5% CO_2 atmosphere in Dulbecco's modified Eagle's medium (DMEM) supplied with 10% fetal bovine serum and ampicillin(100 U/ml) / streptomycin (100 μg/ml).

All the transient trnsfection/resporter experiments were carried out in duplicate. Both the tested promoter fragments that are used to drive the CpG free firefly luciferase gene and the pRL-CMV vector (Promega) where *Renilla reniformis* luciferase is driven by the promoter of CMV were transfected into cells. The luciferase activities of the cell extracts from the recipient cells were quantified by a single-tube assay (Dual-luciferase reporter assay system, Promega, USA), and the promoter potency of the tested fragments were presented as the mean and standard deviation of the ratio of the firefly luciferase over the renilla luciferase activity of the duplicates and plotted against the tested constructs. All the experiments were carried out, at least, three times, and the result from one representative experiment was presented.

To assess the effects of the methylation status of CpG over the promoter potency in the transient transfection assay, the plasmid was methylated *in vitro* with each of following three methylases: M.SssI, M. HpaII and M.Hha I methylases according to the manufacturer's instruction (NEB, Boston, USA). The reaction was verified by Hpa II and Hha I digestion as specified prior to the reporter assays.

RNA preparation and expression profiling by a semi-quantitative PCR

The total RNA was prepared from cultured cells with Trizol reagent following the manufacturer's instruction (Invitrogen, USA). Then the total RNA was reverse transcribed with oligo-dT as primer and Superscript II™ RNase H-Reverse Transcriptase (Invitrogen, USA). Semi-quantitative PCR was carried out with a pair of primers, each of which resides at different exons of the target gene (Tab. 1). The PCR cycling parameters are 94 ºC for 3 min, following by 25-30 cycles of three steps: 94ºC for 30 s, 60ºC for 30 s, and 72ºC for 40 s, and 72ºC 5 min at last. The PCR products were separated on an agarose gel and the ethidium bromide stained bands were visualized under UV illumination.

Methylation specific PCR (MSP) [4, 15]

2 μg DNA in 50 μl TE was incubated with 5.5 μl of 3 M NaOH at 37ºC for 10 min, followed by a 16 h treatment at 50ºC after adding 30 μl of freshly prepared 10 mM hydroquinone and 520 μl of freshly prepared 3.6 M sodium bisulfite at pH 5.0. The DNA was desalted using a home-made dialysis system with 1% agarose (detailed protocol will be provided upon request). The DNA in the desalted sample (approximately 100 μl in volume) was denatured at 37ºC for 15 min with 5.5 μl of 3 M NaOH followed by ethanol precipitation with 33 μl 10 M NH₄Ac and 300 μl ethanol. After washing with 70% ethanol, the gently dried DNA pellet was dissolved with 30 μl TE at 65ºC for 10 min and stored at -20ºC till use. PCR reaction in a volume of 15 μl with 50 ng or less template DNA with FastStart Taq polymerase (Roche, Germany) was carried out in the follwing condition: 4 min at 94ºC with 30 cycles of 92ºC for 15 s, 58ºC for 15 s and 72ºC for 20 s. The PCR products were separated by 1.2% ethidium bromide containing agarose gel electrophoresis with 1×TAE and visualized under UV To verify the PCR results, representative bands from each target were gel-purified and cloned into pGEM-T Easy vector (Promega, USA) followed by automatic DNA sequencing provided by Biocolor

Tab. 1 The primers used in this paper

Primer Name	GenBank No	Sequence (5' to 3')	Length of PCR product(bp)	Note
mager	U82670	taccaccaacacgacga	1080	Cloning full length promoter
magpfb		gaagatcttcacagcaacctctgattc		
RVprimer3	/	gagcagacaagcccgtca	997	The franking primers of bridge-based PCR
pMOD-B		gagcatcagtgaaggcaatg		
MOD-L	/	taactatgcggcatcagagc	1798	PCR the CG free luc from pMOD-Luc::Sh
MOD-R		gctctagattatttgccacccttcttgg		
magea1bdf	NM_004988	ggccgaaggaacctgac	345	MAGE-A1 RT-PCR
magea1dt		tctggggaggatctgttgac		
NBA1	X00351	aagtactccgtgtggatcgg	616	β-actin RT-PCR
NBA2		tcaagttgggggacaaaaag		
MAGEA1-Uf		gtttggttgaaggaatttga	305	
MAGEA1-Ut		acccacaaccctccctctta		
MAGEA1-Mf		gttcggtcgaaggaatttga	303	
MAGEA1-Mt	U82670	ccacaaccctccctcttaaa		Methylation specific PCR
MAGEA1-U5		tttggtttttgttaggaaatatttg	274	
MAGEA1-U3		taaaaacaacccaaactaaaacatc		
MAGEA1-M5		attcggttttcgttaggaaatattc	275	
MAGEA1-M3		taaaaacaacccaaactaaaacgtc		
mageCG1-L		aaccgTatcctgcctgggtggg	413	
mageCG1-R		cAggatacggttcccgccagg	583	
mageCG2-L		gggTaccAggatgtttcctggcgg	436	
mageCG2-R		atccTggtAcccggatgtgacgcc	556	
mageCG3-L	U82670	catccTggTacccggatgtttcctgg	440	For making the CG1-5 mutants
mageCG3-R		ggtAccAggatgtgacgccactgact	553	
mageCG4-L		atgcAcaagtcagtggcgtcacatc	465	
mageCG4-R		tgacttgTgcattgtggggcagaga	531	
mageCG5-L		cActggaatatttggggctctctatt	623	
mageCG5-R		aatattccagTgccgcccttgctgc	432	

(Shanghai, China).

Demethylating treatment of Hep 3B cells with 5-Aza-2'-deoxycytidine

When the culture reaches to the 50% in confluence, 5-Az'-deoxycytidine (Sigma A3656) was added at the final concentration of 10 and 20 mM, respectively for 3. Both DNA (for MSP analysis) and RNA (for the expression assessment by RT-PCR) were extracted and analyzed as previously described [12, 15].

Electrophoresis mobility shift assay

Nuclear extracts were prepared from cultured cells by a modified protocol of the Dygnan [16]. Two complementary strands of the oligonucleotides at an equal molar input were 5' kinase labeled with [γ-^{32}P]ATP (5000 Ci/mM) as previously described. After heat-in-

activating the kinase at 65°C for 10 min, NaCl was added into the reaction to a final concentration of 0.1 M. The reannealling process to form double-stranded oligonucleotides molecule was carried out by gradual cooling down for 2 h. 5 μg of the nucleoprotein extract was incubated at 4°C or 10 min with a 19 ml solution containing 1 mg poly dI:dC (Roche), 12.5 mM Herpes, pH 7.9, 6.25 mM MgCl$_2$, 50 mM KCl and 10% glycerol, followed by an further 20 min incubation with 10 mM of the newly added ^{32}P-end-labelled double stranded oligonucleotides. The DNA-protein complexes were analyzed in a cold room (-7°C) by electrophoresis on a non-denaturing 4% polyacrylamide gel with 0.5×TBE as running buffer (1×TBE: 89 mM Tris, 89 mM boric acid, 8 mM EDTA, pH 8.0). 100 fold molar excess of the individual double-stranded oligonucleotides were included in the binding reaction for competition analyses.

