African Journal of Biotechnology Vol. 9(25), pp. 3699-3713, 21 June, 2010
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2010 Academic Journals

Full Length Research Paper

Isolation and characterization of developmentally regulated novel target site from embryonic chick heart

S. Neha[1], Shyamal K. Goswami[2] and R. Vibha[1]*

[1]Department of Biotechnology, Jaypee Institute of Information Technology University, Noida 201307, India.
[2]School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India.

Accepted 5 April, 2010

Differential gene expression is the primary determinant of numerous biological processes such as cell differentiation, proliferation, organogenesis, tumor progression and apoptosis. In that context, regulatory proteins play pivotal roles and determine cell fate in all physiological conditions of differentiation, development and disease. As these regulatory proteins are present in extremely small amount in cells, their isolation and identification from tissues is impracticable. These regulatory proteins bind to the target sites (DNA elements) located in the promoter, enhancer and other regulatory region of the genome. These target sites control the developmental expression of genes. In the present paper we have isolated and characterized one novel target site (GTGTT) which is developmentally expressed during chick heart development.

Key Words: Novel regulatory target sites, gene expression, promoter, transcription, gene specific transcription factors, bioinformatics, heart, development.

INTRODUCTION

In order for the cell to utilize the information contained within its genome, individual gene, each of which represents a single unit of information, has to be expressed in a coordinated manner (Wyeth and Albin, 2004). Co-ordination of gene expression determines the cellular transcriptome, which in turn specifies the nature of the proteomes and defines the cellular function. Control of gene transcription is one of the major regulatory mechanisms operative in a particular cell type. Transcription regulation in eukaryotes is a coordinated process and requires the concerted functions of multiple proteins or transcription factors (Gross and Oelgeschläger, 2006).

In metazoan organisms, formation of a preinitiation complex at the right time and at the right promoter is a prerequisite to executing the correct programs of mRNA synthesis. This involves the interplay of many gene 9.5 regulatory DNA elements located in the promoter enhancer regions of various tissue and developmental stage restricted genes (Bryan and Tjian, 2000).

The expression pattern of numerous genes changes dynamically depending on the developmental stage and the differentiation state of the cell. Transcription factors regulate cellular events at the gene expression level by communicating signals to the general transcription machinery that forms a preinitiation complex (PIC) at class II core promoters. Recent data strongly suggests that preinitiation complex formation at distinct promoters reflect the spatiotemporal profile of gene expression in multicellular organisms and that core promoter recog-nition by distinct factors is an additional level of transcriptional regulation (Hochheimer and Tjian, 2003). Until recently, the general transcription machinery that makes up the PIC was thought to be largely invariant. However, during recent years, several homologues of basal transcription factors including additional member of TATA box protein (TBP) family such as TBP-related factors

*Corresponding author. E-mail: vibha.rani@jiit.ac.in. or vibhasandeep@rediff.com.

Abbreviations: PIC, Preinitiation complex; **TBP,** TATA box protein; **GSTFs,** gene specific transcription factors; **NBPs**, nucleic acid binding proteins; **PCR,** polymerase chain reaction; **EDTA,** ethylenediaminetetraacetic; **SDS,** sodium dodecyl sulfate; **EPD,** eukaryotic promoter database.

(TRFs-TRF1, TRF2 and TRF3) have been identified in metazoan organisms. In addition, a large number of tissue-specific transcription factors have been described that control expression of cell-type specific genes and most of these are enhancer-binding transcription factors (Boyle et al., 2006).

One of the key questions in gene regulation is how specific sets of genes are selected for expression during cell growth, differentiation, apoptosis or in response to environmental cues. Gene expression program during embryonic development is highly orchestrated and precise (Yang et al., 2007). Therefore, how the array of distinct developmental programs is established by a genome of fixed size remains a question yet to be answered. Nonetheless, formation of a pre initiation complex at the right time and at the right promoter is presumably the mechanism by which such a diverse developmental program is executed. Furthermore, initiation of specific set of genes and their transcription at the right level is directed by an array of cis-regulatory DNA element and their cognate gene specific transcription factors (Corry and Underhill, 2005). Although eukaryotic gene expression is initiated by the transcription apparatus assembled at the core promoter region (TATA sequence/ downstream promoter elements), the extent of transcription is controlled by the gene specific transcription factors operating at the upstream regulatory regions of the promoter. The gene specific transcription factors (GSTFs) constitute the largest and the most divergent class of DNA-binding proteins (Grondin et al., 2007). Gene specific transcription factors mediate their effects by diverse mechanisms such as tissue-restricted activity, existence of different isoforms, and formation of functional complex by hetero and homo dimerization. They are characterized by their specific binding to distinct DNA elements found in the regulatory regions of different genes.

DNA sequencing data derived from the human genome suggests that the total number of genes is ~ 30,000 to 40,000, only fractions of which have yet been annotated. As more and more genomes are sequenced, organisms are increasingly represented by a list of genes although there is little knowledge as to how these expressions are coordinated (Venter et al., 2001). Thus the major challenge of the post-sequencing era is to understand gene regulation in a global scale and to assign the function to tens of thousands of genes. Integrative approaches such as computational analysis of the gene regulatory regions, their expression profiling and functional characterization are required for gene discovery and pathway elucidation (Saurabh and Martin, 2002). In this context, subsets of genes which encode nucleic acid binding proteins (NBPs) are most important to analyze for their function. It is estimated that 1.5% of the human genome encodes proteins. The remaining of the genome contains various regulatory signals that control their functions including expression. These regulatory signals are typically protein-target sites (sequence motifs), which are too short and degenerate having variability in nucleotide preference at certain positions. Nonetheless, there is much evidence that these regulatory elements occur in clusters especially for regulating transcription (such as enhancers/silencers containing multiple cis elements spread over a few hundred base pairs). It is believed that such clusters of signals are individually weak but collectively strong (Birnbaum et al., 2001). Organisms have devoted a significant fraction of their DNA to encrypt cis-regulatory programs that control and coordinate gene expression at the level of transcription. The output(s) of the cis-regulatory programs depend on the cellular context and extra-cellular inputs. Typically, an external stimulus activates a signal transduction pathway which leads to the modification of the activities of several transcription factors which in turn target enhancers/regulatory regions of certain genes, effecting their expression (Ian, 2007).

