Send Orders for Reprints to reprints@benthamscience.ae

260 *Current Bioinformatics,* **2018**, *13*, 260-272

RESEARCH ARTICLE

In-silico Evidences of Regulatory Roles of WT1 Transcription Factor Binding Sites on the Intervening Sequences of the Human *Bcl-2* Gene

Gholampour-Faroji Nazanin[1,3], Haddad-Mashadrizeh Aliakbar[1,2,*], Mirahmadi Mahdi[4,5], Monhemi Hassan[3], Shahreki-Mojahed Safoora[3], Bahrami Ahmad Reza[1,2], Housaindokht Mohammadreza[1], Matin M. Maryam[2,6], Zomorodipour Alireza[7], Dadjoo Parisa[3] and Nohtani Mahdi[2]

[1]Recombinant Proteins Research Group, Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad, Iran; [2]Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran; [3]Structural Biology and Bioinformatics Research Group, Khayyam Bioeconomy Institute (KBI), Mashhad, Iran; [4]Department of Cancer Research, Nastaran Center for Cancer Prevention (NCCP), Mashhad, Iran; [5]Stem Cell and Regenerative Medicine Research Group, Iranian Academic Center for Education, Culture and Research (ACECR), Khorasan Razavi Branch, Mashhad, Iran; [6]Novel Diagnostics and Therapeutics Research Group, Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad, Iran and [7]National Institute for Genetic Engineering and Biotechnology, Tehran, Iran

Abstract: ***Background:*** Intervening sequences (introns) have significant effects on genomic regulations and molecular evolution. So, it deserves a deeper analysis for better understanding the possible regulatory roles of these regions.

Objective and Method: Accordingly, the intron 2 (In-2) of the human B-cell lymphoma 2 *(hBcl2*) gene, with regard to the size of the In-2 as well as critical roles of the gene in the homeostatic of the cellular balance, was analyzed by using *in-silico* approaches to identify In-2 transcription factor binding (In2-TFBs) motifs.

Results: Our analysis revealed 966 motifs of 118 different TFBs types which were scattered throughout both the strands of the complete sequence of the gene, in particular on the In-2, with significant pattern of distribution and repetition. Distribution pattern of these motifs revealed that most of them were accumulated in narrow regions of the In-2, far from the area of the splicing sites. Moreover, it was observed that except for WT1-TFBs, Gfi-1-TFBs, GAGA-TFBs, all other motifs were sporadic, with irregular and random distribution. Among these motifs, WT1-TFBs showed the highest frequencies which were situated in four neighboring regions of the In-2, by a close linear relationship to Sp1-TFBs. Furthermore, the sequence logos of the WT1-TFBs showed that they ranged in size from 22 up to 45 bps and were enriched with G and T nucleotides. Meanwhile, the binding affinity of WT1-TF to WT1-TFBs revealed significant differences compared to the other sequences of the gene as negative control.

Conclusion: In general, this data provides supporting evidences for the existence of regulatory regions in the intronic sequences of the *hBcl2* gene especially in the In-2, and also represents new targets for WT1-TF which might contribute to *hBcl2* regulation and apoptosis process.

ARTICLE HISTORY

Received: June 3, 2016
Revised: February 15, 2017
Accepted: April 10, 2017

DOI:
10.2174/1574893612666170428114551

Keywords: *hBcl2* gene, intron, gene regulation, TFBs, WT1.

1. INTRODUCTION

Programmed cell death (PCD), and its morphological equivalent "apoptosis" play an important role in a wide variety of physiological processes, and also during the development of cancer [1-5]. PCD is carried out in a complicated regulatory process by variety of genes. *Bcl-2* family, with 25 known genes including both anti- and pro-apoptotic members, is a major regulator of the programmed cell death machinery with opposing biological functions [4, 6-8]. Defects in this machinery could lead to evasion of PCD and contribute to many pathological conditions during development and also in the progression of cancer [5, 9]. Hence, resistance of malignant cells to anti-cancer agents may be, in some cases, due to dysregulation of apoptotic pathways, *e.g*. *Bcl2* or *Iap* over expression [4, 10-11]. In this regard, an increased expression of anti-apoptotic proteins, *e.g*. BCL2 is regularly found in malignant cells, contributing to their clonal expansion by conferring increased survival ability [9, 12]. *Bcl2* is the first member of this family which was discovered and found to be an oncogene. Aberrant expression of *Bcl2* con-

*Address correspondence to this author at the Recombinant Proteins Research Group, Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran; Tel: +985138803796; Fax: 00985118762227; E-mail: a.haddad@um.ac.ir

1574-8936/18 $58.00+.00

tributed in numerous known diseases such as melanoma, chronic lymphocytic leukemia, breast, lung and bladder carcinoma, as well as in schizophrenia and autoimmunity via overriding apoptosis [13-17]. It has been shown that the BCL2 modulates cellular viability through regulating intracellular calcium homeostasis, cellular redox state, lipid peroxidation, as well as cytochrome c released from mitochondria by protein-protein interaction [18-21]. Accordingly, it has been demonstrated that introduction of *Bcl2* gene into the eukaryotic cells protects the recipient cells from apoptosis [22]. However, in stressed situations, *Bcl2* expression apparently fails to protect cells from apoptosis [22]. Therefore, simple expression of *Bcl2* may not be enough to protect cells [17, 23]. These data suggest that there are other unknown pathways that have impact on *Bcl2* expression. In this regard, up-regulation of *bcl2* in some malignancies has been attributed to posttranscriptional and posttranslational modifications, including poly(A) tail sequences context, BCL2 phosphorylation and protein-protein interactions [17, 20]. On the other hand, it is has not yet been known whether the regulation of *blc2* and PCD machinery might be controlled by intervening sequence of this gene. Numerous experimental analyses revealed that intronic sequence context and structures (ISCSs) and intronic derived regulatory elements (IDREs) have significant effects on post genomic regulations and molecular evolution in a broad range of organisms including nematodes, insects and mammals [24-31]. Hence, it would be essential to understand which intronic regions of *Bcl2* gene are potentially involved in gene regulation. In this regard, the present study was designed to investigate the sequence context of the In-2 of the *hBcl2* gene in detail, in order to underpin regions with potential regulatory elements. The *hBcl-2* gene located at the long arm of 18 chromosomes, spans over 196 kbp in length and similar to most eukaryotic genes is interrupted by intervening sequences, including 3 exons and 2 introns [32]. In-2 of this gene with 189322 bp length is one of the biggest intervening sequences among all eukaryotic genes which occupies 96% length of the *hBcl2* gene. Therefore, involvement of this sequence on gene regulation is inevitable, and detection of regulatory areas in this region consequently would shed light on details of *hBcl2* gene regulation and may be used toward development of diagnostic and therapeutic tools.

2. MATERIALS AND METHODS

2.1. Sequence Extraction

The complete nucleotide sequence of the *hBcl2* gene was retrieved from Gene Bank of National Center for Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov) with accession number NC_000018. On the other hand, sequences of 7 genes of human BCL2 family (*Bclx, Mcl1, Bclg, Bak, Bax, Bad* and *hmgb1*), 5 genes of human chromosome 18 (*Fech, Greb1L, Ncp1, Smad4, Kc6*), and sequences of human *Brca1*, *Brca2* and coagulation factor VIII (*hFVIII*) genes were retrieved from Gene Bank with different gene IDs including, 100196057, 4170, 79370, 578, 581, 572, 3146, 2235, 80000, 4864, 4089, 641516, 403437, 474180 and 2157, as negative controls. Moreover, the isoform A of Wilms tumor 1 (WT1) transcription factor (TF) retrieved with the accession number NP_000369, from Protein Database of NCBI.

2.2. Programs Used for *in-silico* Characterization

The complete nucleotide sequences of the *hBcl2* gene, and the complete sequences of all of the negative controls were scrutinized, with several *in-silico* programs, for detection of Cis-acting regulatory elements in their contexts. For each analysis, FASTA format of the sequences was used. The output was saved into a word document and/or a portable data file (PDF) for exact scrutiny and review. The names and functions of the programs used are:

1) NSITE: is a program for potentially disclosed TFBs throughout the selected sequences based on statistical estimation of expected number of a nucleotide consensus pattern in a given sequence. In the output of this program, we can gain a pattern of TFBs, Site ID as well as Site Name, its position in the sequence, description of the motif and corresponding binding factor name from the original database (if it exists) and also its expected value (*E*-Value) [33].

2) FPROM/Human promoter prediction: Presence of any promoter areas throughout the gene was investigated based on potential transcription start positions by linear discriminate function combining characteristics describing functional motifs and oligonucleotide composition of these sites using FPROM program [33].

3) GC content: (http://www.biologicscorp.com/tools)-GC content, which is usually calculated as a percentage value of GC-ratio, was calculated for each sequence based on Biologics Corp program.

4) CpG Plot: Given the significant relationship among the CpG islands, promoters and TFBs, any possible correlations among them were investigated using CpG Plot programs [34].

5) Sequence logos: The Sequence Logos of detected regulatory regions were generated using the WebLogo server [35], by depicting stacks of letters. The height of each letter within a stack is proportional to the base frequency at that position, and the letters are sorted by size, with the tallest (*i.e.* most frequent) on top. The height of the stack is the sequence conservation measured in bits of information; 1 bit measures the choice between two equally likely possibilities.