To evaluate methylated CpG effects on the protein-DNA binding, the reannealed double-stranded oligonucleotide: MAGE07 (cold)

was methylated by M.HhaI methylase in presence of S-adenosyl methionine (NEB, Boston, USA).

RESULTS

The methylation status of the promoter CpG island of the *MAGE-A1* gene correlates with the expression profile

DNA methylation is an important step for the transcriptional silencing, and regarded as the early event of a cascade of the biochemical change that lead chromatin condensation [1]. Approximately 40-50% of the protein coding genes have, at least, one CpG island at their 5 proximity, indicating transcription of which is under the control of changes in the methylation status. The *MAGE-A1* gene is such a gene, containing a typical CpG island spanning from -85 to +170, where the initiation nucleotide is defined as position: +1 (Fig. 1A). With the methylation specific PCR using the primers 1-4 (Fig. 1A), that specifically addressed the methylation status of the CpG at +26 and +30, we found that while they were fully methylated in the normal liver tissues from the healthy liver donors, both CpGs were at least partially demethylated in 75% of HCC [4]. Among the liver cancer cell lines, both BEL-7402 (Lane 1) and SMMC-7721 (Lane 2) heterozygously methylated, while both HepG2 (Lane 3) and Hep3B (Lane 4) cell lines were fully methylated (Fig. 1B). A semi-quantitative PCR analysis was carried out to assess the expression state of the *MAGE-A1* gene in parallel with the β-actin control in each cell lines (Fig. 1C). While the methylated cell lines HepG2 and Hep3B cells (Lanes 3 and 4) did not express, those containing unmethylated alleles (BEL-7402, Lane 1 and SMMC-7721, Lane 2) expressed the *MAGE-A1* gene abundantly, compatible to that of the β-actin control. Therefore, the correlation of the *MAGE-A1* gene with methylation status of the promoter CpG (to be more specific, the CpG at +26 and +30) island in this set of liver cancer cell lines was established. There are 17 CpG within the CpG island (-85 to +170). It is logical to believe that each CpG may contribute to the promoter active status to different extents, as the other suggested in a similar study in a panel of the human leukemia cell lines [11]. We carried out the methylation specific sequencing of this whole region with primers 5-8 in Fig. 1A, by which the methylation status of each CpG within this region can be determined. The PCR fragments were T-vector cloned and four (the methylated) from HepG2 and seven clones (the methylated: three and the unmethylated: four) from BEL-7402 cells were sequenced. As shown in Fig. 1D, the CpGs at -30, +14, +17, +26 and +30 were fully methylated (4/4) and -43, as well as -8 were methylated in 3/4 clones, suggesting that the methylated form of these sites (at least, some of them) may be instrumental to the silenced status

of the *MAGE-A1* gene in HepG2 cells. In BEL-7402 cells, four clones from the unmethylated allele contained no methylated CpG and three clones from the methylated allele assumed a different pattern from the HepG2's. The CpG were fully methylated: -81, -74, -60, -8, +17, +26, +30 and +169; 2/3 methylated: -30 and +14; and 1/3 methylated: -52. Since BEL-7402 cells expressed the *MAGE-A1* gene at a high level, the impart of the methylated allele (more specifically, which methylated CpG) to the level of transcription can not be assessed.

5 aza-2'-deoxylcytidine is a commonly used DNA methyl transferase inhibitor and is able to lead to the genome wide demethylation in almost all the cell culture systems tested. With this inhibitor, we tested whether the forced demethylation of the promoter CpG island would result in the transcriptional reactivation of the *MAGE-A1* gene in Hep 3B cells. As Shown in Fig. 1, upon 3 d treatment (with 10 and 20 mM, respectively), the density of the methylated band was reduced while the unmethylated band became evident (Fig. 1E) and the *MAGE-A1* RNA became detectable (Fig. 1F). Therefore, the transcription of the *MAGE-A1* gene is indeed dependent on the DNA-methylation mediated control.

The minimal promoter of the *MAGE-A1* gene in liver cancer cells is located at the region spanning from -56 to +263

With the DNA methylation mediated mechanisms as the central focus of this study, we made all the reporter constructs with the CpG free luciferase gene, as the methylated CpGs within the coding region have been shown capable of leading to complete transcription silence (unpublished observation). A CpG deficient luciferase reporter gene construct based upon pGL3-basic (Promega, USA) was made by PCR cloning of the CpG free luciferase gene from commercially available reporter pMOD-Luc::Sh (Invivogen, USA) to create the PGL3-MOD.