DNA element targeted by sequence specific binding proteins are small, of variable length (4 bp and above) and often show limited degeneracy at various nucleotide positions, therefore making it difficult to be identified in the background of the entire genome, especially in metazoan cell. It is thus expected that efficient identification of novel regulatory target sites and their cognate factors will definitely contribute to the comprehensive understanding of gene regulation.

Embryonic development and organogenesis is a highly orchestrated process involving precise and integrated control of gene expression (Barker, 2007). It involves an enormously complex gene regulatory network. Lineage specification and differentiation of various cell types underlie the basic mechanism of organogenesis (Ahuja et al., 2007). Since a functional circulatory system is essential for the progress of embryogenesis, heart formation is one of the first signs of organogenesis within the developing embryo and cardiac myocytes are amongst the earliest differentiated cell types. This process is conserved from flies to man (Nemer and Nemer, 2001). During the recent years, heart development has come up as a paradigm of cell differentiation and organogenesis (Keller et al., 2007).

In the present study, we have identified one novel developmentally regulated target sites isolated from embryonic chick heart by using methodology reported by Sindhu et al. (2004). We have previously described an efficient method of isolation of a library of DNA elements targeted by sequence specific proteins.

MATERIALS AND METHODS

Materials

Oligonucleotides(5'CGAGGTCGACGGTATCGNNNNNNNNNNGATCCACTAGTTCTAGAGC3',and ATAGTGTGTT) used in study were

synthesized from Genosys, USA. Polymerase chain reaction (PCR) was done using KlenTaq DNA polymerase. All chemicals were purchased from Sigma Aldrich unless mentioned otherwise.

Screening of "developmentally regulated target sites

Synthetic oligonucleotides used for the construction of a pool of random decanucleotide sequences was as described by Nallur et al. (1996). Single stranded oligonucleotide was amplified by PCR using primers encompassing the terminal sequences (M1 and M2, primer synthesized from pBluescript vector). The PCR product was then digested with HincII and SpeI and cloned into pBluescipt plasmid vector (+/-) digested with the same enzymes. The frequency of recombination was checked by PCR amplification of randomly picked up bacterial colonies using the same primer. More than eighty percent of transformants contained the recombinant plasmids and were pooled in batches. The cognate binding sites were then selected from the pre pool by four consecutive steps of gel-mobility shift followed by PCR. Plasmid DNA isolated from each pool of recombinants was amplified and 1 ng of amplified DNA was used for binding with 40 µg nuclear protein prepared from 12 day old chick embryonic heart. The reaction was then resolved on 8% acrylamide gel and the entire lane starting from 1 cm above the unbound oligonucleotide (identified by ethidium bromide staining) to 0.5 cm below the well was excised, crushed into small pieces and eluted overnight in a buffer containing 500 mM ammonium acetate, 10 mM magnesium chloride ($MgCl_2$), 1 mM ethylenediaminete-traacetic (EDTA), 0.1% sodium dodecyl sulfate (SDS). The eluted DNA was then precipitated using 100 ng of poly dI-dC as a carrier and amplified by a second set of internal primer pairs obtained from the pBluescript vector. The selection process was then repeated followed by the amplification with a further internal primer pair. Use of sequentially internal primer pairs during selection-amplification cycles helped in obtaining cleaner PCR products after each rounds of selection and also increases the kinetics of interaction between the target sites and the proteins. After four rounds of selection, the eluted material was finally amplified by the innermost primerpairs and cloned into the pBluescript vector at the Hinc II and SpeI sites as before (for the construction of the pre pool) (Sindhu et al., 2004). The recombinants thus obtained were then pooled and saved as the "Catalogue of developmentally regulated target sites".

DNA sequencing

The nucleotide sequencing of selected target sites were manually done using sequencing kit from Epicenter Biotechnologies, USA. The samples were run on 6% acrylamide gel containing 6 M urea.

Preparation of nuclear extracts from chick embryos

12-days-old chick embryos were taken for nuclear extract preparation. All steps were conducted at 4°C. Brain, heart, liver and skeletal tissues were taken out from chick embryo and added to the prechilled hand homogenizer. One ml of buffer A (20 mM Hepes (Sigma), 20% glycerol (Qualigen), 10 mM NaCl (Sigma), 1.5 mM $MgCl_2$ (Sigma), 0.2 mM EDTA (Sigma) 0.1% triton-X100 (Sigma), 100 mM phenylmethylsulphonyl fluoride (PMSF) (Sigma), 1 mM dithiothreitol (DTT) (Sigma), 10 mg/ml Leupeptin (Sigma), 10 mg/ml Aprotenin (Sigma)) was added in each homogenizer having individual tissue. The tissues were homogenized in ice. The homogenates were centrifuged at 2500 rpm at 4°C for 15 min. Supernatants were discarded, pellets were then resuspended in 300 - 500 µl of buffer B (20 mM Hepes , 20% glycerol, 500 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1% triton-X, 100 mM PMSF, 1 mM DTT, 10 mg/ml Leupeptin, 10 mg/ml Aprotenin) and incubated in ice for 60 min with intermittent tapping (at every 10 min interval). The homogenates were then centrifuged at 13,000 rpm at 4°C for 15 min. The supernatants were aliquoted and snap freezed at -70°C. Bradford (1976) method was used to estimate the protein concentration in the samples.