2.3. Data Validation

To discriminate between positive and false positive results, data validation was performed. In this regard, at the first step, data without references and *E*-values up to 10^{-3} were eliminated. Subsequently, the frequencies, varieties, distribution patterns as well as spatial connections of detected regulatory elements were investigated throughout the In-2 of the *hBcl2* gene, as another parameter of validation. On the other hand, the accuracy of selected data from target genes was then assessed via comparing the results obtained from similar sequences in GC content, as well as sequences of the related genes of the *hBcl2*, as negative control. Moreover, the selected regions of the gene, as WT1-TFBs areas, were then subjected to Molecular docking program in order to determine the binding affinity of WT1-TF to these regions. Molecular docking was performed by NPDock [36]. The

NPDock is designed for protein-nucleic acid complex structure modeling. NPDock is implemented as a computational workflow which consists of the GRAMM program [37], DARS-RNP and QUASIRNP statistical potentials [38], a counterpart of QUASI-RNP for scoring protein-DNA complexes (QUASI-DNP), and tools for clustering, selection and refinement of models [36]. In the first step, the GRAMM program is used to perform a rigid body global search for the generation of geometrically plausible protein-nucleic acid complex structures (decoys). The decoys are scored and ranked using statistical potentials. The best-scored decoys are then clustered and the largest clusters are selected. In the final step, a Monte Carlo Simulated Annealing procedure is used to optimize the protein-nucleic acid interactions [36]. In this regard, 3D structure of WT1-TF and regulatory regions were obtained and generated from Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) with PDB ID: 4R2S, and DNAtools program [39], respectively. Furthermore, the occurrences frequencies of the single nucleotide polymorphism (SNP) throughout the WT1-TFBs areas of the *hBcl2* gene were assessed as supplementary evidence of the importance of these regions. Accordingly, the SNP with occurrences in the *hBcl2* gene were gathered from dbSNP (http://www.ncbi.nlm.nih.gov/snp) and F-SNP (http://compbio.cs.queensu.ca/F-SNP) database, and then their location throughout the gene was investigated.

2.4. Statistical Analysis

Correlations among the frequencies and distribution patterns of TFBs as well as SNP throughout the selected genes and regions were investigated by Pearson's correlation coefficient. Moreover, the energy of binding affinity of WT1-TF to WT1-TFBs regions as well as negative control sequences were carried out from ten situations of the best interaction. The generated data was expressed as mean_standard deviation. One-way analysis of variance (ANOVA) was used to analyze the binding affinity of WT1-TF to WT1-TFBs. ANOVA was done by a Tukey post hoc test. Statistical significance was considered at $P < 0.001$. All statistical analyses were carried out with SPSS 22.0 (SPSS Inc., Chicago, IL, USA).

3. RESULT

3.1. TFBs Throughout the *hBcl2* Gene

Our survey led to the discovery of 966 motifs of 118 different types of TFBs scattered throughout the two strands of the complete sequence of the *hBcl2* gene. More in-depth analysis showed that most of these motifs are situated on negative strand of the In-2 of the gene with or without linear relationship to intronic promoter as well as CpG islands throughout the region (Fig. **1**). As illustrated in this figure, distribution patterns of these elements throughout the In-2 of the gene seem to be random without any significant relationship to each other, so that they are scattered with dependency to the length of the regions. However, the frequencies of these elements in deferent regions of the In-2 of the gene showed significant pattern of distribution. As shown in Fig. **1**, Modern most of the detected TFBs are situated in positions far from regions of the splicing sites of the intron, especially within range 100 up to 120 Kbp of the In-2, with a significant correlation to the intronic promoter as well as CpG islands in this area.

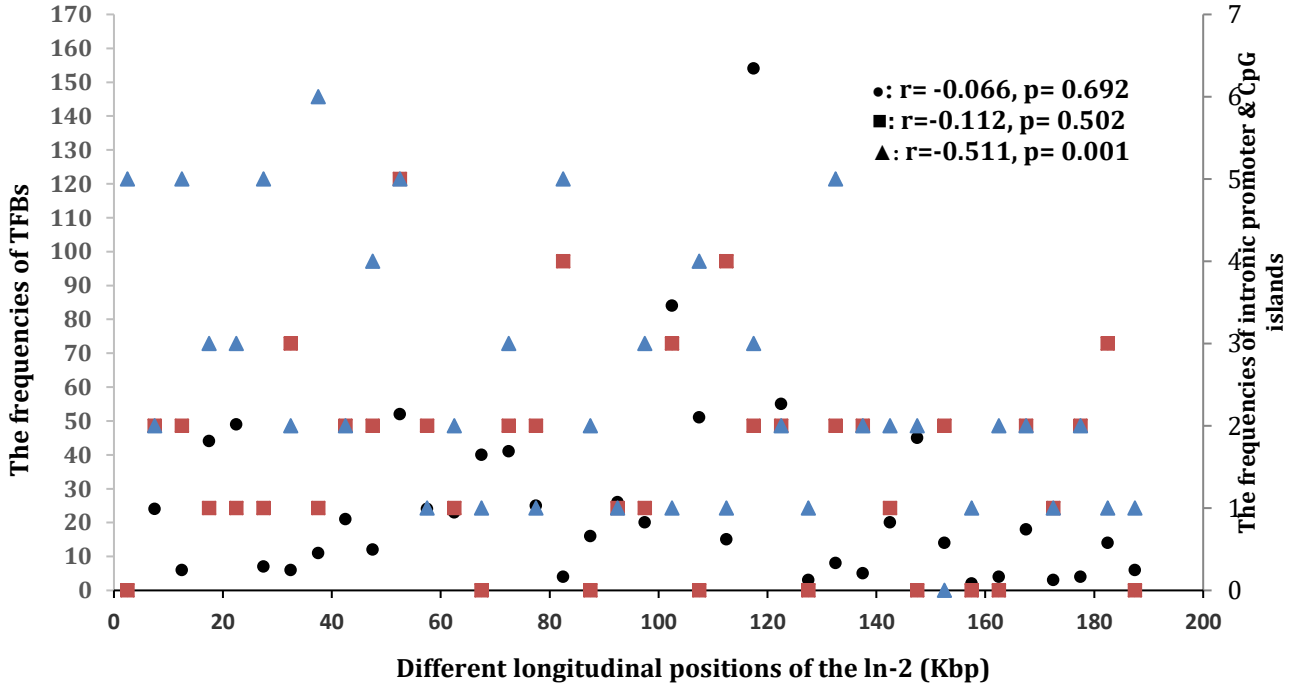

Fig. (1). Pearson correlation among distribution patterns and frequencies of TFBs, Intronic promoters as well as CpG islands throughout the In-2 of the *hBcl2* gene. ● = distribution and frequencies of TFBs. ■ = distribution and frequencies of intronic promoters regions. ▲ = distribution and frequencies of CpG islands. As shown in this figure, there is not any correlation between the frequencies of TFBs as well as intronic promoters to their occurrences in similar longitudinal positions of the ln-2 of the gene. However, the occurrences frequencies of CpG islands showed a relative correlation in the similar longitudinal positions of the ln-2 of the gene.

3.2. Screening the In-2-TFBs

In order to evaluate the accuracy of TFBs which are situated throughout the In-2 of the gene, detected motifs without citation and suitable *E*-Value have been eliminated. The results of this process led to remaining 549 motifs from 113 different types of TFBs (Table **1**). As shown in this table, about 84.5% of these motifs are situated on positive strand of the In-2 of the gene, and most of them have been repeated sporadicly, which could be a cause of incident detection. Nevertheless, some of them possess repetition of up to 141repeat such as WT1, Gfi-1, AT-Alu, CACA-pseudobeta-2, apoB-reducer-site-b, LUN RS, GAGA-V1bR, HIF-1-Noxa, RAP1-mdscan-motif, BRCA1-ERE and HIOMT-A-

Table 1. Variation and repetition of the In-2-TFBs on two strand of the In-2 of the *hBcl2* gene.

