Characterization of regulatory sequences in alternative promoters of hypermethylated genes associated with tumor resistance to cisplatin

MOHAMMED A. IBRAHIM-ALOBAIDE[1], ABDELSALAM G. ABDELSALAM[2,3], HYTHAM ALOBYDI[4], KAKIL IBRAHIM RASUL[5], RUIWEN ZHANG[6] and KALKUNTE S. SRIVENUGOPAL[1]

[1]Department of Biomedical Sciences and Cancer Biology Research Center, School of Pharmacy, Texas Tech University Health Sciences Center, Amarillo, TX 79106, USA; [2]Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, Doha, Qatar; [3]Department of Statistics, Faculty of Economics and Political Sciences, Cairo University, Giza 12613, Egypt; [4]Biomedica, LLC, Sterling Heights, MI 48310, USA; [5]Weill Cornell Medical College and Hamad Medical Corporation, Doha, Qatar; [6]Department of Pharmaceutical Sciences and Cancer Biology Research Center, School of Pharmacy, Texas Tech University Health Sciences Center, Amarillo, TX 79106, USA

Abstract. The development of cisplatin resistance in human cancers is controlled by multiple genes and leads to therapeutic failure. Hypermethylation of specific gene promoters is a key event in clinical resistance to cisplatin. Although the usage of multiple promoters is frequent in the transcription of human genes, the role of alternative promoters and their regulatory sequences have not yet been investigated in cisplatin resistance genes. In a new approach, we hypothesized that human cancers exploit the specific transcription factor-binding sites (TFBS) and CpG islands (CGIs) located in the alternative promoters of certain genes to acquire platinum drug resistance. To provide a useful resource of regulatory elements associated with cisplatin resistance, we investigated the TFBS and CGIs in 48 alternative promoters of 14 hypermethylated cisplatin resistance genes previously reported. CGIs prone to methylation were identified in 28 alternative promoters of 11 hypermethylated genes. The majority of alternative promoters harboring CGIs (93%) were clustered in one phylogenetic subclass, whereas the ones lacking CGIs were distributed in two unrelated subclasses. Regulatory sequences, initiator and TATA-532 prevailed over TATA-8 and were found in all the promoters. B recognition element (BRE) sequences were present only in alternative promoters harboring CGIs, but CCAAT and TAACC were found in both types of alternative promoters, whereas downstream promoter element sequences were significantly less frequent. Therefore, it was hypothesized that BRE and CGI sequences co-localized in alternative promoters of cisplatin resistance genes may be used to design molecular markers for drug resistance. A more extensive knowledge of alternative promoters and their regulatory elements in clinical resistance to cisplatin is likely to usher novel avenues for sensitizing human cancers to treatment.

Introduction

The success of platinum drugs in the treatment of various types of cancer has been challenged by the hamper of intrinsic and acquired resistance. Cell and molecular biology research over several decades has provided insight into this problem and has demonstrated that alterations in drug influx and efflux, detoxification via glutathione and glutathione transferases and enhanced DNA repair (1-3) are all involved, either singly or in combination, in cisplatin resistance. However, it remains puzzling that genomic expression patterns of platinum-sensitive/-resistant determinants in several cell lines, including squamous cell carcinoma of the head and neck, bladder and ovarian cancers, do not always correlate with the results of biochemical analyses (4-6). Additionally, epigenetic changes in the DNA methylation profiles during cancer development are associated with the acquisition of cisplatin resistance (6-9). However, the problem of platinum resistance at the genomic and epigenomic levels has not been clearly defined. Therefore, it is paramount to gain a better understanding of gene expression machinery components that control platinum resistance genes. A promising approach to

these questions appears to be the characterization of regulatory elements, CpG islands (CGIs) and transcription factor-binding sites (TFBS) in alternative promoters of genes that have been validated and their association with cisplatin resistance proven.

It has been well established that alternative promoters are involved in the transcription of nearly half of all eukaryotic genes and are considered one of the main molecular characteristics of eukaryotic genomes (10,11). Overall, alternative promoters have been suggested to add significant flexibility and greater diversity to the regulation of gene expression (10-16). The functional significance of alternative promoters and their role in disease development and progression remains unclear, except for a few well characterized genes, such as the tumor suppressor p53, lymphoid enhancer-binding factor 1, insulin growth factor 2 and guanine nucleotide-binding protein α-stimulating genes, where these sequences are aberrantly activated, developmentally regulated or silenced (11). In addition, alternative promoters have been shown to regulate the expression of paired box homeotic gene 6 in various tissue types and play crucial roles in the development of the eye, brain, olfactory system and endocrine pancreas (17).