Gel shift assay

Protein-DNA Binding reactions were carried out in 40 µl volume. Binding reactions contain (20 mM HEPES pH 7.9, 5% glycerol, 60 mM NaCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT), 4 µl (20 µg) nuclear extract, 1 µg poly dIdC and ^{32}P- labeled probe (40,000 cpm). All the components were mixed thoroughly and incubated in ice for 30 - 45 min. For competition, 100 fold molar excess of cold (unlabeled) DNA (self or nonself) was added to the binding reaction prior to the addition of the labeled probe. Cold competitor was first added, incubated in ice for 10 min and then radiolabeled probe was added in the reaction tube. DNA-protein complexes were separated from the free probe by electrophoresis in 6 - 8% native acrylamide gel.

DNAse I foot printing assay

Bovine pancreatic deoxyribonuclease I (DNAse I; Sigma) was stored as a 30 unit/µl solution in 5 mM sodium acetate (pH 4.0), 1 mM CaCl2 and 50% glycerol. DNA was amplified with pBluescript specific T3 and T7 primers. DNA was digested with PauI enzyme, electroeluted and then klenow labeled. The stock solutions of DNAse I were kept at -20˚C and freshly diluted to the desired concentration immediately prior to use. Foot printing experiments were performed in 40 µl reaction volume. 40 – 50 µg of 12-day-old chick nuclear heart and liver extracts were incubated with the labeled DNA fragments (Klenow labeled) for 60 min. Poly dIdC was added to avoid nonspecific binding. After 60 min of incubation at 4°C, digestion was initiated by the addition of DNAse I solution whose concentration was adjusted to yield a final enzyme concentration of about 0.01 unit/ml in the reaction mixture. For DNAse I activity, cofactor solution containing 10 mM $MgCl_2$, 5 mM $CaCl_2$ was added to the reaction mix. DNAse I treatment was given at 37°C. After the required period of time (varies from 30 - 70 s), the reaction was stopped by adding stop buffer (1% SDS, 200 mM NaCl, 20 mM EDTA). Samples were then phenol-chlorophorm extracted and precipitated by 0.1 vol of 1 M NaCl and 2.5 vol of absolute alcohol. Samples were resuspended in loading dye containing 95% formamide solution, 0.1 mM EDTA and tracking dyes (bromophenol blue and xylene cyanol.), heat denatured at 95˚C for 5 min and electrophoresd in 6% acrylamide gel containing 5 M urea.

South western blotting

South western blotting was used to detect the DNA-protein interaction. 12-days-old chick embryonic heart, liver and brain nuclear extracts were prepared and quantitated according to the Bradford (1976). Bovine serum albumin (BSA) was used as the standard. Polyacrylamide gel electrophoresis under denaturing condition (in the presence of 0.1% SDS) was performed according to the method of Laemmili, (1970). 40 µg of nuclear extracts from 12-day-old chick embryonic brain, liver and heart were loaded with equal amount of loading dye (5% SDS, 5 mM Tris.Cl pH 6.8, 2M DTT, 20% glycerol) onto the gel. The proteins were stacked at

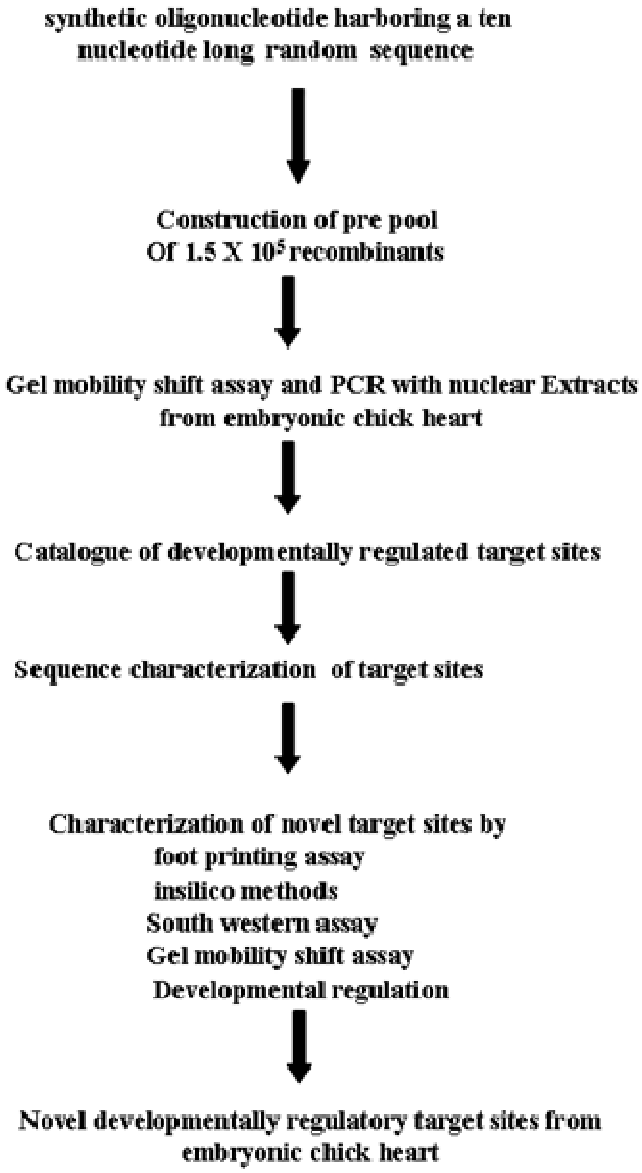

Figure 1. A schematic presentation of the strategy for isolation and characterization of novel developmentally regulatory target site (GTGTT) from embryonic chick heart.