We assessed the promoter competence of the *MAGE-A1* gene in BEL-7402 and HepG2 cells, which differed in both the methylation status and expression profile of the endogenous *MAGE-A1* gene. As shown in Fig. 2A, the promoter fragment spanning -397 to +263 created a 85% of the level of the luciferase activity by the control promoter, where both SV40 promoter and enhancer were used to drive the firefly luciferase reporter in BEL-7402 cells and a much reduced level (42%) in HepG2 cells. Despite of being reduced, not completely inert status of the *MAGE-A1* promoter in HepG2 cells was not a total surprise. The remaining activity of the promoter may reflect the fact that the silenced epigenetic signatures registered in the endogenous *MAGE-A1* gene (DNA methylation, histone modification and chromatin remodeling) are
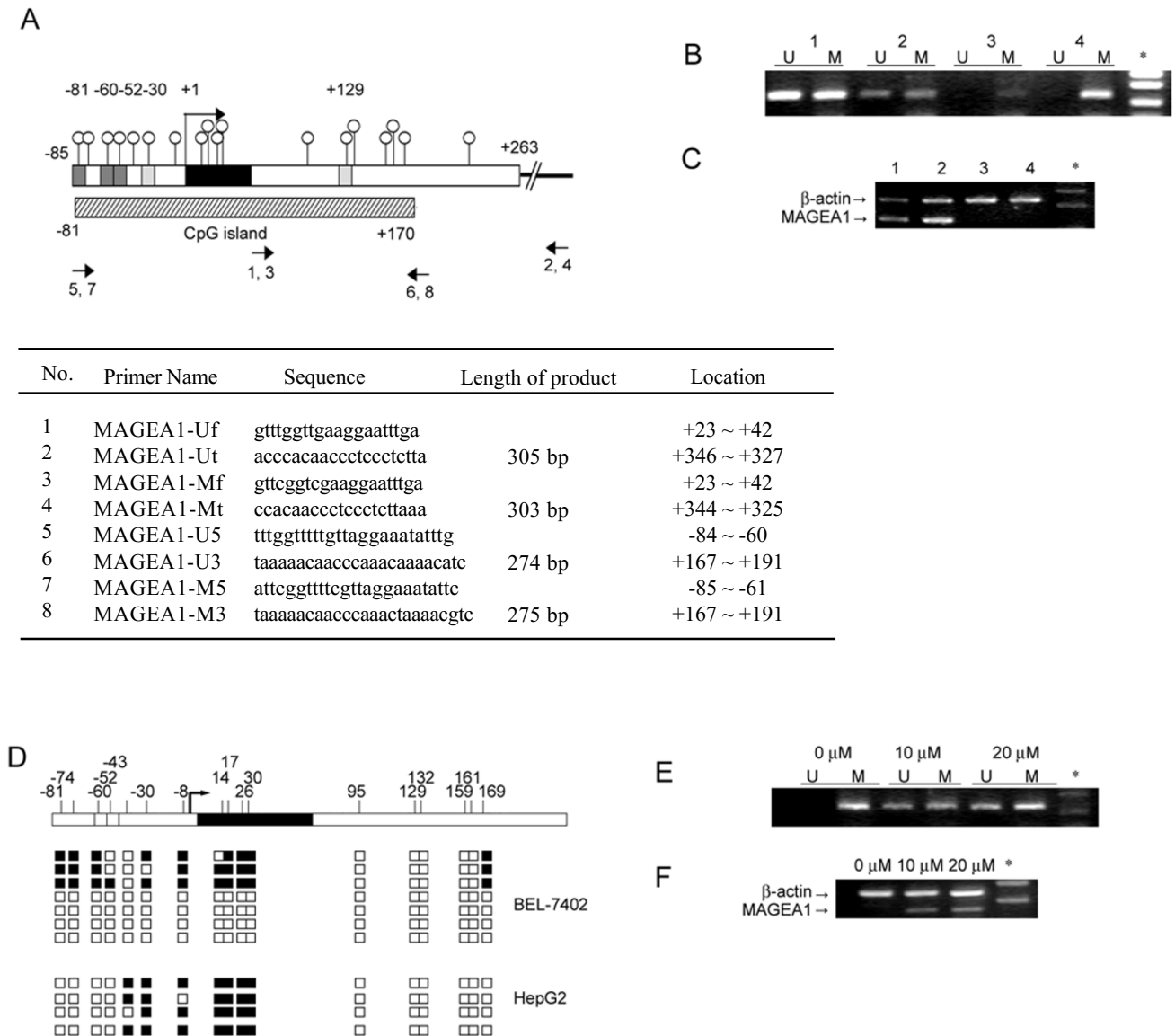
No.	Primer Name	Sequence	Length of product	Location
1	MAGEA1-Uf	gtttggttgaaggaatttga		+23 ~ +42
2	MAGEA1-Ut	acccacaaccctccctctta	305 bp	+346 ~ +327
3	MAGEA1-Mf	gttcggtcgaaggaatttga		+23 ~ +42
4	MAGEA1-Mt	ccacaaccctccctcttaaa	303 bp	+344 ~ +325
5	MAGEA1-U5	tttggtttttgttaggaaatatttg		-84 ~ -60
6	MAGEA1-U3	taaaaacaacccaaacaaaacatc	274 bp	+167 ~ +191
7	MAGEA1-M5	attcggttttcgttaggaaatattc		-85 ~ -61
8	MAGEA1-M3	taaaaacaacccaaactaaaacgtc	275 bp	+167 ~ +191

Fig. 1 The methylation and expression status of the *MAGE-A1* gene in liver cancer cell lines. (**A**) The line drawing of the *MAGE-A1* promoter. Arrow: the transcription initiation site (+1). The CpG island (from -81 to +170) is presented as the stripped bar. The mushroom type signs above the line drawing represents the CpG dinucleotide. The dark bar is the 1st exon, the fainter grey bars represent the Hha I site (GCGC) and the grey bars represent Hpa II sites (CCGG). Sequence and location of the primers (1-8) as well as the size of the PCR products are also given. (**B**) The methylation status of the *MAGE-A1* gene in liver cancer cell lines with primers 1-4 is presented. * the size marker: DL2000; M: the primers for the methylated CpG site; U: the primers for the unmethylated CpG site. (**C**) The semi-quantitative PCR analyses for the expression profile of the *MAGE-A1* gene in liver cancer cells. (**D**) The methylation status of each CpG site of the *MAGE-A1* promoter (-85 to +263) in BEL-7402 and HepG2 cells. The PCR fragments with primers 5-8 were T-vector cloned and sequenced. Each line represents the methylation status from a single clone. The location of each CpG dinucleotide is indicated above the line drawing. The filled bar refers to the 1st Exon. The empty box: the unmethylated CpG, and the filled box: the methylated CpG. (**E**) and (**F**) Demethylation of the promoter CpG island of the *MAGE-A1* gene by 5 aza-2'- deoxylcytidine resulted in activation of the transcription of the *MAGE-A1* gene. The Hep3B cells were treated for three days with 10 and 20 μM 5'-aza-2'-deoxycytidine, respectively, followed by analyses for both DNA methylation and expression status of the *MAGE-A1* gene. E, the methylation status of the promoter CpG island with primers 1-4; F, the expression profiles prior and posterior to the 5 aza-2'-deoxylcytidine treatment in Hep 3B cells.
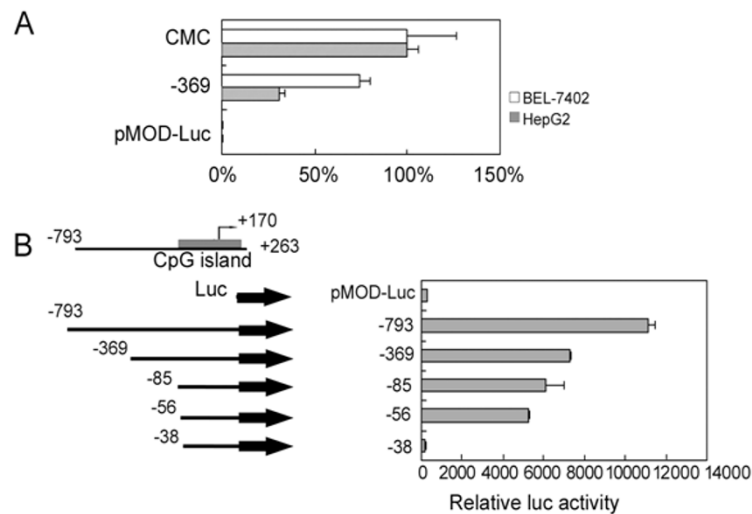
Fig. 2 The transient transfection analysis of the *MAGE-A1* promoter with the CpG free luciferase constructs. (**A**) The promoter potency of the *MAGE-A1* gene in BEL-7402 (expressing) and HepG2 (non-expressing) cell lines. The reporter constructs where the CpG free luciferase gene was driven, respectively by the SV40 promoter/enhancer (control) and the -369 to +263 fragment of the *MAGE-A1* gene were tested for the promoter potency in the expressing (BEL-7402) and non-expressing (HepG2) cell lines by transient transfection assay. The relative potency of the *MAGE-A1* promoter was compared by taking the activity of the control promoter arbitrarily as 100%. pGL3-Mod, the promoterless control. (**B**) The minimal promoter of the *MAGE-A1* promoter defined by 5' progressively deletion mutagenesis. The line drawing represents the *MAGE-A1* promoter (-793~+263) with a CpG island from -81 to +170 and a schematic illustration of the *MAGE-A1* reporter constructs used in transient transfection analysis of promoter activity in BEL-7402 is shown. The arrow indicates the transcription initiation site. The promoter-reporter constructs were co-transfected with a control plasmid (pRL-CMV) and assayed 24 h later. The data is shown as the means ± SD from three separate experiments.