Site Name	R.	St.	Site Name	R.	St.	Site Name	R.	St
prolactin USII	1	-	Sp1-alpha-actin_(6)	1	+	LUN RS	18	+
CuE3.3	1	-	BP1-beta-globin_(1)	1	+	6-16-ISRE	1	-
AT-Alu	20	+	GAGA-eve_(2)	2	+	Pho5-UASp2	1	+
HTLV-pa-BS.3	1	-	Krox-20-Hox-2.3	2	-	STAT6-lymphotoxin-alpha	1	+
MHCI/H2K	2	+	WT1-selected genomic sequence3	50	-	HAND2/E12 CS	1	+
CACA-pseudobeta-2	26	+	WT1-selected genomic sequence4	50	+	MTIIA-MREa	1	+
LexA CS2	1	-	WT1-selected genomic sequence5	41	+	TH-dyad	1	-
GR-HMTIIa-II	1	+	KBF1-MHC class I H-2Kk	1	+	GAGA3-V1bR	12	+
Antp-Ubx	1	+	Sp1-cytochrome C	1	+	GAGA5-V1bR	23	+
eve RS2	1	+	C/EBP-HBV_(1)	1	+	ANT2-silencer-core	1	-
ftz RS1	1	+	A-gamma-globin-3_-enhancer-4	1	+	IA-1-IA-1	1	+
Ubx-Ubx	1	+	BCR-undefined-site-9	1	+	beta5-integrin-FP3	1	+
CP1-MLP.2	1	-	IRE site I	1	-	ApoD-APP	7	-
C/EBP-SV40 core-C	1	-	sTnI -undefined-site	1	+	TRvalpha2-68DR	1	+
Y box (1)	1	+	Ad-E3-kappa-2	1	-	aldolase-C-site-824	1	-
c-Myb-F1-L-2	1	+	albumin-undefined-site-2	1	+	Spi-1-scavenger-receptor	3	+
gERE	3	+	GATA-1-FP-4	1	+	HIF-1-Noxa	12	+
BGP1 RS2	2	+	C/EBP-TAT_(8)	1	+	ovIFNT-p7	1	-
HTF4-HIV-E-box-I	1	-	DHS B_(1)	1	+	PAX2-WT1-Site-III	1	-
AluA	6	+	distal region II/ D	1	-	CDP-dNRE-1	1	-
GBF-B3-DG17	1	+	HBP-1a-nitrate reductase	1	+	Gfi-1-ELA2-9	1	+
OV-220	1	-	uteroglobin-element-II	1	+	Gfi-1-AZU-1	19	+
GA1_(1)	1	+	PR-artificial sequence-1	1	+	Gfi-1-AZU-4	1	+
IR-site_C	3	+	c-Myb-HTLV-I_(3)	1	-	Gfi-1	16	+
GR-intron-site-1	1	-	GC box (WiF1)	1	+	Gfi-1-AAT	26	+
C-rich element	1	+	IL-2/NRE-A	1	-	Gfi-1-ACT	23	+
betae4	1	+	Sp1-prosaponin	3	+	Gfi-1-Jak3	14	+
DR+2_(2)	1	-	Sp1-alpha-dystrobrevin	1	+	Gfi-1-P21	15	+
mADA-site_A	1	+	HIOMT-A-E4	17	+	Flc-SOC1	1	+
apoB-reducer-site_b	11	+	PPAR-rApo-CIII	1	+	Pax6-c-maf-E	1	-
globin-HS2-GTGGG-site	1	-	CLC-GA-element	1	+	Barx2-Bx98(HBS1)	1	+
rice-actin-MPE	4	+	PTTG-Element-C	1	+	p53-MCK	1	+
TTF-2-TPO-Z	1	-	Sp1-COL2A1-2	1	-	Sp1-KDR/flk-1-IV	1	-
ETF-beta-actin_(1)	1	+	SFF-CLB2	1	+	rGH-T3RE-2	1	-
ETF-beta-actin_(2)	1	-	BRCA1-ERE	14	+	Nmp4-COL1A1-site-A	1	+
apoE-B1-III	1	+	ER-beta-ERE	1	-	Nmp4-COL1A1-site-B	1	+
GATA-1-gamma-globin_(1)	1	-	Alu-ERE	4	+	FAC1-PS-1	8	+
leghaemoglobin glb3-undefined	1	-	RAP1-mdscan-motif	17	+	FAC1-PS1-2	6	+

R: Repetition, St: strand

E4. As could be calculated from the table, among these elements, WT1-In2-TFBs, Gfi-1-In2-TFBs and GAGA-In2-TFBs with 141, 115 and 37 repetition have the highest frequencies.

3.3. Transcription Factors with Binding Site Throughout the In-2

Our investigation revealed that 60 different types of TFs could bind to In-2-TFBs (Table **2**). As shown in this table, corresponding TFBs of these elements originally situated on various genes of different organisms, however the originality of some of them were not determined. However, among these factors Sp1, WT1, CRX, ER, RAP1, LUN, GAGA, HIF-1alpha, Gfi-1, FAC1 showed the highest corresponding position with or without mismatch in the In-2 of the *hBcl2* gene.

3.4. Distribution Pattern of the In-2-TFBs with the Highest Frequency

As it is obvious in Table **1**, WT1-In2-TFBs, Gfi-1-In2-TFBs and GAGA-In2-TFBs have the highest repetitions throughout the In-2 of the gene. Distribution pattern of these elements throughout this region showed considerable schema (Fig. **2**). As shown in this figure, no correlation was seen between the occurrence frequencies of the Gfi-1-In2-TFBs to

Table 2. Characteristics of the TFs with binding sites on the *Bcl2* gene.

F. Name	Org/Species	Gene	N.	F. Name	Org/Species	Gene	N.
IgPE-1	M musculus	Ig mu	1	CRX	R norvegicus	HIOMT	3
fI/fII	HTLV-I	……	1	CRX	R norvegicus	HIOMT	17
LexA	Multiple	……	1	PPAR	……	……	1
GR	H sapiens	MTIIa	2	KKLF	R norvegicus	CLC-K1/CLC-K2	1
even-skipped	Artificial	……	1	SFF	S cerevisiae	CLB2	1
Ftz	……	……	1	ER	H sapiens	BRCA1	19
CP1	Adenovirus	major late	1	RAP1	……	……	17
C/EBP	R norvegicus	aminotransferase	3	LUN	……	……	18
c-Myb	……	c-myc	2	DRAF1/DRAF2	H sapiens	IFI 6-16	1
BGP1	……	……	2	Pho4	S cerevisiae	Pho5	1
HTF4	HIV-1	……	1	STAT6	H sapiens	lymphotoxin alpha	1
GBF	D discoideum	DG17	1	HAND2/E12	……	……	1
GA-BF	S cerevisiae	ura3	1	MTF-1	H sapiens	MT-IIA	1
Sp1	H sapiens	insulin receptor	12	rITF2/CDP2	R norvegicus	tyrosine hydroxylase	1
Antp	Ultrabithorax	……	1	GAGA	……	……	37
RAR-alpha1	……	……	1	IA-1	H sapiens	IA-1	1
TTF-2	R norvegicus	thyroperoxidase	1	RXR-alpha	……	……	1
ETF	G gallus	beta actin	2	Brn-1	R norvegicus	aldolase C	1
ETF	G gallus	beta actin		Spi-1	……	……	3
GATA-1	H sapiens	G-gamma globin	1	HIF-1alpha	H sapiens	Noxa	12
BP1	H sapiens	beta-globin	1	Ets-2	O aries	IFN-tau	1
Krox-20	……	……	2	CDP	MMTV genome	……	1
WT1	……	……	141	Gfi-1	……	……	114
KBF1	M musculus	H-2K(k)	1	Flc	A thaliana	SOC1	1
NF-kappaB	Adenovirus	E3	1	Pax6	R norvegicus	c-maf	1
HBP-1a	……	……	1	Barx2	……	……	1
Sp factors	O cuniculus	Uteroglobin	1	P53	……	……	1
PR	……	……	1	T3R	R norvegicus	GH	1
WiF1	……	……	1	Nmp4	R norvegicus	COL1A1	2
Nil-2a	……	……	1	FAC1	……	PS1	14

F: factor, N: number of corresponding TFBs.

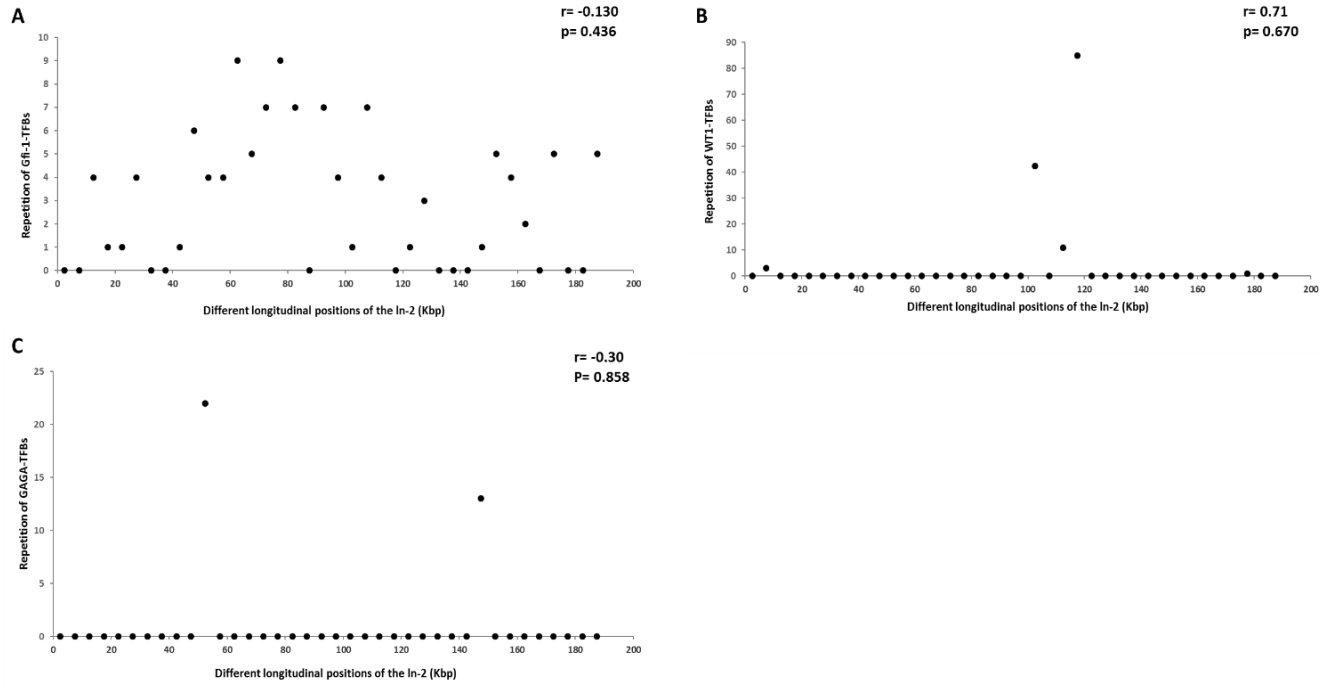

Fig. (2). Pearson correlation among distribution pattern and frequencies of the In-2-TFBs. Gfi1-TFBs (A), WT1-TFBs (B) and GAGA-TFBs (C) throughout the different longitudinal positions of the In-2 of the *hBcl2* gene. As shown in this figure, it did not indicated any Pearson correlation among the occurrences frequencies of the motifs in different areas of the In-2 of the gene. However, the motifs distribution indicates a relative correlation among the occurrences of Gfi1-TFBs (A), WT1-TFBs (B) in the intron, so that they are restricted in specific regions of the intron, such as WT1-TFBs, which are huddled in a region of the intron from 100 up to 120 kbp.