The TFBS and the CGIs targeted for DNA methylation, significantly affect the functional activities of alternative promoters (6-9,18,19). To date, the role and molecular characteristics of alternative promoters have not been investigated in platinum drug resistance. The present study mapped the distribution of TFBS and CGIs in 48 alternative promoters involved in the regulation of 14 hypermethylated genes, known to be associated with the development of cisplatin resistance.

Materials and methods

Alternative promoters of cisplatin resistance genes. A total of 14 genes known to contribute to cisplatin resistance and to be specifically hypermethylated in resistant cell lines, but not in the sensitive parental cell lines (9), were selected for this study. The genes were screened for the presence of alternative promoters and regulatory elements (CGIs and TFBS) within these sequences. The chromosomal map locations of cisplatin genes were searched in the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/) and the alternative promoters for each gene were searched in the Transcriptional Regulatory Element Database (TRED; https://cb.utdallas.edu/cgi-bin/TRED/tred.cgi?process=home). The length of the promoter sequence was set to 1,000 base pairs (bps), which were numbered along the sequence of the promoter relative to the transcription start site (TSS). The sequence of bps upstream the TSS was numbered with a negative sign, whereas downstream bases were identified with a positive sign.

Search for regulatory elements in the alternative promoters. The alternative promoters of 14 cisplatin resistance genes were analyzed for TFBS, namely, TATA-8 (TATAWAWR) and TATA-532 (HWHWWWWR, excluding HTYTTTWR, CAYTTTWR, MAMAAAAR and CTYAAAAR) elements, initiator (INR; YYANWYY), CCAAT and its inverted sequence TAACC, B recognition element (BRE; SSRCGCC) and downstream promoter element (DPE; RGWCGTG) sequences (20,21). CGIs in alternative promoters were searched in a 100-bp window (N=100) moving across the sequence at

1-bp intervals. The parameter sets used to search for CGIs in the alternative promoters were as follows: observed/expected (O/E) CpG ≥0.6 and %GC >55% (22-25). The O/E ratio for CpG was calculated according to the equation reported by Gardiner-Garden and Frommer (22).

Construction of phylogenetic tree. To investigate the evolutionary associations among the 48 alternative promoter sequences and to group the related sequences into specific categories, a phylogenetic tree was constructed using the Phylogenetic Tree application in the MATLAB Bioinformatics toolbox to probe the interrelationships and linking the tree nodes. The alternative promoter sequences and their derivatives were drawn using a hierarchical diagram.

Statistical analysis. The promoter sequences were analyzed using Perl, C++ and Excel software. The independence of each promoter element was examined using the Fisher's exact probability test.

Results

Hypermethylated genes in acquired cisplatin resistance. Chang *et al* (9) published an elegant study of specific genes that selectively undergo methylation in cisplatin-resistant cell lines, but not in their cisplatin-sensitive isogenic parent cell counterparts. The authors found that a set of 14 genes were silenced in the cisplatin-resistant cells and confirmed that promoter methylation, as determined by sodium bisulfite sequencing, accounted for the lack of gene expression. A number of these genes were also reactivated by azacytidine treatment of resistant cells (9). We used this panel of methylated genes for analysis of alternative promoters and their regulatory characteristics in cisplatin resistance. A list of these genes and their designations is provided in Table I. We also mapped the locations of these genes to their chromosomes and these sites are represented in Table I.

Genomic context of alternative promoters of cisplatin hypermethylated genes. A search through TRED identified 48 promoters for the 14 hypermethylated cisplatin resistance genes reported by Chang *et al* (9). These promoters were analyzed to identify the CGIs and TFBS (Fig. 1 and Table II). The number of identified alternative promoters for each gene varied between 2 and 6 (Fig. 1A) and were located on 9 chromosomes (Fig. 1B). Chromosome 1 and the long arm 'q' were found to host a higher number of genes and alternative promoters compared to other chromosomes and short arms (Fig. 1B). Of the 14 cisplatin resistance genes, 5 (35.71%) were located on chromosome 1, while of all the promoters investigated, 18 (37.5%) were located on this chromosome. However, the distributions of alternative promoters and genes were almost equal on forward and reverse strands of the chromosomes: 6 genes (43%) were located on the reverse strands and 8 genes (57%) on the forward strands (Fig. 1C); 26 promoters (54%) were on the reverse strands and 22 (46%) on the forward strands (Fig. 1D).