pH 6.8 in stacking gel containing 4% acrylamide (Sigma), 0.106% N, N' methylene bis acrylamide (Sigma), 0.125 M Tris.Cl, pH 6.8, 0.01% TEMED (Amresco) and 0.1% ammonium per sulphate (Amresco). The resolving gel was made of 12% acrylamide, 0.33% N, N' methylene bis acrylamide, 0.375 M Tris Cl. PH 8.8, 0.01% TEMED and 0.1% ammonium per sulphate. The protein samples were electrophresed in running buffer consisting of 0.025 M tris base, 0.192 M glycine, pH 8.3 and 0.1% SDS. After complete run, the gel was transferred to nitrocellulose (Hybond $^{R+}$) membrane for 16 h in transfer buffer containing 0.025 M tris base, 0.192 M glycine (Merck), pH 8.3 and 15% methanol (Qualigen). After complete transfer, the membrane was removed from transfer apparatus and blocked by blotto (50 mM Tris.Cl pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT and 5% skimmed milk powder) for 2 h at 4°C with gentle agitation. Blot was then washed thrice in washing solution (50 mM Tris.Cl pH 7.5, 50 mM NaCl, 1 mM EDTA, and 1 mM DTT). After that, membrane was blocked in binding buffer (10mM Tris.Cl, pH 7.5, 25 mM NaCl, 10 mM $MgCl_2$, 5 mM EDTA, 1 mM DTT) with 200 fold molar excess of poly dIdC for 2 h at 4°C. After poly dIdC incubation, membrane was soaked in binding buffer for 8 h and then soaked in binding buffer containing 10^6 cpm/ml radioactive probe for overnight at 4°C. Membrane was finally washed four times in binding buffer (each wash for 8 min at room temperature). After washing, the membrane was air-dried and exposed to x-ray film or phophorimager.

RESULTS

Screening of "developmentally regulated target sites"

During the past decade, number of laboratories independently developed methods for the isolation of transcription factor binding sites from an unbiased pool of randomized oligonucleotides. However, majority of these studies involve purified recombinant protein, ensuring its efficient binding and selection of the cognate target sites. Nonetheless, methodologies for the isolation of novel target sites by using total nuclear protein have also been described (Nallur et al., 1996). However, such approaches are presumably of limited efficiency due to sub-optimum interaction between multiple transcription factors (and other sequence specific DNA binding proteins) with their cognate binding sites present in the milieu (1 X 10^6 variants or more) sequences. Such constraint is further compounded for less abundant proteins. Our laboratory had developed an efficient method of isolating a large variety of DNA sequences targeted by sequence specific DNA binding proteins.

Briefly, a pre-pool of about 1.5 X10^5 of randomized decanucleotide sequences were constructed in the pBluescript plasmid vector. Independent pool (1 - 1.5 X 10^4) of sequences were selected by four consecutive rounds of gel-mobility shift and PCR by using embryonic chick heart nuclear extract. After four consecutive rounds of selection, we further cloned the embryonic heart enriched target sites in the same vector and prepared the "Catalogue of developmentally regulated target sites." Figure 1 shows a schematic presentation of the strategy adopted for isolation and characterization of novel regulatory target site (ATAGTGTGTT) from embryonic chick heart.

Sequence characterization of regulatory target sites isolated from embryonic chick heart

We randomly picked up around 200 recombinant clones harboring protein target sites and their nucleotide sequences were determined by conventional method see. Figure 2 shows the sequencing profile of five representative