lacking in the tested system with the exogenously introduced promoter constructs. Therefore, the otherwise completely normal transcriptional machinery for the *MAGE-A1* gene remains operational in HepG2 cells.

To narrow down the region where the crucial CpG subjected to the DNA methylation mediated control over the *MAGE-A1* gene expression locates, i.e., to define the minimal promoter, we assessed the promoter potency of a panel of the 5' progressively deletion mutants (-793, -397, -85, -56, -38 to +263, respectively) in BEL-7402 cells.

As shown in Fig. 2B, although the fragment spanning from -793 to +263 led to the highest level of the reporter activity, -369 displayed 2/3 of that of the former, the mutants with the more 5' progressive deletion: the -85 and the -56, did not displayed significantly reduced promoter activity. The -38 mutant lost all the promoter activity. In conclusion, the region between -56 and +263 represented the minimal promoter of the *MAGE-A1* gene.

The methylation status of the -81 and -30 CpG profoundly affected the promoter activity of the *MAGE-A1* gene in cells

There are seventeen CpG within and one outside of the

CpG island (from -81 to +163). Since the CpG may be present within or outside the cis-elements recognized by the key transcription factors, the functional consequence of the methylation status of each CpG would differ drastically. The reporter construct (-369 to +263) containing additional six CpG within the region between -369 to -85 were respectively *in vitro* methylated by one of the following three methylases: M. Sss I (capable of transferring the methyl group to the *C in C*pG, 26 CpG in this region), Hha I methylase (capable of transferring methyl group to the *C in 5'-G*CGC-3' at -30 and +129) and Hpa II methylase (capable of transferring methyl group to the *C in 5' C*CGG-3', at -81, -60 and -52, respectively). The promoter activity was assayed in comparison with their non-methylated counterparts in BEL-7402 cells. As shown in Fig. 3A, the *in vitro* methylation was complete, as both Hha I and Hpa II could cut the parental, but not the methylated plasmids. The full methylation of all the CpG sites by M.SssI abolished all the promoter activity. However, the methylation at all the Hpa II sites (-81, -60 and -52) or all the Hha I sites (-30 and +129) only reduced the promoter potency down to the approximately 1/3 of the level of that with the unmethylated (Fig. 3B), implying

Tab. 2 The bioinformatics analysis of the sequences embracing the Hpa II or Hha I sites for the putative cis-elements and mutation to destroy the restriction sites

Site	Mutation	Factor	Cognate sequences	A	C	G	T
-81	C→A	v-Myb	tcCGGTTccc	17	21	1	1
		v-Myb	cCGGTTccc	6	7	9	8
		Ik-1	ccggTTCCCgcca	0	2	13	3
		Elk_1	ccggTTCCCgccag	10	29	0	0
		RFX-1	cgGTTCCgccaggaaac	2	5	3	11
		c-Ets-1(p54)	cggtTCCCGc	1	4	4	5
-59	G→T	Elk-1	aaacATCCGggtgc	1	2	23	5
		c-Ets-1(p54)	aacaTCCGGg	7	0	9	5
		USF	tccgGGTGCc	0	3	1	1
-55	G→A	USF	tccgGGTGCc	11	0	12	
		Elk-1	gtgccCGGATgtga	2	9	6	5
		Elk-1	gtgcccGGATGtgacg	5	1	1	0
		Elk-1	gtgccCGGATgtga	2	23	2	1
-52	C→A	Elk-1	gtgcccGGATGtgacg	5	2	0	0
		c-Ets-1(p54)	cCCGGAtgtg	11	9	0	1
		GATA-3	ccgGATGTga	3	5	5	5
		USF	cggaTGTGAc	2	2	7	0
		AP-1	acTGACTtgcg	3	2	7	3
-30	C→T	AP-1	acTGACTtgcg	7	6	10	4
		Staf	tgacttgcgcattgTGGGGcag	1	2.9	15.5	74.6
		Nkx2-5	GACTTgc	5	3	2	1
		BSAP	gacttgcgCATTGtggggcagagagaag	1	1	0	1
		SOX-9	gcgcATTGTggggc	2	29	10	22
+129	C→T	Oct-1	ccaaaTATTCcagcg	10	11	6	13
		Ik-3	aataTTCCAgcgc	9	3	13	0
		Elk-1	aataTTCCAgcgcc	5	11	3	12
		Oct-1	aataTTCCAgcgcc	1	0	1	4
		c-Ets-1(p54)	atatTCCAGc	3	3	2	7
		Hand1/E47	tattCCAGCgccgccc	25	0	2	2

Columns:

Site: the location of CpG to be mutated; Mutation: the wild-type→the mutated nucleotide; Factor: the putative factor, the underlined C is destined to be mutated; Cognate sequence: the sequence recognized by the factor in the cell left to. The Capitals refer to the core consensus. The right most four columns (A, C, G, or T) refer the occurrence of headed nucleotide (A, C, G, or T) at the underlined nucleotide that needs to be changed. (http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi).

that methylation of all these five sites indeed contributed significantly to the transcriptional silencing status of the *MAGE-A1* gene in cell.