CGIs and TFBS. CGIs were identified in 28 alternative promoters of 11 cisplatin resistance genes, while 3 genes, namely the C4-binding protein β (C4BPB), chromosome 8

Table I. Description of functional activity and map locations for 14 genes associated with cisplatin resistance.

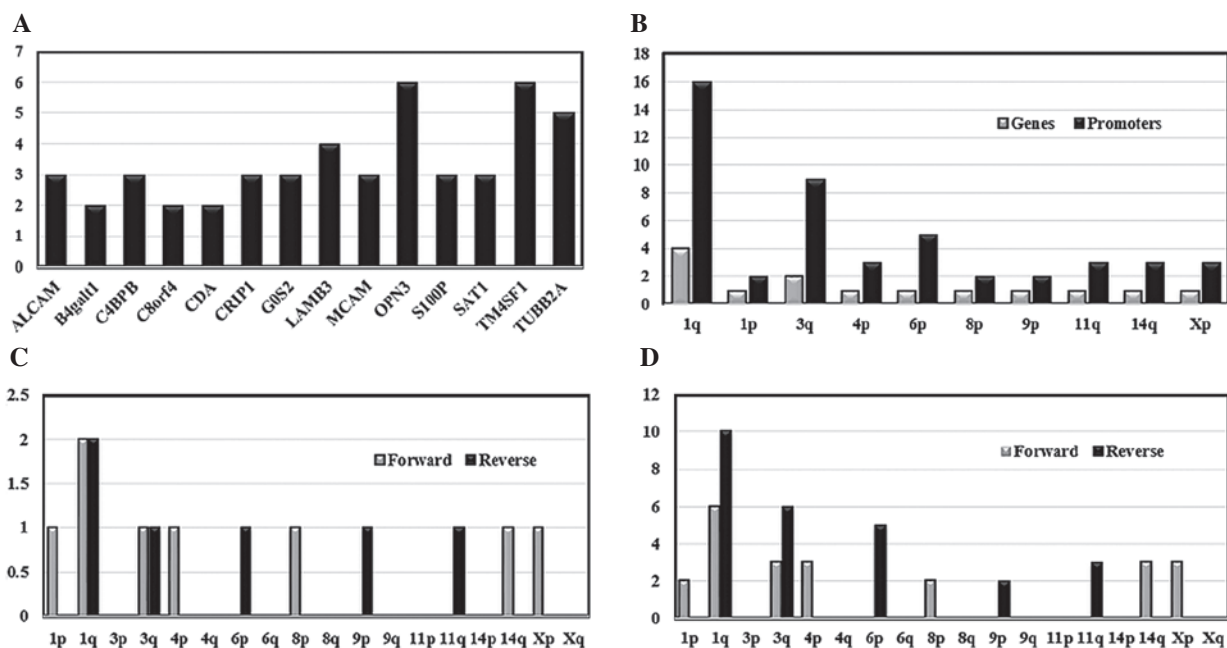| Genes	GenBank ID	Description	Map location
ALCAM	NM_001627	Activated leukocyte cell adhesion molecule	3q13.1, forward strand
B4GALT1	NM_001497	UDP-Gal:βGlcNAc β1,4-galactosyltransferase, polypeptide 1	9p13, reverse strand
C4BPB	NM_000716	C4-binding protein β	1q32, forward strand
C8orf4	NM_020130	Chromosome 8 open reading frame 4	8p11.2, forward strand
CDA	NM_001785	Cytidine deaminase	1p36.2-p35, forward strand
CRIP1	NM_001311	Cysteine-rich protein 1 (intestinal)	14q32.33, forward strand
G0S2	NM_015714	G0/G1 switch 2	1q32.2-41, forward strand
LAMB3	NM_000228	Laminin β3	1q32, reverse strand
MCAM	NM_006500	Melanoma cell adhesion molecule	11q23.3, reverse strand
OPN3	NM_014322	Opsin 3	1q43, reverse strand
S100P	NM_005980	S100 calcium-binding protein P	4p16, forward strand
SAT1	NM_002970	Spermidine/spermine N1 acetyltransferase 1	Xp22.1, forward strand
TM4SF1	NM_014220	Transmembrane 4 L six family member 1	3q21-q25, reverse strand
TUBB2A	NM_001069	Tubulin β2A class IIa	6p25, reverse strand