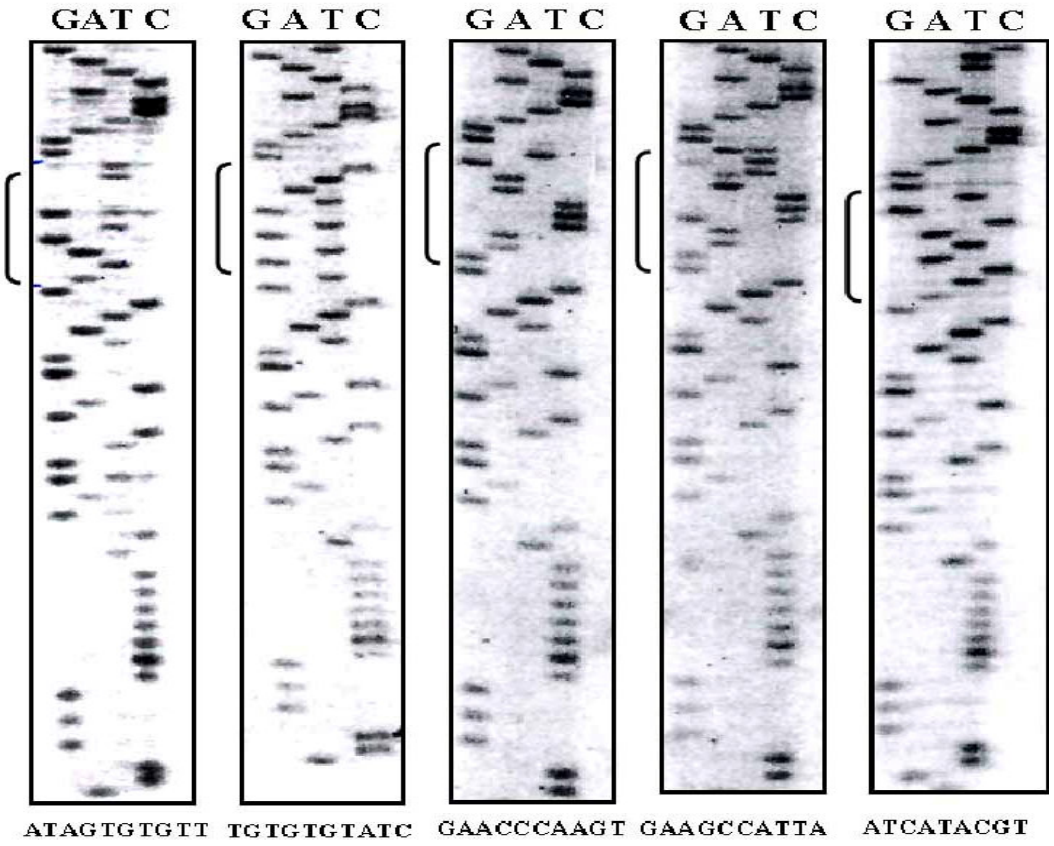

Figure 2. Sequencing of recombinant clones. Cyclic sequencing of pBluescript multiple cloning sites harboring 10 nucleotide long randomized sequences were performed by dideoxy method. Around 200 recombinant clones were sequenced. Sequencing profile of five representative clones is shown. Decanucleotide sequence is marked by bracket. Four lanes are marked as G, A, T, C. ATAGTGGTGTT (sequence of our interest is shown).

clones harboring the randomized sequence. Following sequencing, identity of each sequence was established by comparing with the existing transcription factor binding site database "TRANSFAC" (http://www.gene-regulation.com). Following such comparison, one hundred sequences were "known" (previously reported in the context of various promoters expressing in both cardiac and non-cardiac tissues). The remaining ninety sequences were novel (not listed in TARNSFAC database). Table 1 represents a complete list of such analysis.

Characterization of novel target sites by foot printing assay

After establishing the identity of novel binding sequences, we further characterized them by determining the protein binding sequence. It may be noted that the randomized sequences originally used for selection were ten-nucleotide long. However, that does not necessarily mean that their cognate proteins targeted the entire sequence. The lengths of regulatory target sites are variable (GATA, CANNTG (E-box), TGACTCA (AP-1), TTATTTTTA (MEF2)). It was thus expected that many of the target sites we had selected were smaller than ten base pairs (the length used for their selection). We therefore randomly picked up some of those sequences and identified the exact protein binding sequences by DNAse I foot printing analysis using nuclear extracts from 12-day-old chick embryonic heart and liver. DNAse I protected sequences ranged from pentanucleotide, heptanucleotide, entire decanucleotide, a bipartite tetra-nucleotide sequence with a two nucleotide spacer, or entire decanucleotide sequence along with part of the vector sequences (data not shown). Figure 3 represents the foot print analysis of one such novel target site ATAGTGTGTT (CLONE No. 2 in Table 1). Pentanucleotide region "GTGTT" was found to be protected by the protein extracted from chick embryonic tissues.

In the present study we extensively characterized binding activity of one such novel target site "ATAGTG TGTT" by experimental as well as *insilio* parameters.

Table 1. Compilation of nucleotide sequences of protein target sites. Based on the TRANSFAC database search, each binding sites were classified as known or novel (-).

Clones	Nucleotide sequence	Target sites
Clone 1	GAATCGGTGC	GFI-1
Clone 2	ATAGTGTGTT	-
Clone 3	GTACAGTGTT	-
Clone 4	GTATCCTAGT	GATA
Clone 5	ATGATTGATG	AP-1
Clone 6	TAAGAGTCG	-
Clone 7	TTAGTGTGTC	-
Clone 8	TGAGAGTGTC	-
Clone 9	TACGGGGAGG	IK2
Clone 10	ATAGTGTGTT	-
Clone 11	AATTCAATG	OCT-1
Clone 12	TGAACTCTTG	-
Clone 13	GTACCTAAGC	DELTAF-1
Clone 14	GTCGGTAAGC	-
Clone 15	TACGGCTGCG	AP-4
Clone 16	GTATCTGCT	GATA
Clone 17	CACTAGGTCG	-
Clone 18	ATACTTTTT	-
Clone 19	TGAGGAGTTA	-
Clone 20	TACGGGGAGG	MZF
Clone 21	CATTGGCCC	NF-1
Clone 22	ATCTTGGGCT	-
Clone 23	TAACTACATA	-
Clone 24	TGTAATTGG	CAAT
Clone 25	AGGCACGTGC	MYO-D , USF
Clone 26	TTACATTCCG	OCT-I
Clone 27	GGTAGTATGG	-
Clone 28	GCTAAGTTGT	-
Clone 29	TGTAGCGGGG	vMZF
Clone 30	GAGGGCAATCG	IK-1,2
Clone 31	CATTGGCCC	CAAT, S8
Clone 32	TAAGAGTCG	-
Clone 33	TCTTAATCCT	NKx-2.5
Clone 34	GTGTCCAGG	-
Clone 35	TTACATTCCG	VBP
Clone 36	TGTAATTG	NF-1, CAAT
Clone 37	TCTTGTCG	-
Clone 38	AAAATTCTCT	-
Clone 39	TGTTGGCTCC	NF-1
Clone 40	TGCATACTGC	OCT-1
Clone 41	AGGCACGTGC	ARNT-1
Clone 42	CCTCTGGTGG	-
Clone 43	CTGCATTTGC	-
Clone 44	CCTCTGGTGG	-
Clone 45	CTAAGTCTGT	-
Clone 46	GTCTGTGCGTG	AHRARNT
Clone 47	ATGGGCGTT	-
Clone 48	GGACAAGCTG	AP-4

Table 1. Contd.