similar longitudinal positions of the ln-2 of the gene (Fig. **2-A**). However, it seems that there are significant correlations among the occurrence frequencies of the WT1-In2-TFBs (Fig. **2-B**), and GAGA-In2-TFBs (Fig. **2-C**) in the ln-2 of the gene. So that, they are situated in several narrow regions of the In-2 of the gene.

3.5. The Occurrence of WT1-TFBs in the Related and Non-Related Genes of the *hBcl2*

The results of our analysis for the presence of the WT1-TFBs throughout the human *Bclx*, *Mcl1*, *Bclg*, *Bak*, *Bax*, *Bad*, *hmgb1*, *Fech*, *Greb1L*, *Ncp1*, *Smad4*, *Kc6*, *Brca1*, *Brca2* and *hFVIII* genes, as negative controls, led to the detection of various numbers WT1-TFBs in these genes without any correlation to the length of the genes (Fig. **3**). So, *Greb1L* gene with the highest length has only 116 WT1-TFBs and the *bad* gene which is 14877 bp long showed 96 of this factor in their context, whilst there were not any WT1-TFBs in the context of *Bclg and* human *Brca2* genes with 49816 and 63366 bp in the length, respectively.

3.6. CG Content and the Frequencies of WT1-TFBs

In order to determine whether there is a relationship between the occurrence frequencies of the WT1-TFBs and GC content of the studied genes, GC content of the genes were investigated. This assessment revealed a variance of 36 up to 57% in the GC content of the investigated genes (Fig. **4**). As shown in this figure, there are not any linear relationships among the occurrence frequencies of the WT1-TFBs to the CG content of these genes.

3.7. The Accurate Position of the WT1-TFBs on the In-2 of the *hBcl2* Gene

With a closer inspection of the locations of the WT1-TFBs on the In-2 of the gene, we realized that these motifs are scattered on both the strands of the In-2 of the gene, however 65% of them were situated on the positive strand (Table **3**). Meanwhile, it was found that these motifs were scattered on eleven areas (L1 up to L11) of both the strands of the In-2 with various frequencies. On the other hand, spatially correlation among the sequence context of these areas showed overlapping in four regions including A, B C and D. Table **3** shows L1, L5 and L9 areas that in total have 42TFBs in its place, situated from 104492 up to 104559bp of the intron, as A region. In this regard, L2 and L6 as B region, showed spatial correlation to each other in the area of the In-2 from 114366 up to 114420bp with 11 TFBs. Moreover, C region is the area of the intron from 118107 up to 118848 bp comprising L3, L7 and L10 locations with 21 TFBs. Meanwhile, L4, L8, L11 locations which in total have 48 TFBs in its place, situated from 119011up to 119923bp of the intron, as D region.

3.8. The Features of the Sequence Context of the WT1-TFBs

Closer attention to the occurrence frequencies of the WT1-TFBs to the length of the corresponding locations reveals cooperative structure in five locations including L1 (1), L2 (2), L5 (5), L6 (6) and L9 (9) areas (Fig. **5**). As shown in this figure, these areas with short length related to other locations comprise more number of the WT1-TFBs in its context, as overlapped structure to each other. On the other hand, a

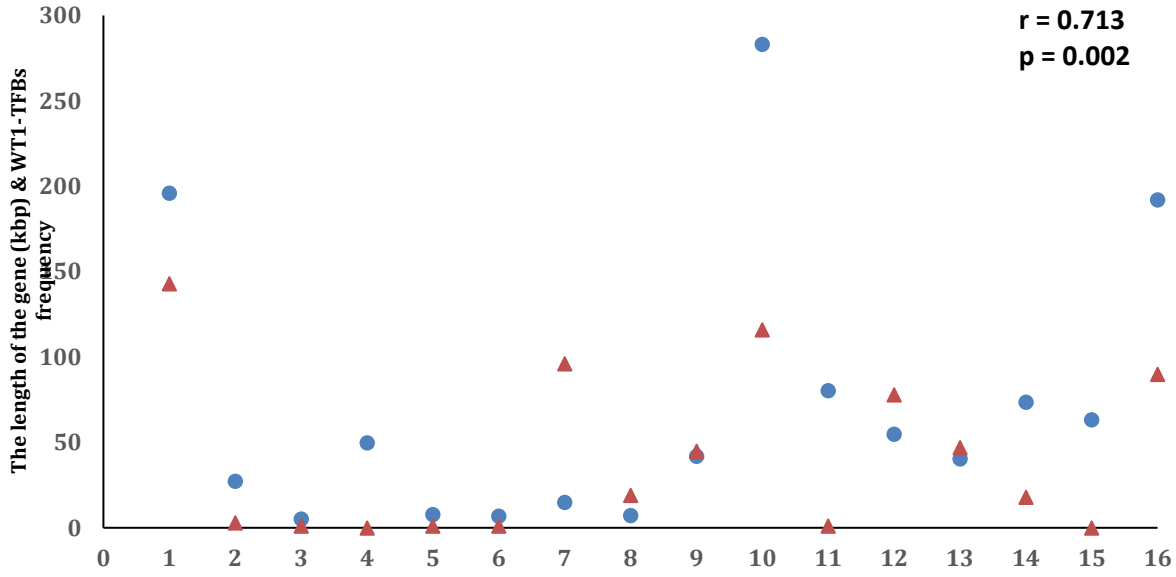

Fig. (3). Pearson correlation among the occurrence frequencies of WT1-TFBs to the length of human *Bcl2* (1), *Bclx* (2), *Mcl1* (3), *Bclg* (4), *Bak* (5), *Bax* (6), *Bad* (7), *hmgb1* (8), *Fech* (9), *Greb1L* (10), *Ncp1* (11), *Smad4* (12), *Kc6* (13), *Brca1*(14), *Brca2* (15) and *hFVIII* (16) genes. As shown in the figure, there are not any correlation among the frequencies of WT1-TFBs and the length of the genes. (● = length of the gene, ▲ = frequency of the WT1-TFBs).

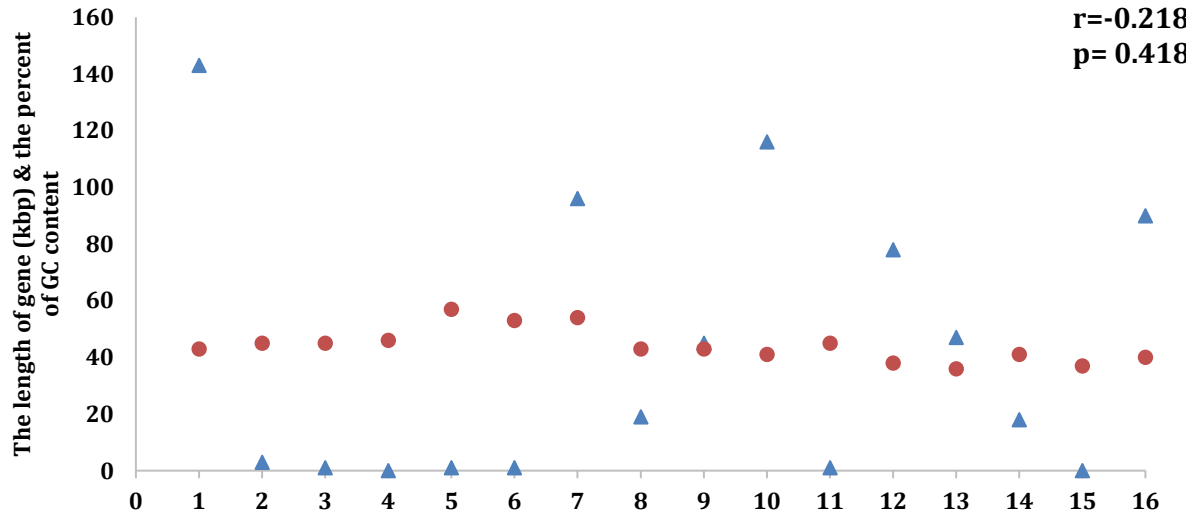

Fig. (4). Pearson correlation among the occurrence frequencies of the WT1-TFBs to the GC content of the human *Bcl2* (1), *Bclx* (2), *Mcl1* (3), *Bclg* (4), *Bak* (5), *Bax* (6), *Bad* (7), *hmgb1* (8), *Fech* (9), *Greb1L* (10), *Ncp1* (11), *Smad4* (12), *Kc6* (13), *Brca1* (14), *Brca2* (15) and *hFVIII* (16) genes. (● = CG content of the gene, ▲ = frequency of the WT1-TFBs).

deeper inspection in the sequence context of the WT1-TFBs showed that some of them had nucleotide mismatched in some position to the original sites. This investigation revealed that these sites, with *E*-value equal to zero, had various lengths ranging in size from 22 up to 45 bps. However, most of these sites were located in two groups with 36 and 45 nucleotides in length with G and T nucleotides in the sequence context (Fig. **6**).