Figure 1. Analysis of 48 alternative promoters for the selected 14 genes associated with cisplatin resistance. (A) Number of alternative promoters for each gene. (B) Map locations of the 14 cisplatin genes and their alternative promoters on the chromosomes and the chromosome arms (p or q). (C) Number of cisplatin resistance genes on the p and q arms of the reverse and forward strands. (D) Number of alternative promoters of cisplatin genes on the reverse and forward strands of the p and q arms of chromosomes. ALCAM, activated leukocyte cell adhesion molecule; B4galt1, UDP-Gal:βGlcNAc β1,4-galactosyltransferase, polypeptide 1; C4BPB, C4-binding protein β; C8orf4, chromosome 8 open reading frame 4; CDA, cytidine deaminase; CRIP1, cysteine-rich protein 1; G0S2, G0/G1 switch 2; LAMB3, laminin β3; MCAM, melanoma cell adhesion molecule; OPN3, opsin 3; S100P, S100 calcium-binding protein P; SAT1, spermidine/spermine N1 acetyltransferase 1; TM4SF1, transmembrane 4 L six family member 1; TUBB2A, tubulin β2A class IIa.

open reading frame 4 (C8orf4) and cytidine deaminase (CDA) genes, exhibited no CGIs (Table II). As shown in Table II, the INR and TATA-532 sequences prevailed over TATA-8 sequences. We also observed that the alternative promoters of cisplatin resistance genes were rich in TATA-8 and TATA-532 sequences, although these varied significantly in their frequencies. For example, the four promoters of opsin 3 (OPN3), namely 44257, 2157, 114168 and 118881, which lack CGIs, contained 13-35 TATA-532 elements. Our analysis also demonstrated

that each of two promoters, melanoma cell adhesion molecule (MCAM)-117653 (lacked CGI) and S100P-31251 (harbored two CGIs), contained 10 TATA-8 sequences. Furthermore, the alternative promoters that harbored CGIs contained TATA-532 sequences in a wide range, from 1 to 22. The CCAAT and its inverted sequence TAACC, BRE and DPE sequences were not as frequent as INR and TATA-532 sequences. Another interesting observation was associated with BRE sequences: we noticed that the identified 48 BRE sequences were present
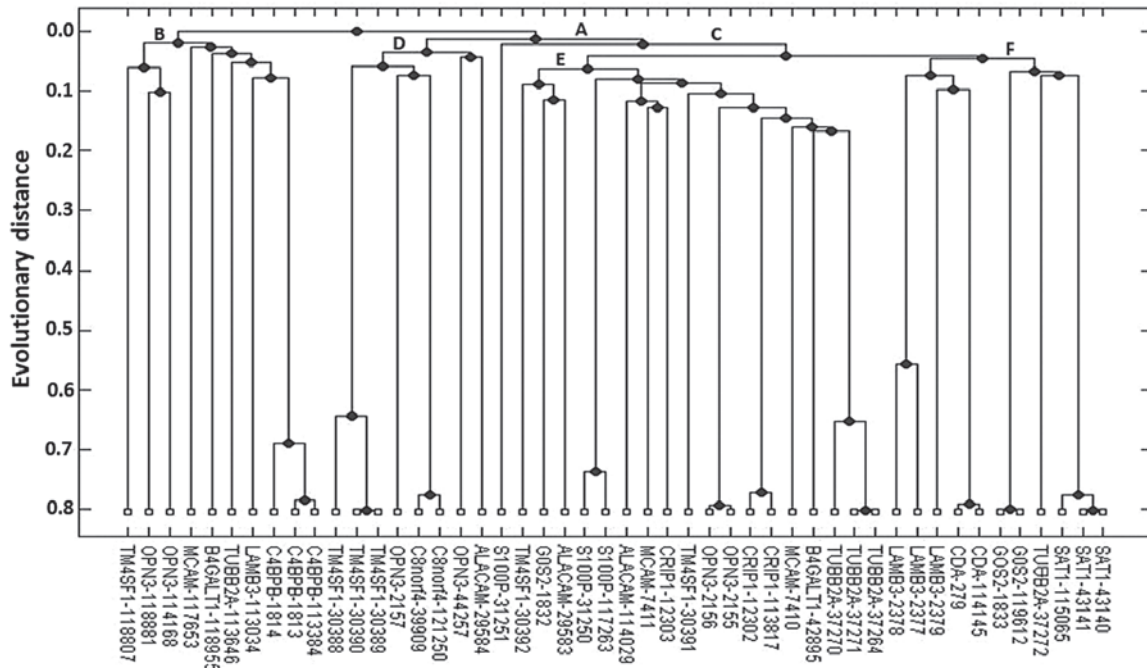
Figure 2. Phylogenetic tree generated for the 48 DNA sequences of the alternative promoters of 14 cisplatin resistance genes.

in 20 promoters of 9 cisplatin genes that harbored CGIs and were absent in other alternative promoters that lacked CGIs.