Clones	Nucleotide sequence	Target sites
Clone 49	GGACAAGCGT	-
Clone 50	CTGTTTCGGC	-
Clone 51	TTTTCTTTCA	NF-AT
Clone 52	CTACAGTGTC	-
Clone 53	AGTTGCAGCT	AP-4
Clone 54	TTTTCTCTCG	-
Clone 55	TTTTCTTTC	NF-AT
Clone 56	TAGGCGCCGC	-
Clone 57	TAGTCTCACT	-
Clone 58	CAAGGTTGTA	-
Clone 59	TGAAGTATCC	GATA
Clone 60	TACTCTCACT	-
Clone 61	TTATTGGGT	NKX-2.5
Clone 62	TGTTAGGGT	CAAT, NFY, S8
Clone 63	TCAGGTGTAG	E47, MYOD
Clone 64	ATTCGTTGGT	-
Clone 65	TTGGCCCTGG	NF-1
Clone 67	GGGTGATTT	GFI-1 , PADS
Clone 68	CTCCCTATTT	IK2
Clone 69	TTAGTTTGC	-
Clone 70	TTAGTCTGC	-
Clone 71	TGAAGTATCC	GATA
Clone 72	TGTGTGTATC	GATA,
Clone 73	TTCAGTTATC	LMO2COM
Clone 74	GAAGCCATTA	CAAT, S8
Clone 75	GAAGACATTA	-
Clone 76	CTCCCTATTT	IK2
Clone 77	GCATATGGCA	OCT-1
Clone 78	CTGCTTGCGC	-
Clone 79	ATCATACGT	TCF -II
Clone 80	AGACCAACCC	-
Clone 81	GGGTGATTTA	GFI-1
Clone 82	CTTCCTCTAG	NRF-2
Clone 83	GAACCCAAGT	-
Clone 84	GAACCCATTA	S8
Clone 85	GAAGCCATTA	CAAT, S8
Clone 86	TGGCAGAAC	NF-1
Clone 87	ATGTGTTGGG	-
Clone 88	AGTTCTCAT	-
Clone 89	TGGCAGTAC	NF-1
Clone 90	CGGAACTCG	-
Clone 91	GATCACCGGG	-
Clone 92	CGGAACTCG	-
Clone 93	GTCATCACGC	AP-1, TCF-II
Clone 94	ATTAAGGGGT	-
Clone 95	CATGGGGCCG	-
Clone 96	CTGGCTTCAG	NF-1
Clone 97	GGTACGGTT	v & c myb
Clone 98	TTTCATAGAG	TCF
Clone 99	CGTAAGTATG	-

Table 1. Contd.

Clones	Nucleotide sequence	Target sites
Clone 100	CGTCCTACA	-
Clone 101	GTGATGTGTG	v-PADs
Clone 102	AGTTAGGAC	-
Clone 103	TGCCGAGAGC	-
Clone 104	GGTCAAGTCT	RORA1, ER
Clone 105	ATGATTGATG	AP-1, PBX-1
Clone 106	GGCCGAGAGC	-
Clone 107	TAAGAGCTAG	-
Clone 108	TAAGAGCTTG	-
Clone 109	ATGATTGATG	CDPCK3H, TCF-II
Clone 110	TGGCCCACAC	NF-1
Clone 111	TACGACGAGG	-
Clone 112	AATTCAATAG	-
Clone 113	TTACATTCCG	OCT-1, VBP
Clone 114	TGACGTATCC	GATA
Clone 115	AGTTGCAGCT	AP-4
Clone 116	CTGCTTGCGC	-
Clone 117	ATTAACGG	V-MYB
Clone 118	ATTTAACGG	V-MYB
Clone 119	ATTTAACGG	V-MYB
Clone 120	ATTTAACGG	V-MYB
Clone 121	ATTTAACGG	V-MYB
Clone 122	ATTTAACGG	V-MYB
Clone 123	ATTTAACGG	V-MYB
Clone 124	GTGGTTCACC	-
Clone 125	CAATCGGTGG	GFI-1
Clone 126	CAATCGGTGG	GFI-1
Clone 127	TGAATATGCC	OCT -1
Clone 128	ACGGCGGGCT	-
Clone 129	ACGGCGGGCT	-
Clone 130	TGCTGGGCA	-
Clone 131	TGCTGGGCA	-
Clone 132	ACGGCGGCT	-
Clone 133	TTATTCGTG	-
Clone 134	GGCTTTGTG	-
Clone 135	GTACTTTTG	-
Clone 136	CTGGCGGGG	AP-4, SP-1
Clone 137	CTGGCGGGG	AP-4, SP-1
Clone 138	TTGGATGTG	-
Clone 139	AAGTATAGAT	TATA
Clone 140	GCGGTTCGCG	-
Clone 141	GGCTCATTCC	-
Clone 142	GTACTTTTG	-
Clone 143	ATGGGGGTGA	SREBP
Clone 144	TCGTGCTTGT	-
Clone 145	GGATGGGGT	-
Clone 146	GTACTTTTT	-
Clone 147	TGATATGAT	GATA, TCF-II
Clone 148	CGGCGTAGTG	-
Clone 149	TACTTGTTCC	-

Table 1. Contd.