To determine which one of these five CpGs methycation of which contributes most to the of gene silencing, we made five mutants, each of which was individually mutated with the precaution to avoid any detrimental effects to the sequence-specific DNA-protein interactions, accord-ing to bioinformatics analysis (Tab. 2). The promoter potency of these *in vitro* methylated mutants (Fig. 3C) was compared with the unmethylated counterparts. As shown in Fig. 3D, the methylated -60 and -52 CpGs (-81 was mutated, CG1)did not cause any reduction in promoter activity. But the methylation of both -81 and -60 CpGs (-52 site was mutated, CG3), however, reduced the promoter petency significantly. Due to the unexpected low
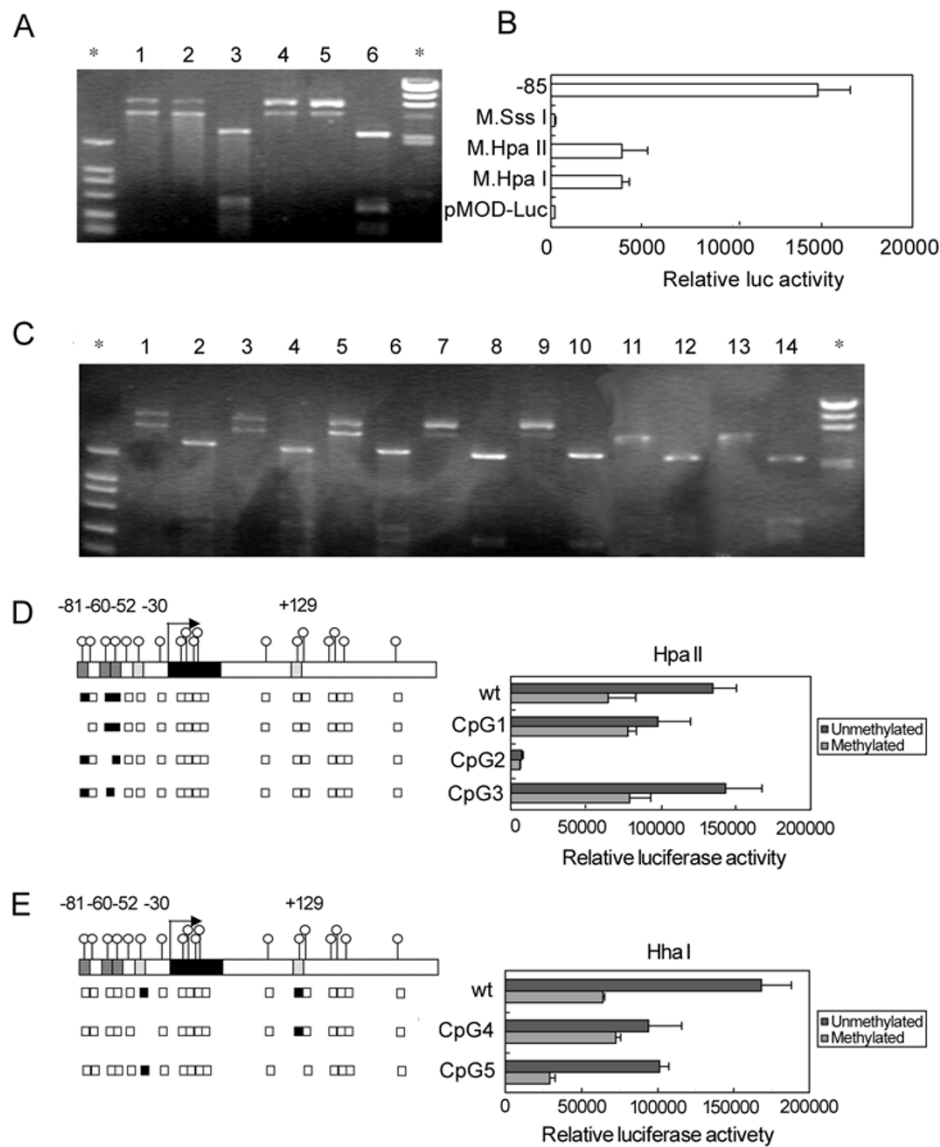
Fig. 3 Identification of the crucial CpGs responsible for the DNA methylation mediated transcription silencing of the *MAGE-A1* gene in BEL-7402 cells. (**A**) The plasmid containing the *MAGE-A1* promoter (-85 ~ +263) and the CpG free luciferase gene was methylated, respectively, by M.SssI (methylates the underlined C within 5'CG3'), M. HpaII (methylates the underlined C within 5' CC-GG 3') or M.HhaI (methylates the CpG underlying C within GCGC), and verified with HpaII or HhaI digestion (refractory to the CpG methylation). Lanes 1~3: digested with HpaII, lane 1: methylated by M.SssI, lane 2: methylated by M.HpaII, lane 3: unmethylated; lanes 4~6: digested with HhaI, lane 4: methylated by M.SssI, lane 5: methylated by M.HhaI, lane 6: unmethylated. (**B**) Equal amount of the methylated and unmethylated -85 to +263 reporter constructs were transfected into BEL-740 cells, respectively, and the promoter activity was measured and plotted. (**C**) Verification of the methylation of each of the five mutants (CG1-5) and the wild-type control: -369 ~ +263 fragment treated with Hpa II or Hha I methylase. Lanes in odd numbers: the *in vitro* methylated with designated methylase; the even numbers: the unmethylated samples. The lanes 1-2, and 5-10, digested by Hpa II; and lanes 3-4 and 11-14, digested with Hha I. *, size markers. Lanes 5-6, the CpG mutant at -81(CG1), lanes 7-8, the CpG mutant at -62 (CG2), lanes 9-10: the CpG mutant at -50 (CG3); lanes 11-12, the CpG mutant at-30 (CG4), and lanes 13-14, the CpG mutant at +129 (CG5). (**D**) The methylation effects of -81 (CG1), -60 (CG2) and -52 (CG3) over the promoter activity. Both unmethylated and Hpa II methylase methylated wild type, and CpG mutants at -81, -62 and -50 were transfected into BEL-7402 cells and the reporter activities were measured and compared. (**E**) The methylation effects of -30 (CG4) and +129 (CG5) over the promoter activity. Both unmethylated and Hha I methylase methylated wild type, and CpG mutants at -30, and +129 were transfected into BEL-7402 cells and the reporter activities were measured and compared.

activity of the -60 mutant, CG2, the effect of the methylation of -81 and -52 cannot be accurately determined. Therefore, the methylation status of -81 may be the one which contributed significantly to the reduction of promoter activity.

As shown in Fig. 3E, when both -30 and +129 CpGs (the wild type) were methylated, the promoter activity was reduced to about 1/3 of the unmethylated ones. It was interesting to notice that the methylated +129 CpG (the -30 mutant, CG4), while the methylated -30 CpG (+129 mutant, CG5) produced the same effects as the methylated parental, suggesting that the methylated -30 CpG might indeed contribute greatly to the methylation mediated silencing of the *MAGE-A1* gene *in vivo*.