3.9. Interaction Among WT1-TF to WT1-TFBs Regions of the In-2 of the *hBcl2* Gene

The results of the molecular docking among the WT1-TF to the WT1-TFBs regions of the In-2 of the gene led to significant outcomes, about the circumstance of the interaction as well as amount of the binding affinity compared to the negative controls (Figs. **7** and **8**). As shown in Figure 7, in the circumstance of interaction among WT1-TF to the WT1-TFBs regions, transcription factor positioned in the major grooves of the WT1-TFBs regions perfectly, with numerous strong hydrogen bonds (Fig. **7-A** and **C**). However, this situation during the interaction of the WT1-TF to the negative control was quite different, that performed a sketchy interaction between the WT1-TF and negative sequence, especially in the minor groove without strongly hydrogen bonds (Fig. **7-B**). On the other hand, evaluating the binding affinity of the WT1-TF to corresponding sites throughout the WT1-TFBs regions provided confirming data about the circum-

Table 3. The accurate position of the WT1-TFBs areas throughout the In-2 of the *hBcl2* gene. (A: Location of the area of the WT1-TFBs

Position		A.	R.	St.	Position		N.	R.	St.	Position		N.	R.	St.
Start	End				Start	End				Start	End			
104498	104563	L1	16	-	104492	104566	L5	16	+	104497	104559	L9	10	+
114368	114415	L2	6	-	114366	114420	L6	5	+	118107	118151	L10	1	+
118114	118848	L3	9	-	118112	118812	L7	11	+	119086	119923	L11	30	+
119011	119449	L4	16	-	119035	119123	L8	2	+					

R: Repetition of WT1-TFBs, St: strand)

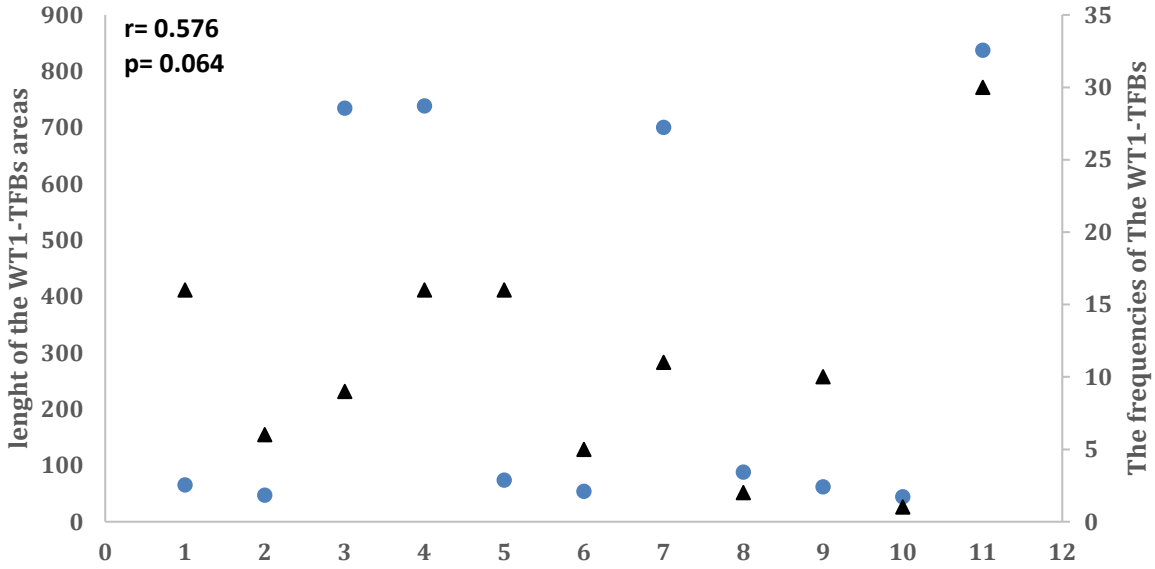

Fig. (5). Pearson correlation among the frequencies of the WT1-TFBs to the different areas of the In-2 of the *hBcl2*. As shown in the figure, there is not any correlation among the concentration of WT1-TFBs with corresponding length of the locations, which could be implying the cooperative structure of TFBs in some of the WT1-TFBs locations (● = length of the WT1 areas (L1 up to L11), ▲ = frequency of the WT1-TFBs).

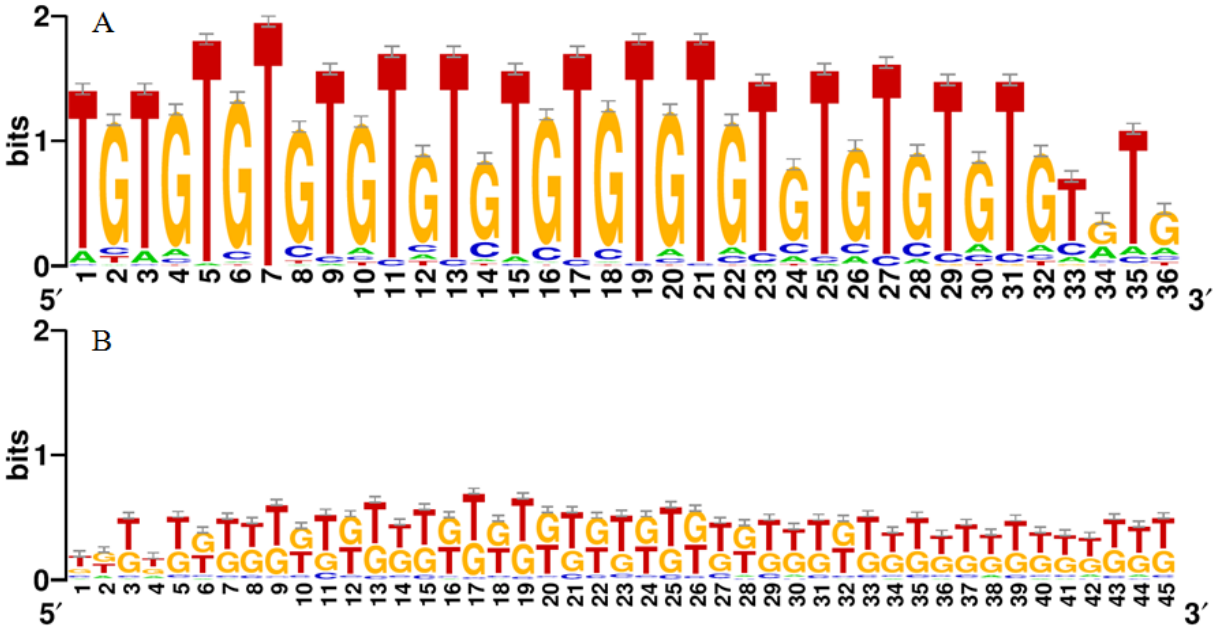

Fig. (6). Sequence logos of the WT1-TFBs in the In-2 of the *hBcl2* gene with 50 (A) and 91 (B) repetition. As showed in this figure, WT1-TFBs in the In-2 of the *hBcl2* gene are riches of G and T nucleotides in their context.