Clones	Nucleotide sequence	Target sites
Clone 150	GGATGGGGTG	-
Clone 151	CCGGCGGGG	AP-2
Clone 152	CAATCGGTGG	GFI-1
Clone 153	TTTGTTGTG	VSTY
Clone 154	GAACCCAAGT	-
Clone 155	GAACCCAAGT	-
Clone 156	GAACCCATTA	S8
Clone 157	GAAGCCATTA	CAAT, S8
Clone 158	TGGCAGAAC	NF-1
Clone 159	ATGTGTTGGG	-
Clone 160	AGTTCTCAT	-
Clone 161	TGGCAGTAC	NF-1
Clone 162	CGGAACTCG	-
Clone 163	GATCACCGGG	-
Clone 164	CGGAACTCG	-
Clone 165	ATTTAACGG	V-MYB
Clone 166	TGCATACTGC	OCT-1
Clone 167	ACGGCGGCT	-
Clone 168	CCGGCGGGG	AP-2
Clone 169	AAGTATAGAT	TATA
Clone 170	CGGAACTCG	-
Clone 171	GATCACCGGG	-
Clone 172	GAAGCCATTA	CAAT, S8
Clone 173	ATTTAACGG	V-MYB
Clone 174	TTAGTGTGTC	-
Clone 175	TTATTGGGT	NKX-2.5
Clone 176	CGTAAGTATG	-
Clone 177	TGCATACTGC	OCT-1
Clone 178	CATGAAAAC	-
Clone 179	GGACAAGCTG	AP-4
Clone 180	TTAGTGTGTC	-
Clone 181	TTAGAGTG	-
Clone 182	TAAGAGTCG	-
Clone 183	TAAGTGTGTC	-
Clone 184	TTATTGGGT	NKX-2.5
Clone 185	TATGAGTCG	-
Clone 186	ATGGGGGTGA	SREBP
Clone 187	CGGAACTCC	-
Clone 188	TAAGA GTCG	-
Clone 189	GAACCCATTA	S8
Clone 190	TTATTGGGT	NKX-2.5

Identification of similar binding sites in the pool

cis-Regulatory target sequences are known for various degrees of variability in the nucleotide bases at different positions. It was also observed that many of the known sequences isolated were different from their canonical

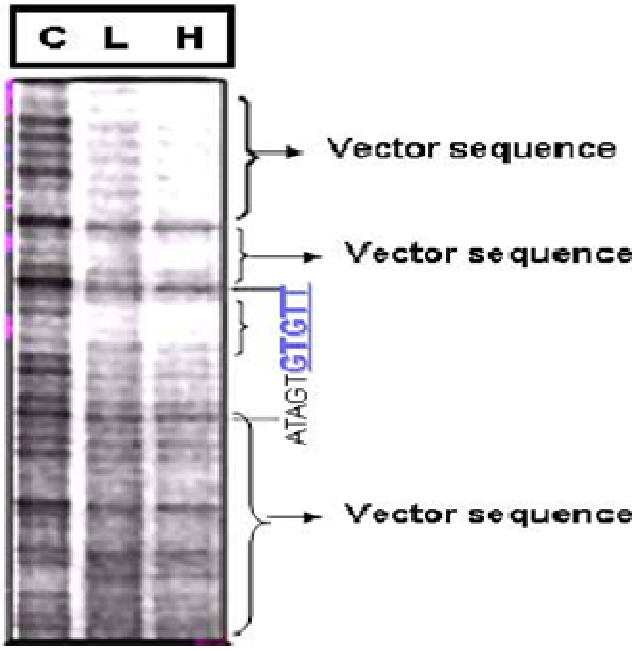

Figure 3. DNAse I Foot printing analysis of candidate clone (2). Left lane is the control digestion (without any nuclear protein), middle lane is digestion with liver extract and the right lane is digestion with heart extract. Protected region is shown by a bracket and the cognate sequence is marked by blue letters and the unprotected sequences are shown in black letters. Protein binding sequences from the vector is also marked. (Vector sequences just flanking the core sequence were also having the protein binding sites that gave the similar protected region in all the clones).

counterparts. We therefore tested whether the binding sequences identified by DNAse I foot printing also occurred in other sequences (unknown category) present in the pool of novel sequences. The number of such binding sequences showed multiple homologues.

To further establish the functional correlations of these homologous binding sites, elctrophoretic gel mobilityshift-competition assay was performed using 12-day-old chick embryonic heart nuclear extract. Radiolabeled probe was prepared from the sequences (ATAGTGTGTT)) characterized by foot printing and its relatedness was tested by competition with fifty-fold molar excess of unlabeled homologous sequences. As shown in Figure 4, clone number 2 (ATAGTGTGTT) upon foot printing showed a protected region GTGTT. Upon sequence comparison, It was found that at least two other sequences (# clone 3 and # clone 87) also harbored the same sequence and two more had similar sequences, clone 181 GAGTG, # clone 187 GAGTC) (Table 1). To find their functional correlations, a gel mobility shift-competition assay was performed using radiolabeled # clone 2 (ATAGTGTGTT) as probe and fifty fold molar excess of those sequences as the competitors. It was observed that sequence # clone 87 (ATGTGTTGGG) and clone 3 (GTACAGTGTT) effectively competed the complex formed by # clone 2, sequence. As expected, the other two sequences; clone 181 (TTAGaGTg) and clone 188 (TAAGaGTcG) which harbored variants of GTGTT, only partially competed the clone 2 specific complex. These data therefore indicate that the same binding protein with different affinities might have targeted these four sequences and led to their isolation.

In-silico foot print analysis for novel target site (GTGTT)

Following experimental identification of exact target sites and their homologous sequences, we also did a computational analysis of the entire pool of sequences to club them into groups of related sequences and thereby determine the exact binding sites (*in silico* foot printing). To that purpose, we used the program Pratt (http://bioweb.pasteur.fr/seqanal/motif/meme/Prattwwww) (Jonassen et al., 1995, 1997) and derived fuzzy sequence patterns, which included at least five sequences. It allows the discovery of conserved patterns only when they exist.

We thus identified twenty such patterns (Table 2). As expected, some of those derived patterns were either corresponding to known transcription factor binding sites such as NFAT-1, NF-1, Nkx 2.5 and GATA or those novel sites which we have already identified by DNAse I foot printing analysis. GTGTT was also found in the analysis which further validates our data. It needs to be noted that in order to avoid any mutual bias in identifying novel sequences, we did the experimental analysis (DNAse I foot printing) prior to the computational analysis. These data therefore established that the novel target site/s can be efficiently used for the identification of conserved sequence motifs, which are likely to be the protein binding sites (transcription factor binding sites).