Phylogenetic tree analysis of alternative promoter sequences. A phylogenetic tree was created from a set of 48 alternative promoter sequences of the 14 cisplatin resistance genes (Fig. 2). Two main clustering patterns were observed, namely A and B, which may be further divided into subclasses and secondary classes within each category. The first main class (A) included 38 alternative promoters and the second main class (B) included 10 promoters. Furthermore, the first main class (A) may be divided into two subclasses, namely C and D; subclass C contained 30 alternative promoters, of which 26 contained CGIs. Groups E and F were identified in subclass C, which were represented by 19 and 11 alternative promoters, respectively. Group E included 18 alternative promoters with CGIs and 1 promoter (S100P-117263) without CGIs; group F included 8 alternative promoters with CGIs: SAT-43140, -43141 and -115065, TUBB2A-37272, G0S2-119612 and -1833 and LAMB3-2377 and -2378. The small subclass (D) contained 8 promoters, with 7 promoter sequences lacking CGIs: TM4SF1-30388, -30390 and -30389, OPN3-2157 and -44257 and C8norf4-39909 and -121250. Of the 10 alternative promoters in class B, 9 did not have CGIs: C4BPB-113384, -1813 and -1814, LAMB3-113646, B4GALT1-118955, MCAM-117653, OPN3-114168 and -118881 and TM4SF1-118807. The remaining promoter in this class, TUBB2A-113646, was characterized by the presence of 22 sequences of TATA-532 type and harbored CGIs. The evolutionary distances between the two major classes A and B and subclasses C and D were in the range of 0.0125 and 0.03125, respectively, exhibiting weak evolutionary ties between them; however, stronger ties were observed among the alternative promoters in group E and F within subclass C,

which were composed mainly from alternative promoters with CGIs (93%). The alternative promoters that lacked CGIs were distributed in two unrelated groups, B and D.

Discussion

We selected a panel of genes specifically known to undergo methylation in cisplatin-resistant cell lines (9) and characterized the alternative promoters and regulatory sequences associated therewith. The exact mechanisms by which these genes contribute to cisplatin resistance has not been fully elucidated; however, a number of these were shown to be inducible by cisplatin treatment in sensitive cells and silenced in sensitive isogenic cells (9). Recent findings support a prominent role for alternative promoters in cell type and human tissue type-specific gene expression (26). Considering that the transcription machinery utilizes alternative promoters for regulating differential transcription (10,16) and the aberrant use of one alternative promoter over another may result in disease, including cancer (11), we hypothesized that cisplatin resistance may be mediated by a differential usage of alternative promoters with variable regulatory sequences, TFBS and CGIs. Transcription factors and their binding sites in a given promoter are key elements in controlling the rate and extent of mRNA synthesis (19,27). However, the interaction between transcription factors and *cis*-regulatory modules, which contain the TFBS in promoter sequences, has not been clearly determined (27-33). Seven types of TFBS, namely INR, TATA-8, TATA-532, BRE, DPE, CCAAT and TAACC are well recognized (20,21). Our results demonstrated that 11 alternative promoters have TATA-8 sequences and 47 promoters contain TATA-532 sequences. It has been reported that ~76% of human core promoters lack TATA-like elements (20) and only 10-20% of promoters contain the TATA sequences (34). Of note, our study was not restricted to core

Table II. TFBS and CGIs in the 48 alternative promoters of 14 genes associated with cisplatin resistance.