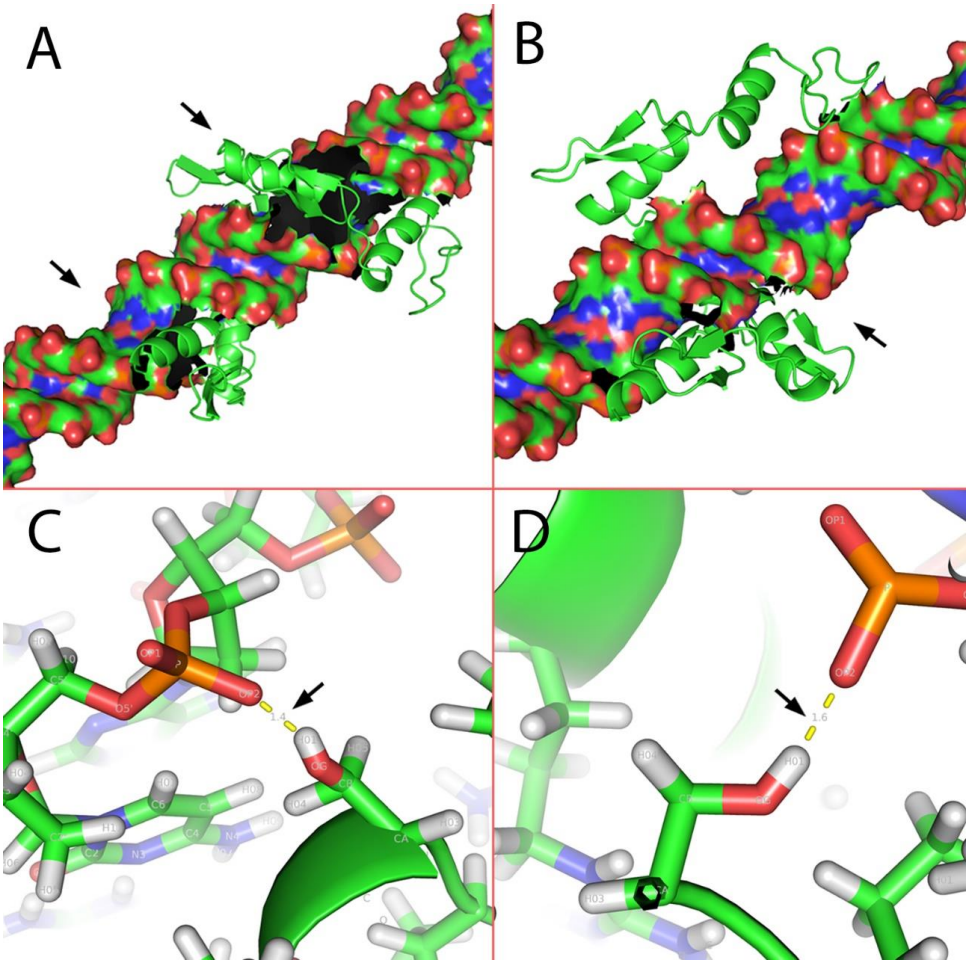

Fig. (7). Top structures of the interaction of the WT1-TF to the WT1-TFBs regions of the In-2 of the *hBcl2* gene (**A**), and sequences of negative control (**B**). As shown in the A and B figures with arrows, WT1-TF is positioned in the major groove of DNA in the WT1-TFBs regions, whilst this factor is located as partially in minor groove of the sequences of negative control. **C**) Enlarged view of a part of the interaction of the WT1-TF to the WT1-TFBs regions. As shown in this image with arrows, very strong hydrogen bonds are formed during interaction with 1.4 and 1.6 angstrom in the lengths among Ser444 with T29 and Ser476 with T26.

stance of interaction (Fig. **8**). As shown in this figure, the average binding energy of the interaction of the WT1-TFs to the WT1-TFBs regions were in lower level of the energy compared to negative control.

3.10. Sp1-TFBs Throughout the In-2 of the *hBcl2* Gene

Due to the cooperative effects of the WT1-TF with Sp1-TF, distribution pattern of this factor throughout the In-2 of the *hBcl2* gene was investigated. The result of this survey revealed that the Sp1-TF had 12 homologous corresponding TFBs throughout the In-2 of the gene with a close linear relationship to WT1-TFBs locations, in some cases (Table **4**).

3.11. Assessment the Occurrences of the Mutation Throughout the WT1-TFBs Regions

The result of our investigation further continued with assessment of 9285 SNPs occurring throughout the *hBcl2* gene, up to now. This analysis showed that the 96% of these mutations were situated in In-2 of the gene which could imply to their occurrence dependency of them to the length of this region. However, the occurrence frequencies of these SNPs throughout the WT1-TFBs regions seem to be independent of the length of these regions, compared to the other parts of the In-2. Therefore, more than 2% of these mutations occur in these regions with various densities (Fig. **9**). As illustrated in figure 9, the occurrence frequency of these mutations throughout the various regions of the WT1-TFBs seems to be independent of the length, so that the C region has the highest density of the mutation, whilst the D region is the longest area.

4. DISCUSSION

Eukaryotic gene expression relies on accurate interaction of Cis and Trans acting elements in a highly coordinated fashion [40-41]. Therefore, exact prediction of Cis-regulatory elements as well as regulatory regions encoded within sequence context of DNA such as TFBs throughout the genomes and in different regions of a gene, is a crucial step towards understanding the regulatory mechanism of gene expression. In this regard, over 30 years ago, DNase I hypersensitive sites have been used extensively to mark these regions and elements throughout the genomes [42]. Nowadays, advanced technologies such as ChIP-chip, ChIP-

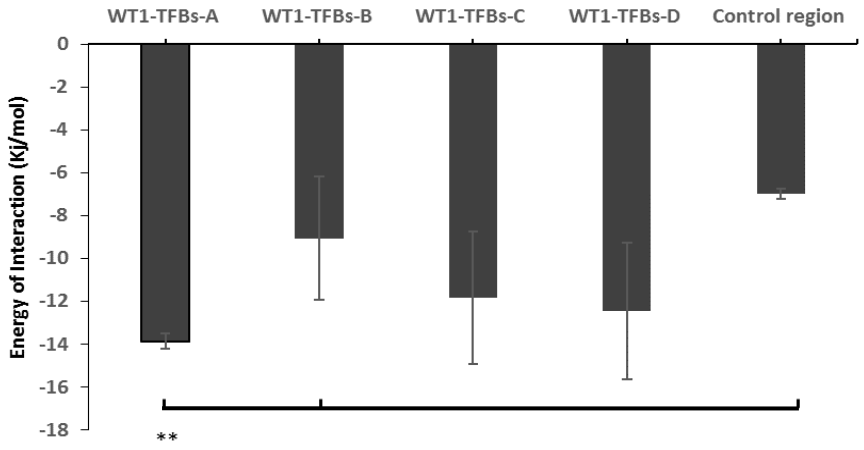

Fig. (8). The average energy of the interaction of WT1-TF to WT1-TFBs. The average energy of the interaction of WT1-TF to WT1-TFBs regions (A, B, C and D) of the In-2 of the hBcl2 gene compared to control region. **P ≤ 0.001. Error bars indicate standard deviation.

Table 4. The frequency and position of the homologous Sp1-TFBs throughout the In-2 of the *hBcl2* gene.

Site Name	Position		St.	Factor Name	Org/Species	Gene
	Start	End				
IR-site_C	57800	57823	+	Sp1	H sapiens	insulin receptor
IR-site_C	70912	70935	+	Sp1	H sapiens	insulin receptor
IR-site_C	118407	118430	+	Sp1	H sapiens	insulin receptor
Sp1-alpha-actin_(6)	134354	134372	+	Sp1	H sapiens	cardiac alpha-actin
Sp1-cytochrome C	83832	83852	+	Sp1	R norvegicus	cytochrome C
Sp1-prosaponin	109727	109745	+	Sp1	M musculus	prosaponin
Sp1-prosaponin	137558	137576	+	Sp1	M musculus	prosaponin
Sp1-prosaponin	176330	176348	+	Sp1	M musculus	prosaponin
Sp1-alpha-dystrobrevin	134734	134745	+	Sp1	M musculus	alpha-dystrobrevin
Sp1-COL2A1-2	50971	50947	+	Sp1	H sapiens	COL2A1
beta5-integrin-FP3	50222	50238	+	Sp1/Sp3	M musculus	beta5-integrin
Sp1-KDR/flk-1-IV	16337	16327	-	Sp1	H sapiens	KDR/flk-1

St: strand

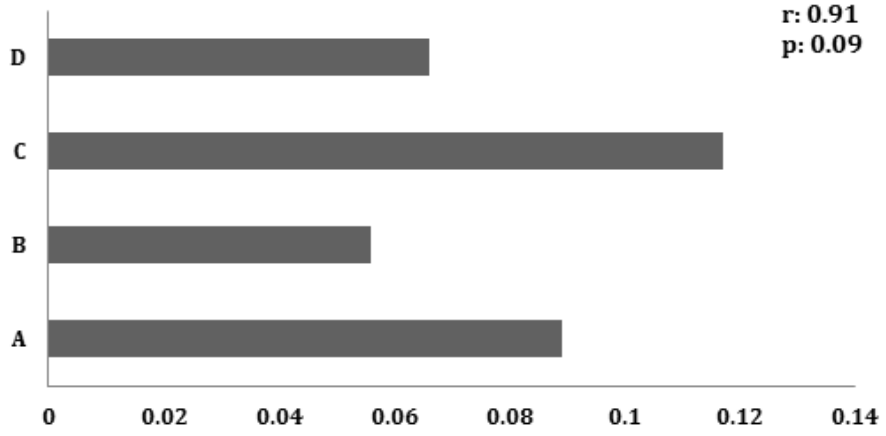

Fig. (9). The density of the SNP throughout the WT1-TFBs regions (A, B, C & D) of the In-2 of *hBcl2* gene. As showed in this figure there isn't any Pearson correlation among the concentration of WT1-TFBs to the corresponding regions, however the occurrence of them to the length of the regions seem to be significantly. (This data is obtained from dividing the number of mutations to the length of WT1-TFBs regions).