Methylated status of the -30 CpG was destructive to a novel DNA-protein interaction

Bioinformatics analysis showed there are a number of cis-elements around -30 position recognized by the known transcription factors, ETS [17], CREB [12, 18], ATF [19] AP-1 [20] and SOX9 [21] (Fig. 4A). As the -81 CpG locates outside any known cis-element, while -30 CpG dinucleotide resides at the overlapped regions between AP-1 and SOX9 consensus (Fig. 4A), we experimentally asked whether the crucial DNA protein interaction with the -30 CpG may indeed occur in cell. Both AP-1 and SOX9 sites fall into the category of the highest score, 100% for the core match, and >0.9 for matrix match (http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi) (Fig. 4B). Therefore, it is logical to assume that one of these two factors may be responsible for the biological consequences observed so far (Figs 1-3). To test this hypothesis, we used the gel mobility shift assay in conjunction with the sequence-specific competition to test the sequence binding capacity of the nuclear proteins of BEL-7402 cells for 1, the wild type sequence in the CpG island of the *MAGE-A1* gene (5' ACTGACTTGCGCATTGTGG-GGC-3', -37 to -19, MAGE07), 2, AP-1 consensus (5'-GTGTGATGACTCAGGTTTG-3') and 3, SOX9 consensus (5'-CCATTGTTCT-3'), respectively. As shown in Fig. 4C, there was a specific binding by the nuclear proteins from BEL-7402 cells, which was completely eliminated by a 100 fold excess of the cold unlabeled same oligonucleotides (MAGE07). The mutated oligonucleotide at -30 from the wild-type CpG to TpG (CG4) remains fully potent to abolish the binding between the MAGE07 and the nuclear proteins. However, the methylated MAGE07 oligonucleotides by Hha I methylase failed to compete, indicating that the observed sequence-specific interaction of DNA-protein is indeed sensitive to the methylated status of the -30 CpG in this particular setting. But the pr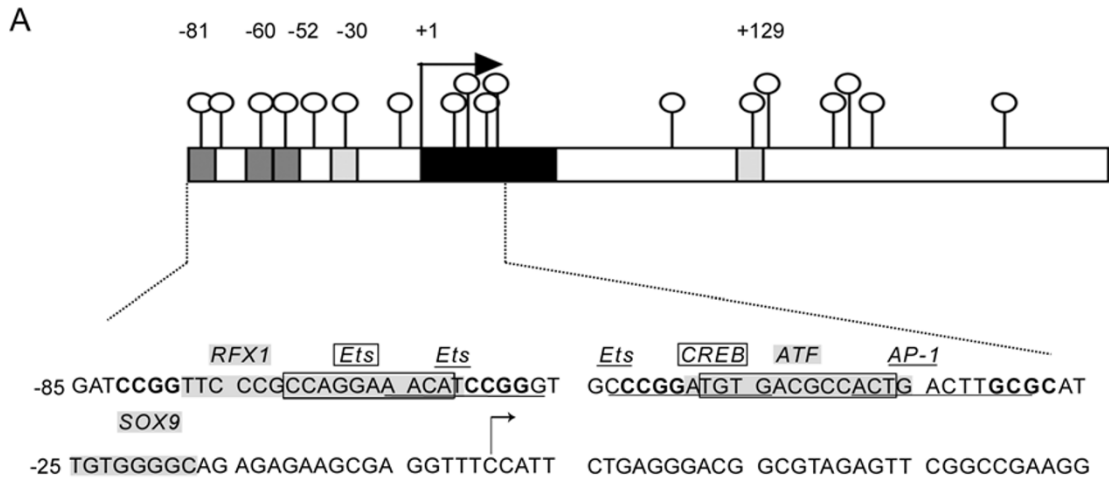otein identity of this interaction remains a mystery, as neither AP-1 nor SOX9 consensuses could eliminate the binding (Fig. 4C and 4F), suggesting that neither of these two known transcription factors are responsible. The mutant oligonucleotide (mAP-1), where the Ap-1 part was substituted, was capable to block the MAGE07 binding. However, the mutant oligonucleotide (mSOX9), where the SOX9 part was substituted failed to do so. These observations confirmed the sequence: 5'-GCGCATTGTGGGGC-3', as the crucial cis-element (Fig. 4C). The indication for the novel DNA-protein interaction was further supported by the band-shift analyses using the Ap1 and SOX9 probes, which showed that although the 100 fold excess of the cold probes efficiently eliminated the binding to the cognate hot probes, neither methylated nor unmethylated MAGE07 failed (Fig.4D and 4E).

DISCUSSION

The normal methylation pattern should be closely monitored and well-maintained throughout the life span of individual vertebrates [1]. Any disruption of this epigenetic homeostatic mechanism would result in various disease states, including cancer [2]. In addition to the global reduction of the methylation level, the local hypermethylation at the promoter CpG island has won the great recognition, as it represents an alternative mechanism for the inactivation of the tumor suppressors whose etiological roles in carcinogenesis have been well-established. The reverse process, namely, demethylation of the promoter CpG island linked to the tumor specific transcription activation has been rarely reported [4]. The *MAGE-A1* gene was reported transcriptionally reactivated in many tumors including HCC [5]. Co-incidentally, we found that its promoter CpG island was significantly demethylated in HCC, while the liver tissue from the healthy donor were fully methylated [4]. Therefore, the *MAGE-A1* gene is an ideal model system for the mechanistic study of the hepatoma associated demethylation and transcription activation.

The methylation status of the promoter CpG island correlates with expression profile of the *MAGE-A1* gene in liver cancer cell lines
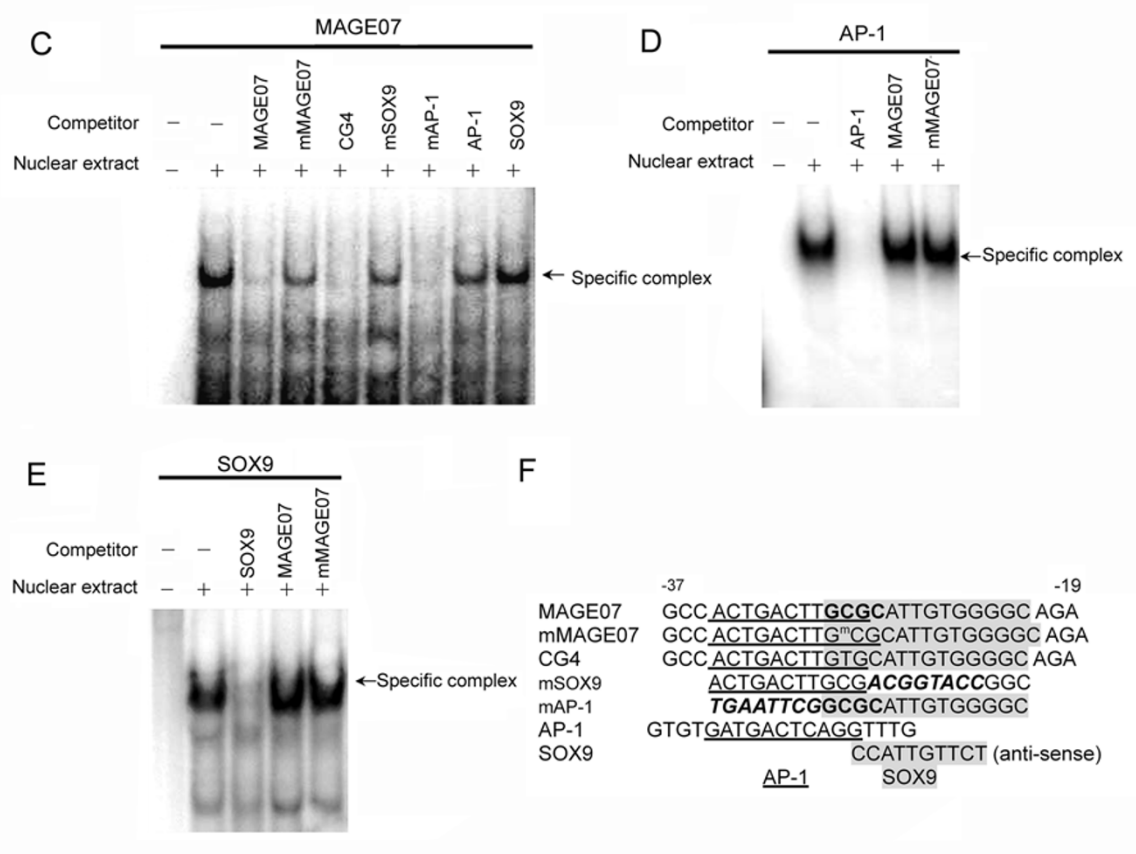
Learnt from the observed correlation of the *MAGE-A1* gene between DNA methylation and expression profile in the liver cancer tissues [4, 5], the similar correlation was expected in the liver cancer cell lines. Indeed, our analysis in four liver cancer cell lines was in a full agreement with the role. The homozygously methylated cell lines : Hep3B and HepG2 did not express the *MAGE-A1* gene, while the heterozygously methylated cell lines expressed a high level of *MAGE-A1* mRNA (Fig. 1 B and 1C). The methylated profile was also supportive (Fig. 1D). Partial demethylation