Genes[a]	Promoter ID	INR	T8	T532	CCAAT	TAACC	BRE	DPE	CGIs
ALCAM	114029	5	0	5	1	0	1	1	+
	29583	7	0	7	1	0	1	0	+
	29584	11	0	8	0	0	0	0	+
B4GALT1	42895	6	0	1	0	2	4	0	+
	113955	5	0	18	0	0	0	0	-
C4BPB	1813	12	0	15	0	1	0	0	-
	1814	11	0	11	0	1	0	0	-
	113384	12	0	15	0	1	0	0	-
C8orf4	39909	10	0	10	0	0	0	0	-
	121250	10	0	10	0	0	0	0	-
CDA	279	12	0	5	0	0	0	0	-
	114145	12	0	5	0	0	0	0	-
CRIP1	12302	4	0	1	0	1	2	0	+
	12303	0	0	1	0	0	1	0	+
	113817	4	0	1	0	1	2	0	+
G0S2	1832	2	0	7	0	0	1	0	+
	1833	3	0	13	0	1	1	0	+
	119612	3	0	13	0	1	1	0	+
LAMB3	2377	7	1	6	0	1	0	0	+
	2378	9	1	8	0	1	0	0	+
	2379	13	1	13	3	3	0	0	-
	113034	6	0	7	1	1	0	0	-
MCAM	7410	4	0	0	0	0	3	0	+
	7411	4	0	2	0	2	0	0	+
	117653	5	10	45	0	0	0	0	-
OPN3	2156	7	0	3	1	0	5	0	+
	44257	12	0	13	0	0	0	0	-
	2155	7	0	3	1	0	5	0	+
	2157	14	0	28	1	1	0	0	-
	114168	8	2	27	2	1	0	0	-
	118881	13	1	35	1	0	0	0	-
S100P	31250	6	0	4	0	0	0	0	+
	31251	4	10	50	0	0	0	0	+
	117263	3	0	4	0	0	0	0	-
SAT1	43141	6	0	9	1	1	3	0	+
	43140	6	0	9	1	1	3	0	+
	115065	6	0	9	1	1	3	0	+
TM4SF1	30388	8	0	15	2	0	0	0	-
	30389	8	0	14	2	0	0	0	-
	30390	7	0	14	2	0	0	0	-
	30391	10	0	2	0	2	0	0	+
	30392	11	0	6	0	0	1	0	+
	118807	7	1	33	2	0	0	0	-
TUBB2A	37269	6	1	1	0	0	3	0	+
	37270	6	1	1	0	0	3	0	+
	37271	6	1	1	0	0	3	0	+
	37272	10	0	14	1	0	0	0	+
	113646	8	0	22	0	0	2	0	+

[a]For gene description, see Table I. TFBS: T8, TATAWAWR; T532, HWHWWWWR (excluding HTYTTTWR, MAMAAAAAR and CTYAAAAR); BRE, SSRCGCC; DPE, RGWCGTG; INR, YYANWYY; CCAAT and its inverted sequence TAACC (20,21). Parameter sets used to search for CGIs in the alternative promoters: observed/expected CpG ≥0.6 and %GC >55% (22-25). + and - indicate presence and absence of CGIs. TFBS, transcription factor-binding sites; CGIs, CpG islands; INR, initiator; T8, TATA-8; T532, TATA-532; BRE, B recognition element; DPE, downstream promoter element.

promoters, which are located ~40 bp up- and downstream of the TSS (35), but included the entire length of 1,000 bp that included 700 bp upstream and 300 bp downstream of TSS, as given by TRED. Furthermore, our results demonstrated that CCAAT, TAACC, BRE and DPE sequences were not as frequent as TATA sequences. Seventeen alternative promoters (35.4%) contained CCAAT; 19 (39.6%) contained TAACC, 20 (41.6%) contained BRE and one (2%) contained DPE. The percentage obtained for CCAAT was similar to that reported previously, reflecting its ubiquity in mammalian promoters (21,36). As the mechanism associated with cisplatin resistance is multifactorial (1-3), our results suggest that the genes encoding cisplatin resistance harbor more than one option to initiate their transcripts to enhance the efficiency of mRNA production.

It has been well documented that CGIs are prone to DNA methylation and this epigenetic mechanism is associated with gene silencing, initiation and maintenance of malignancy and drug resistance (6-9,37-39). Hypermethylation in CGIs and subsequent gene inactivation contributes to cisplatin resistance (9), as described earlier in this study. In the present study, we demonstrated that 11 out of the 14 genes reported by Chang *et al* (9) have 28 promoters harboring CGIs. However, the remaining 3 genes, namely C4BPB, C8orf4 and CDA, lacked discernible CGIs, raising the possibility of other mechanisms. The majority of alternative promoters harboring CGIs (93%) were clustered in one phylogenetic subclass, indicating relatively strong evolutionary ties among them. Thus, it may be hypothesized that the use of promoters with less frequent DNA methylation hotspots may assist the genes in escaping the epigenetic modifications and enable continued and efficient expression. Furthermore, we demonstrated that BRE and CGI sequences co-localized in the alternative promoters of cisplatin resistance genes (Table II) and this property may be utilized to design molecular markers for resistance to cisplatin and/or drugs. Such an approach will involve designing specific primers for amplifying DNA sequences that include BRE and a downstream segment of alternative promoter and a short exon region of the gene in question. A more detailed understanding of the functionality of the alternative promoters in cisplatin resistance is likely to have other applications; for example, it may be possible to design novel anticancer drugs that interact with specific promoter sequences, e.g., G-rich sequences which may form DNA G-quadruplexes (40-42). In addition, future research may make use of novel synthetic biology methods to build molecular models of various components of alternative promoters (43-46) to overcome drug resistance and enable more effective and personalized cancer therapies.