Seq analysis, next-generation sequencing have provided unprecedented opportunity for this purpose [41, 43-44]. Although this process is very interesting, but information about the repertoire of regulatory regions of individual genes cannot be achieved by these methods [45-46]. Subsequently, prediction using multiple genome information based statistical survey models is becoming a hot topic in bioinformatics to overcome these problems for revealing comprehensive catalogues of regulatory DNA elements and regions in individual genes [41, 47]. Bearing in mind, we surveyed complete regions of the *hBcl2* gene, in a systemic analysis, to discover TFBs encoded within sequence context of the intronic and exonic regions, based on *in-silico* investigation. Our analysis revealed 966 motifs of 118 different types of TFBs throughout the both strand of the *hBcl2* gene, in particular in narrow regions of the In-2 sequence context (Fig. **1**). Besides, these analyses detected target sites for 60 different types of TFs within the In-2 of the *hBcl2* gene, with most of them belonging to WT1-TF (Table **2**). Broadly speaking, this data is in agreement with previous thoughts, concerning the presence of the regulatory elements in intronic regions of a gene and regulatory functions of the introns [31, 48-50]. Although these features of TFBs could indicate the potential transcriptional regulatory function of the In-2 of the *hBcl2* gene, however additional evidences are needed to support this conjecture. In this regard, in the next step, the frequencies and distribution pattern of In2-TFBs were investigated in detail. Our results show that most of the In2-TFBs scattered sporadically and occurred only once, however 32 of them showed more than one repetition from 2 up to 141 on the both strands of the In-2 (Table **1**). Thorough analysis of the frequencies and distribution patterns of the selected motifs revealed that WT1-TFBs, Gfi-1-TFBs and GAGA-TFBs, with the highest repetition, have significant and contrasting distribution, in particular in WT1-TFBs features (Fig. **2**), which is consistent with cooperative functions of TFs in binding to regulatory DNA [41, 51]. WT1-TFBs as a target for WT1-TFs have been in the center of attention recently, and they were found to be exposed in promoter regions of a broad range of genes such as *c-myc*, *Egr*, *Wt*, *Pax2*, *trkB*, *Sf1*, *Bcl2*, which could be a conductor either a transcriptional repressor or activator, in a cell and promoter specific manner [52-53]. In this regard demonstrated that, WT1-TFs is a master TF involved in embryonic development and development of several tissues in normal situation and its deregulation is also reported in tumorigenesis and abnormal situations [54-56]. However, established involvement of WT1-TFs in human cancer is very complex, acting as a tumor suppressor in some contexts and as an oncogene in other context [49, 52]. On the other hand, various isoforms of WT1-TFs may explain the different and apparently opposing roles in proliferation and apoptosis [55]. Moreover, it has become evident that a group of WT1-TF involves in RNA metabolism, translational regulation and association with translating polysomes [52, 57], which might be correlated to WT1-TFBs of the In-2 sequence context as a RNA binding target. Taken together, our findings not only are consistent with other reports, but also confirm the regulatory function of the intronic regions [27, 30, 58], and nuclear/cytoplasmic roles of intronic sequence context (ISC) which reported previously [40, 59-60]. In this regard, it has become increasingly clear that ISC can influence various stages of mRNA metabolism such as transcription, nuclear mRNA export and mRNA decay [40, 61-63]. On the other hand, it was demonstrated that WT1-TF might be a specific adaptor protein that links a specific subset of mRNAs for transporting to the target location [64], suggesting probably dual roles of WT1-TFBs in *hBcl2* transcription and post-transcription processes. Moreover, cooperative feature of WT1-FBs was investigated in this study via tracking the frequency (Table **1**), distribution (Fig. **2**), location (Table **3**) as well as their correlation to other TFs and TFBs (Table **2** and **4**). As demonstrated in these figures and tables, WT1-TFBs with the highest frequency, were situated in four neighboring regions of the In-2 with overlapped structure in 11 narrow areas and also with a close linear relationship to Sp1-TFBs. In this regard, there are a wealth of data that the function of WT1-TF may be modulated by physical contact with other regulators such as SP1, Egr1 or with itself [49, 53]. Furthermore, numerous TFs have been determined with overlap feature to the targets of WT1 such as NANOG, GLI1, E2F1, POU5F1/OCT4, SPI-1, YY-1, GATA1, and C/EBP-β [49]. On the other hand, WT1-TFBs of the In-2 of the *hBcl2* gene provide a suitable location for WT1-TF interaction (Figs. **7** and **8**), however sequence logos of these motifs showed a character of the WT1-TFBs contrary to other reports in the length and context (Fig. **5**). In this regard, several binding sites with 4 up to 10-mers have been determined for WT1-TF with GC-rich such as GCGGGGGCG, GNGNGGGNG, GNGNGGGNGNS, and GCGTGGGAGT [49, 53, 65]. However, it is important to note that these sites are determined in promoteric regions of the genes [49], which cannot be inconsistent with our results in the length and context of the WT1-TFBs of the In-2. Moreover, it was demonstrated that each isoform of the WT1-TFs may bind to slightly different sequences in DNA, influenced via dimerization and post-translational modifications [49]. Furthermore, reported that DNA binding sites typically have 4 to 30 base pairs long, but could be reached up to 200 bp in length [65], which is consistent to the length of WT1-In2-TFBs. Meanwhile, the values of the WT1-In2-TFBs were measured by tracking the occurrence of SNP in WT1-TFBs regions (Fig. **9**). These analyses showed that a bulk of SNP occurred in these regions up to now. These data could be another reason for the regulatory function of the intronic regions, especially of the WT1-TFBs regions of the In-2 of the *hBcl2* in line with other reports. Regarding the contribution of ISC in regulation of gene expression, it has been determined that various genetic diseases result from ISC mutation [66-68]. On the other hand, numerous studies have demonstrated that WT1-TF affects the cell cycle and regulates apoptosis by targeting *Bcl-2* family members, either directly or indirectly [69-71], which were approved again by our data.

CONCLUSION

These data indicate a series of new targets of WT1-TFs in intronic regions of the *hBcl2* gene, which could endorse regulatory functions of introns on one hand and the role of detected regions on *hBcl2* gene regulation on the other hand. Moreover, these data indicate that WT1-TFBs regions could be a new candidate for attenuating the biological behavior of apoptosis, and a biomarker for cancer diagnosis. However,

based on these preliminary data, comprehensive studies are needed to analyze WT1-TFBs and their role in the *hBcl2* regulation for a variety of cancers. Taken together, it can be concluded that WT1-TFBs regions may provide a novel diagnostic approach in the majority of hematologic malignancies expressing *Bcl2*.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

This work was financially supported by grant No. 2/29804 from the Ferdowsi University of Mashhad and performed at the Institute of Biotechnology, Ferdowsi University of Mashhad. The authors are grateful to Muhammad Irfan-maqsood for valuable suggestions with respect to the manuscript.

REFERENCES

[1] Heiser D, Labi V, Erlacher M, Villunger A. The Bcl-2 protein family and its role in the development of neoplastic disease. Exp Gerontol 2004; 39: 1125-35.

[2] Jhala NC, Vickers SM, Argani P, McDonald JM. Regulators of apoptosis in cholangiocarcinoma. Arch Pathol Lab Med 2005; 129: 481-6.

[3] Reed JC, Pellecchia M. Apoptosis-based therapies for hematologic malignancies. Blood 2005; 106: 408-18.

[4] Tang Y, Zhang DY, Wu XM. Progress in small-molecule inhibitors of Bcl-2 family proteins. Yao Xue Xue Bao 2008; 43: 669-77.

[5] Tsuda H, Ning Z, Yamaguchi Y, Suzuki N. Programmed cell death and its possible relationship with periodontal disease. J Oral Sci 2012; 54: 137-49.

[6] del Mar Martinez-Senac M, Corbalan-Garcia S, Gomez-Fernandez JC. Study of the secondary structure of the C-terminal domain of the antiapoptotic protein bcl-2 and its interaction with model membranes. Biochemistry 2000; 39: 7744-52.

[7] Belzacq AS, Vieira HL, Verrier F, *et al*. Bcl-2 and Bax modulate adenine nucleotide translocase activity. Cancer Res 2003; 63: 541-6.

[8] Meyer KB. B-Myb as a critical regulator of Bcl-2 in human B cells. Immunology 2005; 114: 23-4.

[9] Lieber J, Kirchner B, Eicher C, *et al*. Inhibition of Bcl-2 and Bcl-X enhances chemotherapy sensitivity in hepatoblastoma cells. Pediatr Blood Cancer 2010; 55: 1089-95.

[10] Oltersdorf T, Elmore SW, Shoemaker AR, *et al.* An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature 2005; 435: 677-81.

[11] Zivny J, Klener P Jr, Pytlik R, Andera L. The role of apoptosis in cancer development and treatment: focusing on the development and treatment of hematologic malignancies. Curr Pharm Des 2010; 16: 11-33.

[12] Jeon BS, Yoon BI. Altered expression of cellular Bcl-2 in the progression of hamster cholangiocarcinogenesis. ScientificWorld Journal 2012; 2012: 385840.

[13] Strasser A, Harris AW, Cory S. E mu-bcl-2 transgene facilitates spontaneous transformation of early pre-B and immunoglobulin-secreting cells but not T cells. Oncogene 1993; 8: 1-9.

[14] Papadimitriou CS, Costopoulos JS, Christoforidou BP, *et al.* Expression of Bcl-2 protein in human primary breast carcinomas and its correlation with multifocality, histopathological types and prognosis. Eur J Cancer 1997; 33: 1275-80.