A

```
        -81   -60 -52 -30   +1                    +129
```

```
                    RFX1        Ets    Ets              Ets    CREB   ATF      AP-1
        -85 GATCCGGTTC CCGCCAGGAA ACATCCGGGT    GCCCCGGATGT GACGCCACTG ACTTGCGCAT
              SOX9
        -25 TGTGGGGCAG AGAGAAGCGA GGTTTCCATT    CTGAGGGACG GCGTAGAGTT CGGCCGAAGG
```

B

Transcription factor	Squence	Core match	Matrix match
AP1	acTGATtgcg rsTGACTnmnw	1.000	0.910
AP1	actgactTGCG rsTGACTmann	1.000	0.933
SOX9	gcgcATTGTggggc nncyATTGTtnnnn	1.000	0.902

R=G or A, M=A or C, S=G or C, Y=C or T, W=A or T, N=A or G or T or C

C


```
                              MAGE07
Competitor        -  -  MAGE07 mMAGE07 CG4 mSOX9 mAP-1 AP-1 SOX9
Nuclear extract   -  +    +      +      +    +     +    +    +
                                                    ← Specific complex
```

D


```
                      AP-1
Competitor        -  -  AP-1 MAGE07 mMAGE07
Nuclear extract   -  +   +     +      +
                                    ← Specific complex
```

E


```
                SOX9
Competitor        -  -  SOX9 MAGE07 mMAGE07
Nuclear extract   -  +   +     +      +
                                    ← Specific complex
```

F

```
                  -37                            -19
MAGE07    GCC ACTGACTTGCGCATTGTGGGGC AGA
mMAGE07   GCC ACTGACTTGᵐCGCATTGTGGGGC AGA
CG4       GCC ACTGACTTGTGCATTGTGGGGC AGA
mSOX9         ACTGACTTGCGACGGTACCGGC
mAP-1         TGAATTCGGCGCATTGTGGGGC
AP-1      GTGTGATGACTCAGGTTTG
SOX9                     CCATTGTTCT (anti-sense)
              AP-1                  SOX9
```

caused by the the specific DNA methyl transferase inhibitor, 5 aza-2'-deoxycytidine, did result in the reactivated transcription of the *MAGE-A1* gene (Fig. 1D).

Despite of the all or none type of expression of the endogenous *MAGE-A1* gene in the BEL-7402 verse HepG2 cells, the exogenously introduced *MAGE-A1* gene promoter remained active in HepG2 cells, only partly reduced in comparison with its activity in BEL-7402 cells (Fig. 4). This may suggest that the chromatin structure mediated control mechanisms are missing completely in the transient transfection assay system with supercoiled plasmid DNA. Therefore, to address this important aspect of the transcriptional control, alternative model system should be considered. However, the question as to the DNA methylation mediated transcriptional silencing remains approachable with the molecular biological methods adopted in this study.

The crucial CpG dinucleotide for transcription regulation of the *MAGE-A1* gene in cell

The primary goals of this study were to define the DNA methylation mediated control over the transcription at the level of the individual CpG dinucleotides within the promoter CpG island of the *MAGE-A1* gene. Three approaches were used, 1), the methylation profiling at the single CpG level in parallel with the expression status of the *MAGE-A1* gene in two liver cancer cell lines (Fig. 1); 2), the identification of the minimal promoter region as well as assessment of the methylation-status related promoter potency of the *MAGE-A1* gene, and 3), the sequence-specific DNA-protein interaction. The bioinformatics analysis was also integrated into each step of the whole process.

By bisuphite sequencing, the methylated status of each of 17 CpGs within the CpG island (-85 to +170) were assessed in both HepG2 and BEL-7402 cells. In HepG2 cells, the CpG at -43, -30, -8, and +14, +17, +26, +30 among 17 CpG were methylated either to 100% or 75%, correlating with the transcription silencing of this gene (Fig. 1C). The methylated allele in BEL-7402 cells assumed a slightly different pattern (Fig. 1C). Since BEL-7402 cells expressed the *MAGE-A1* gene, the impart of the methylation pattern of the methylated allele to the transcription

silencing can not be clearly defined. Nevertheless, the -30 CpG standed out for attention.

By deletion analysis, the minimal promoter was defined to -56 to +263 (Fig. 2B), as the -38 to +263 fragment has lost almost all the promoter potency in the transient transfection assay. To reconcile this observation with the obvious impart of the -30 CpG to the *MAGE-A1* promoter potency (Fig. 1C), the sequence upstream of the -38 may be required for the DNA methylation mediated mechanism to act most effectively.

Within the region spanning from -369 to +263, there were five CpGs embedded in either Hpa II or Hha I sites which needed to be further characterized (Fig. 3). By testing the methylated form of each mutants (CG1-5) where a single CpG of these five was individually mutated, the transcription potency for the -30 CpG was the most noticeable. Its methylation resulted in the most drastic reduction in promoter activity in comparison with the unmethylated counterpart in the transient transfection/reporter assay.