Acknowledgements

References

1. Siddik ZH: Cisplatin: mode of cytotoxic action and molecular basis of resistance. Oncogene 22: 7265-7279, 2003.
2. Kelland L: The resurgence of platinum-based cancer chemotherapy. Nat Rev Cancer 7: 573-584, 2007.
3. Rabik CA and Dolan ME: Molecular mechanisms of resistance and toxicity associated with platinating agents. Cancer Treat Rev 33: 9-23, 2007.
4. Akervall J, Guo X, Qian CN, *et al*: Genetic and expression profiles of squamous cell carcinoma of the head and neck correlate with cisplatin sensitivity and resistance in cell lines and patients. Clin Cancer Res 10: 8204-8213, 2004.
5. Tsunoda T, Koga H, Yokomizo A, *et al*: Inositol 1,4,5-trisphosphate (IP3) receptor type1 (IP3R1) modulates the acquisition of cisplatin resistance in bladder cancer cell lines. Oncogene 24: 1396-1402, 2005.
6. Zeller C, Dai W, Steele NL, *et al*: Candidate DNA methylation drivers of acquired cisplatin resistance in ovarian cancer identified by methylome and expression profiling. Oncogene 31: 4567-4576, 2012.
7. Zhang YW, Zheng Y, Wang JZ, *et al*: Integrated analysis of DNA methylation and mRNA expression profiling reveals candidate genes associated with cisplatin resistance in non-small cell lung cancer. Epigenetics 9: 896-909, 2014.
8. Guo R, Wu G, Li H, *et al*: Promoter methylation profiles between human lung adenocarcinoma multidrug resistant A549/cisplatin (A549/DDP) cells and its progenitor A549 cells. Biol Pharm Bull 36: 1310-1316, 2013.
9. Chang X, Monitto CL, Demokan S, *et al*: Identification of hypermethylated genes associated with cisplatin resistance in human cancers. Cancer Res 70: 2870-2879, 2010.
10. Batut P, Dobin A, Plessy C, Carninci P and Gingeras TR: High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res 23: 169-180, 2013.
11. Davuluri RV, Suzuki Y, Sugano S, Plass C and Huang TH: The functional consequences of alternative promoter use in mammalian genomes. Trends Genet 24: 167-177, 2008.
12. Ayoubi TA and Van De Ven WJ: Regulation of gene expression by alternative promoters. FASEB J 10: 453-460, 1996.
13. Jacox E, Gotea V, Ovcharenko I and Elnitski L: Tissue-specific and ubiquitous expression patterns from alternative promoters of human genes. PLoS One 5: e12274, 2010.
14. Schibler U and Sierra F: Alternative promoters in developmental gene expression. Annu Rev Genet 21: 237-257, 1987.
15. Liu H, Zhang Y, Li S, Yan Y and Li Y: Dynamic regulation of glutamate decarboxylase 67 gene expression by alternative promoters and splicing during rat testis maturation. Mol Biol Rep 37: 3111-3119, 2010.
16. Xin D, Hu L and Kong X: Alternative promoters influence alternative splicing at the genomic level. PLoS One 3: e2377, 2008.
17. Elso C, Lu X, Weisner PA, Thompson HL, *et al*: A reciprocal translocation dissects roles of Pax6 alternative promoters and upstream regulatory elements in the development of pancreas, brain, and eye. Genesis 51: 630-646, 2013.
18. Smith RP, Lam ET, Markova S, Yee SW and Ahituv N: Pharmacogene regulatory elements: from discovery to applications. Genome Med 4: 45, 2012.
19. Cusanovich DA, Pavlovic B, Pritchard JK and Gilad Y: The functional consequences of variation in transcription factor binding. PLoS Genet 10: e1004226, 2014.
20. Yang C, Bolotin E, Jiang T, Sladek FM and Martinez E: Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene 389: 52-65, 2007.
21. Hsu PC, Chao CC, Yang CY, *et al*: Diverse Hap43-independent functions of the *Candida albicans* CCAAT-binding complex. Eukaryot Cell 12: 804-815, 2013.
22. Gardiner-Garden M and Frommer M: CpG islands in vertebrate genomes. J Mol Biol 196: 261-282, 1987.
23. Han L and Zhao Z: CpG islands or CpG clusters: how to identify functional GC-rich regions in a genome? BMC Bioinformatics 10: 65, 2009.
24. Zhao Z and Han L: CpG islands: algorithms and applications in methylation studies. Biochem Biophys Res Commun 382: 643-645, 2009.
25. Hackenberg M, Barturen G, Carpena P, Luque-Escamilla PL, Previti C and Oliver JL: Prediction of CpG-island function: CpG clustering vs. sliding-window methods. BMC Genomics 11: 327, 2010.
26. FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest AR, Kawaji H, *et al*: A promoter-level mammalian expression atlas. Nature 507: 462-470, 2014.
27. Ben-Tabou de-Leon S and Davidson EH: Gene regulation: gene control network in development. Annu Rev Biophys Biomol Struct 36: 191-212, 2007.