[15] Bakhshi A, Wright JJ, Graninger W, *et al.* Mechanism of the t(14;18) chromosomal translocation: structural analysis of both derivative 14 and 18 reciprocal partners. Proc Natl Acad Sci U S A 1987; 84: 2396-400.

[16] Glantz LA, Gilmore JH, Lieberman JA, Jarskog LF. Apoptotic mechanisms and the synaptic pathology of schizophrenia. Schizophr Res 2006; 81: 47-63.

[17] Otake Y, Soundararajan S, Sengupta TK, *et al.* Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA. Blood 2007; 109: 3069-75.

[18] He H, Lam M, McCormick TS, Distelhorst CW. Maintenance of calcium homeostasis in the endoplasmic reticulum by Bcl-2. J Cell Biol 1997; 138: 1219-28.

[19] Sheikh MS, Huang Y. TRAIL death receptors, Bcl-2 protein family, and endoplasmic reticulum calcium pool. Vitam Horm 2004; 67:169-88.

[20] Deng X, Gao F, May WS. Protein phosphatase 2A inactivates Bcl2's antiapoptotic function by dephosphorylation and up-regulation of Bcl2-p53 binding. Blood 2009; 113: 422-8.

[21] Ferdek PE, Gerasimenko JV, Peng S, Tepikin AV, Petersen OH, Gerasimenko OV. A novel role for Bcl-2 in regulation of cellular calcium extrusion. Curr Biol 2012; 22: 1241-6.

[22] Ruvolo PP, Deng X, May WS. Phosphorylation of Bcl2 and regulation of apoptosis, Leukemia 2001; 15: 515-22.

[23] Packham G, White EL, Eischen CM, *et al.* Selective regulation of Bcl-XL by a Jak kinase-dependent pathway is bypassed in murine hematopoietic malignancies. Genes Dev 1998; 12: 2475-87.

[24] de Souza SJ, Long M, Gilbert W. Introns and gene evolution. Genes Cells 1996; 1: 493-505.

[25] Rohrer J, Conley ME. Transcriptional regulatory elements within the first intron of Bruton's tyrosine kinase. Blood 1998; 91: 214-21.

[26] Furger A, O'Sullivan JM, Binnie A, Lee BA, Proudfoot NJ. Promoter proximal splice sites enhance transcription. Genes Dev 2002; 16: 2792-9.

[27] Le Hir H, Nott A, Moore MJ. How introns influence and enhance eukaryotic gene expression. Trends Biochem Sci 2003; 28: 215-20.

[28] Kim DS, Kim TH, Huh JW, *et al.* LINE FUSION GENES: a database of LINE expression in human genes. BMC Genomics 2006; 7: 139.

[29] Polak P, Domany E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. BMC Genomics 2006; 7: 133.

[30] Eddy J, Maizels N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. Nucleic Acids Res 2008; 36: 1321-33.

[31] Haddad-Mashadrizeh A, Zomorodipour A, Izadpanah M, *et al.* A systematic study of the function of the human beta-globin introns on the expression of the human coagulation factor IX in cultured Chinese hamster ovary cells. J Gene Med 2009; 11: 941-50.

[32] Lang G, Gombert WM, Gould HJ. A transcriptional regulatory element in the coding sequence of the human Bcl-2 gene. Immunology 2005; 114: 25-36.

[33] Solovyev VV, Shahmuradov IA, Salamov AA. Identification of promoter regions and regulatory sites. Methods Mol Biol 2010; 674: 57-83.

[34] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000; 16: 276-7.

[35] Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator, Genome Res 2004; 14: 1188-90.

[36] Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM. NPDock: a web server for protein and nucleic acid docking. Nucleic Acids Res 2015; 43: 425-30.

[37] Vakser IA, Aflalo C. Hydrophobic docking: A proposed enhancement to molecular recognition techniques. Proteins 1994; 20: 320-9.

[38] Tuszynska I, Bujnicki JM. DARS-RNP and QUASI-RNP: New statistical potentials for protein-RNA docking. BMC Bioinformatics 2011; 12; 1-16.

[39] Vlahovicek K, Kajan L, Pongor S. DNA analysis servers: plot.it, bend.it, model.it and IS, Nucleic Acids Res 2003; 31: 3686-7.

[40] Skoko N, Baralle M, Tisminetzky S, Buratti E. InTRONs in biotech. Mol Biotechnol 2011; 48: 290-7.

[41] Chen H, Li H, Liu F, *et al.* An integrative analysis of TFBS-clustered regions reveals new transcriptional regulation models on the accessible chromatin landscape. Sci Rep 2015; 5: 8465.

[42] Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. Annu Rev Biochem 1988; 57: 159-97.

[43] Wang J, Bowen NJ, Marino-Ramirez L, Jordan IK. A c-Myc regulatory subnetwork from human transposable element sequences. Mol Biosyst 2009; 5: 1831-9.
[44] Conley AB, Jordan IK. Identification of transcription factor binding sites derived from transposable element sequences using ChIP-seq. Methods Mol Biol 2010; 674: 225-40.
[45] Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 2012; 13: 613-26.
[46] Whitfield TW, Wang J, Collins PJ, *et al.* Functional analysis of transcription factor binding sites in human promoters. Genome Biol 2012; 13: R50.
[47] Hou L, Qian MP, Zhu YP, Deng MH. Advances on bioinformatic research in transcription factor binding sites. Yi Chuan 2009; 31: 365-73.
[48] Nishihara H, Smit AF, Okada N. Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res 2006; 16: 864-74.
[49] Holloway DT, Kon M, DeLisi C. In silico regulatory analysis for exploring human disease progression. Biol Direct 2008; 3: 24.
[50] Li H, Chen D, Zhang J. Analysis of intron sequence features associated with transcriptional regulation in human genes. PLoS One 2012; 7: e46784.
[51] Pape UJ, Klein H, Vingron M. Statistical detection of cooperative transcription factors with similarity adjustment. Bioinformatics 2009; 25: 2103-9.
[52] Ambu R, Vinci L, Gerosa C, *et al.* WT1 expression in the human fetus during development. Eur J Histochem 2015; 59: 2499.
[53] Bagchi D, Andrade J, Shupnik MA. A new role for wilms tumor protein 1: differential activities of + KTS and -KTS variants to regulate LHbeta transcription. PLoS One 2015; 10: e0116825.
[54] Xu J, Wang HW, Li XH, *et al.* WT1 gene expression and its isoform ratio in different cell subsets of normal human bone marrow. Zhongguo Shi Yan Xue Ye Xue Za Zhi 2007; 15: 603-6.
[55] Parenti R, Cardile V, Graziano AC, *et al.* Wilms' tumor gene 1 (WT1) silencing inhibits proliferation of malignant peripheral nerve sheath tumor sNF96.2 cell line. PLoS One 2014; 9: e114333.
[56] Luna I, Such E, Cervera J, *et al.* WT1 isoform expression pattern in acute myeloid leukemia. Leuk Res 2013; 37: 1744-9.
[57] Bor YC, Swartz J, Morrison A, Rekosh D, Ladomery M, Hammarskjold ML. The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron. Genes Dev 2006; 20: 1597-608.
[58] Parris GE. Developmental diseases and the hypothetical Master Development Program. Med Hypotheses 2010; 74: 564-73.
[59] Reed R, Hurt EA. conserved mRNA export machinery coupled to pre-mRNA splicing, Cell 2002; 108: 523-31.
[60] Ying SY, Lin SL. Intron-derived microRNAs-fine tuning of gene functions. Gene 2004; 342: 25-8.
[61] Fong YW, Zhou Q. Stimulatory effect of splicing factors on transcriptional elongation. Nature 2001; 414: 929-33.
[62] Le Hir H, Seraphin B. EJCs at the heart of translational control. Cell 2008; 133: 213-6.
[63] Rose AB. Intron-mediated regulation of gene expression. Curr Top Microbiol Immunol 2008; 326: 277-90.
[64] Dudnakova T, Spraggon L, Slight J, Hastie N. Actin: a novel interaction partner of WT1 influencing its cell dynamic properties. Oncogene 2010; 29: 1085-92.
[65] Borneman AR, Gianoulis TA, Zhang ZD, *et al.* Divergence of transcription factor binding sites across related yeast species. Science, 2007; 317: 815-9.
[66] Schell T, Kulozik AE, Hentze MW. Integration of splicing, transport and translation to achieve mRNA quality control by the nonsense-mediated decay pathway. Genome Biol 2002; 3: REVIEWS1006.
[67] Lin SL, Miller JD, Ying SY. Intronic microRNA (miRNA). J Biomed Biotechnol 2006; 2006: 26818.
[68] Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 2007; 76: 51-74.
[69] Loeb DM, Korz D, Katsnelson M, Burwell EA, Friedman AD, Sukumar S. Cyclin E is a target of WT1 transcriptional repression. J Biol Chem 2002; 277: 19627-32.
[70] Loeb DM. WT1 influences apoptosis through transcriptional regulation of Bcl-2 family members. Cell Cycle 2006; 5: 1249-53.
[71] Xu C, Wu C, Xia Y, *et al.* WT1 promotes cell proliferation in non-small cell lung cancer cell lines through up-regulating cyclin D1 and p-pRb in vitro and in vivo. PLoS One 2013; 8: e68837.