The most likely mechanism for the DNA methylation mediated effects over the transcription potency is that methylation may interrupt the otherwise occurring sequence (with the CpG within)-specific DNA-protein interaction(s). This was indeed the case, the methylated MAGE07 oligonucleotides failed to compete, while its unmethylated counterparts sucessfully competed the binding to the hot MAGE07 probe with the nuclear proteins from BEL-7402 cells (Fig. 4C). Two high scored cis-elements: AP-1 and SOX9 identified by the conventional bioinformatics analysis (http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi) were within this region; the -30 CpG resided at the overlapped strech of these two consensuses (Fig. 4B). Unexpectedly, neither consensus sequences could block the MAGAE07 binding and vice versa (Fig. 4C and 4D). These results not only suggested a novel methylation sensitive protein-DNA binding involved, but also casted serious doubt upon the value of the bioinformatics tools used in this work. To avoid any fortuitous change in the cis-elements bound by the known transcription factors in the database (Tab. 2 and Fig. 3C) while the site-specific mutagenesis to destroy the Hpa II site was intended, the bioinformatics analyses were

▶**Fig 4**, The sequence specific DNA-protein interactions at the -30 CpG containing sequence by the band-shift assay. (**A**) The sequence and the cis-elements bound by the known transcription factors predicted by bioinformatics analysis, within the region spanning from -85 to +35. Both Hha I site (GCGC) and Hpa II sites (CCGG) are highlighted by the bold font. (**B**) The consensus and the real sequence embracing the -30 CpG with the Core match:1 and Matrix match>0.9. (**C-E**) The mobility shift assay for the sequence specific binding between the nuclear proteins from the BEL-7402 cells and the MAGE07 oligonucleotide (-39 to -19, panel C), the AP-1 oligonucleotide (panel D) and the SOX9 oligonucleotide (panel E). 10 fM of the ^{32}P labeled double stranded oligonucleotides was incubated with 5 μg of nuclear proteins without (-) or with 100 molar excess (1 pM) of the cold double stranded oligonucleotides as indicated at the top of each autographs. The specific bands were specified with arrows at the right side of the autograph. (**F**) The sequence information of the competitor double stranded oligonucleotides.

intensively carried out. However, one of five mutagenesis was failed by this criterion, as the promoter potency of one of five CpG mutants (CpG2, -60) was unexpected lost. Therefore, additional precaution should be taken when such a cis-element prediction software is used.

The previous works done by Boon and his colleagues suggested that several elements at the upstream of the -30 were important for the promoter function in a similar assay system with a panel of human leukemia cell lines [9, 10]. For example, three ETS sites at the -70 to -44 (Fig. 4A) appeared to be important and its cognate transcription factor failed to bind the methylated form of the ETS consensus. In this connection, by testing directly the effects of the methylation on the promoter potency, we found none of these sites having any noticeable effects (Fig. 3C), while methylation of the CpG at -80 and -30, respectively, created drastic effects. Such a discrepancy might simply reflect the difference between the cell lines (leukemia verse liver cancer) or assay systems (deletion analysis verse the functional test for the DNA methylation in the transient transfection system).

Our work presented in this paper should pave the way for the further identification of the novel transcription factors binding to the -30 CpG containing sequence. Another interesting direction of the future study on the *MAGE-A1* gene system will concern the DNA methylation, histone modification and chromatin remodeling aspects of gene activation or silencing, a senario that may stars from the change in the methylation status of the promoter CpG island, or the -30 CpG.

ACKNOWLEDGMENTS

REFERENCE

1   Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Gene 2003; **33 Suppl**:245-54.
2   Baylin S, Bestor TH. Altered methylation patterns in cancer cell genomes: cause or consequence? Cancer Cell 2002; **1(4)**:299-305.
3   Neibergs HL, Hein DW, Spratt JS. Genetic profiling of colon cancer. J Surg Oncol 2002; **80(4)**:204-13.
4   Yu J, Zhang HY, Ma ZZ, et al. Methylation profiling of twenty four genes and the concordant methylation behaviours of nineteen genes that may contribute to hepatocellular carcinogenesis. Cell Res 2003; **13(5)**:319-33.
5   Mou DC, Cai SL, Peng JR, et al. Evaluation of MAGE-1 and MAGE-3 as tumour-specific markers to detect blood dissemination of hepatocellular carcinoma cells. Br J Cancer 2002; **86(1)**: 110-6.
6   van der Bruggen P, Traversari C, Chomez P, et al. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. Science 1991; **254(5038)**:1643-7.
7   Barker PA, Salehi A. The MAGE proteins: emerging roles in cell cycle progression, apoptosis, and neurogenetic disease. J Neurosci Res 2002; **67(6)**:705-12.
8   Serrano A, Garcia A, Abril E, et al. Methylated CpG points identified within MAGE-1 promoter are involved in gene repression. Int J Cancer 1996; **68(4)**:464-70.
9   De Smet C, Courtois SJ, Faraoni I, et al. Involvement of two Ets binding sites in the transcriptional activation of the MAGE1 gene. Immunogenetics 1995; **42(4)**:282-90.
10   De Smet C, De Backer O, Faraoni I, et al. The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. Proc Natl Acad Sci USA 1996; **93(14)**:7149-53.
11   De Smet C, Lurquin C, Lethe B, et al. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. Mol Cell Biol 1999; **19(11)**:7327-35.
12   Xu J, De Zhu J, Ni M, et al. The ATF/CREB site is the key element for transcription of the human RNA methyltransferase like 1(RNMTL1) gene, a newly discovered 17p13.3 gene. Cell Res 2002; **12(3-4)**:177-97.
13   Guo JY, Xu J, Mao Q, et al. The promoter analysis of the human C17orf25 gene, a novel chromosome 17p13.3 gene. Cell Res 2002; **12(5-6)**:339-52.
14   Cai L, Zhu JD. The tumor-selective over-expression of the human Hsp70 gene is attributed to the aberrant controls at both initiation and elongation levels of transcription. Cell Res 2003; **13(2)**:93-109.
15   Yu J, Ni M, Xu J, et al. Methylation profiling of twenty promoter-CpG islands of genes which may contribute to hepatocellular carcinogenesis. BMC Cancer 2002; **2(1)**:29.
16   Zhu JD. The myeloid-lineage specific enhancer of the mouse myeloperoxidase gene consists of three cis-elements defined as *in vitro* DNase I footprints. FEBS Lett 1994; **339(3)**:243-8.
17   Fisher RJ, Mavrothalassitis G, Kondoh A, Papas TS. High-affinity DNA-protein interactions of the cellular ETS1 protein: the determination of the ETS binding motif. Oncogene 1991; **6 (12)**:2249-54.
18   Garcia J, Wu F, Gaynor R. Upstream regulatory regions required to stabilize binding to the TATA sequence in an adenovirus early promoter. Nucleic Acids Res 1987; **15(20)**:8367-85.
19   Richter JD. *In vivo* photocrosslinking reveals that transcription factor binding to the mammalian ATF recognition sequence is required for E1A-induced transactivation in injected Xenopus laevis oocytes. Nucleic Acids Res 1989; **17(12)**:4503-16.
20   Struhl K. The JUN oncoprotein, a vertebrate transcription factor, activates transcription in yeast. Nature 1988; **332(6165)**:649-50.
21   Mertin S, McDowall SG, Harley VR. The DNA-binding specificity of SOX9 and other SOX proteins. Nucleic Acids Res 1999; **27(5)**:1359-64.