28. Jeziorska DM, Jordan KW and Vance KW: A systems biology approach to understanding *cis*-regulatory module function. Semin Cell Dev Biol 20: 856-862, 2009.
29. Cameron RA and Davidson EH: Flexibility of transcription factor target site position in conserved *cis*-regulatory modules. Dev Biol 336: 122-135, 2009.
30. Vavouri T and Elgar G: Prediction of *cis*-regulatory elements using binding site matrices - the successes, the failures and the reasons for both. Curr Opin Genet Dev 15: 395-402, 2005.
31. Maston GA, Evans SK and Green MR: Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 7: 29-59, 2006.
32. Maston GA, Zhu LJ, Chamberlain L, Lin L, Fang M and Green MR: Non-canonical TAF complexes regulate active promoters in human embryonic stem cells. Elife 1: e00068, 2012.
33. Jaksik R and Rzeszowska-Wolny J: The distribution of GC nucleotides and regulatory sequence motifs in genes and their adjacent sequences. Gene 492: 375-381, 2012.
34. Martinez E: Core promoter-selective coregulators of transcription by RNA polymerase II. Transcription 3: 295-299, 2012.
35. Gross P and Oelgeschläger T: Core promoter-selective RNA polymerase II transcription. Biochem Soc Symp 73: 225-236, 2006.
36. Bucher P: Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol 212: 563-578, 1990.
37. Camacho-Arroyo I, Hansberg-Pastor V and Rodríguez-Dorantes M: DNA methylation analysis of steroid hormone receptor genes. Methods Mol Biol 1165: 89-98, 2014.
38. Cocozza S, Scala G, Miele G, Castaldo I and Monticelli A: A distinct group of CpG islands shows differential DNA methylation between replicas of the same cell line in vitro. BMC Genomics 14: 692, 2013.
39. Tanaka T, Bai T, Toujima S, *et al*: Demethylation restores SN38 sensitivity in cells with acquired resistance to SN38 derived from human cervical squamous cancer cells. Oncol Rep 27: 1292-1298, 2012.
40. Yang D and Okamoto K: Structural insights into G-quadruplexes: towards new anticancer drugs. Future Med Chem 2: 619-646, 2010.
41. Chen Y and Yang D: Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. Curr Protoc Nucleic Acid Chem 17: Unit17.5, 2012.
42. Hudson JS, Brooks SC and Graves DE: Interactions of actinomycin D with human telomeric G-quadruplex DNA. Biochemistry 48: 4440-4447, 2009.
43. Meng H, Wang J, Xiong Z, Xu F, Zhao G and Wang Y: Quantitative design of regulatory elements based on high-precision strength prediction using artificial neural network. PLoS One 8: e60288, 2013.
44. Brewster RC, Jones DL and Phillips R: Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. PLoS Comput Biol 8: e1002811, 2012.
45. Hosseinpour B, Bakhtiarizadeh MR, Khosravi P and Ebrahimie E: Predicting distinct organization of transcription factor binding sites on the promoter regions: a new genome-based approach to expand human embryonic stem cell regulatory network. Gene 531: 212-219, 2013.
46. Lai FJ, Chiu CC, Yang TH, Huang YM and Wu WS: Identifying functional transcription factor binding sites in yeast by considering their positional preference in the promoters. PLoS One 8: e83791, 2013.