Promoter analysis for the regulatory target sites

We also searched the eukaryotic promoter database (EPD) (http://cmgm.stanford.edu/ help/manual/database/ epd.html) for finding the occurrences of a number of binding sequences as identified above by DNAse I foot printing analysis. Since most of the binding sequences identified (by *in silico* analysis and DNAse I foot printing) were relatively small (4 - 6 nucleotides), it was expected to have a large number of hits (due to randomness). To eliminate the random occurrences, only those, which occurred in conjunction with other known binding sites or more than once in a promoter were considered.

It was found that the novel target site, GTGTT, occurs in a large variety of promoters including those from plants. While a substantial number of them contained it

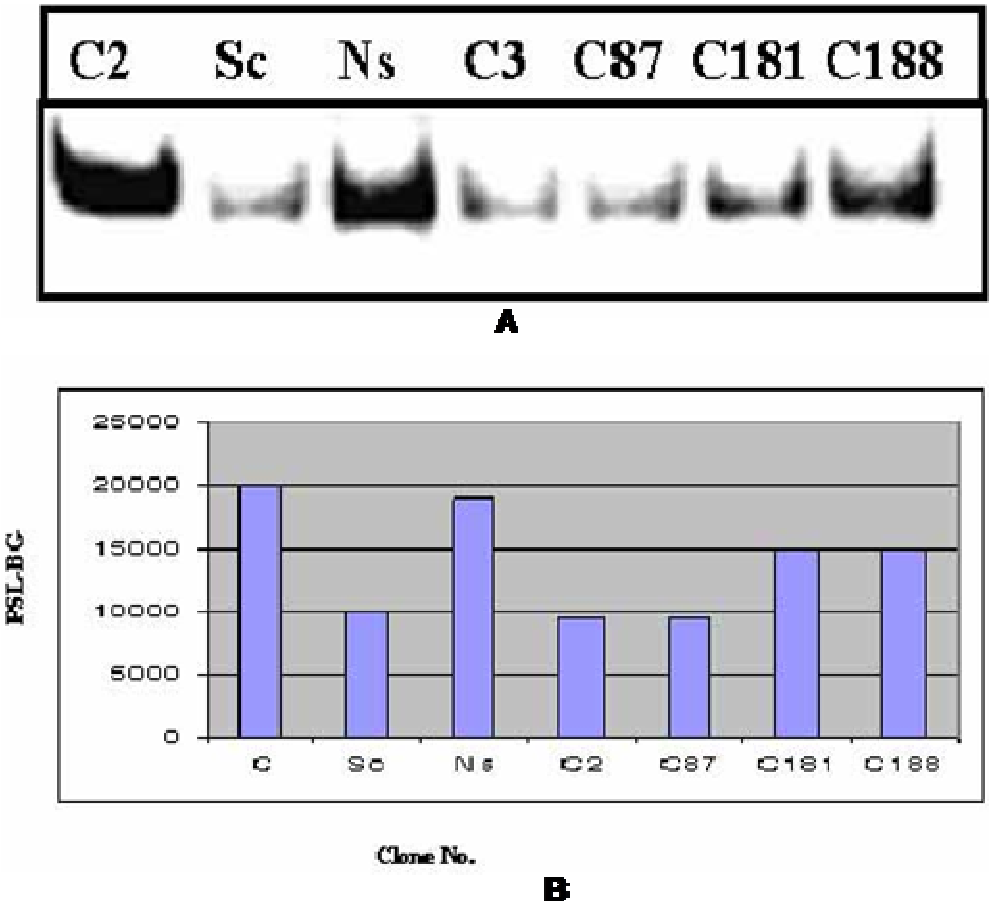

Figure 4. Identification of homologous target site for clone 2 (ATAGTGTGTT) in the pool. Comparison of the DNA binding properties of homologous sequences of clone 2. Panel A: Radiolabeled probe was prepared from clone 2 (ATAGTGTGTT) containing the binding site GTGTT. Fifty-fold molar excess of four independent sequences harboring either completely or partly homologous sequences (Table 1) were used as competitors. The number on each lane represents the sequence number, i.e., 3: GTACAGTGTT, 87: ATGTGTTGGG, 181: TTAGAGTG, 188: TAAGAGTCG. Sc: Specific competitor (fifty fold molar excess of unlabeled self DNA). Fifty-fold molar excess of nonspecific (unrelated sequence) competitor (Ns) was also used; Panel B: quantitation analysis of the binding activities of various clones.

twice, a few contained GTGTT sequence even thrice or more (Table 3). Since the relevance of occurrence of GTGTT sequence once or twice in a particular promoter could not be assessed by such analysis, few other criteria to determine whether those sequences are functionally important were adopted.

Regulatory elements are characterized by their occurrence in clusters and generally, co-regulated genes contain similar regulatory elements. We therefore clustered these promoters (genes) based on their functional relatedness and then looked for other known regulatory sequences immediately neighboring the GTGTT sequence. We thus created clusters in various promoters (interlukin, endocrine hormone promoters).

Figure 5 shows the clustering of conserved target sites in interlukin promoter. The interspecies sequence conservation of our novel regulatory target sites (GTGTT) was also found. Target site was found to be well conserved with the different species and within a gene family (Data not shown).

Characterization of binding activity of Novel (GTGTT) target site by Southwestern analysis

Since our future objective is the molecular cloning of regulatory target site specific proteins, the binding was further confirmed by southwestern assay. Southwestern blotting was carried out by resolving nuclear extracts prepared from chick embryonic tissues (heart, liver and

Table 2. *in-silico* foot print analysis for novel target site.

Sequence	Pattern (score)	Status*
atagTGTGTT	T-G-T-G-T-[CT](24.0745)	FP(GTGTT)
aTGTGTTggg		
atagTGTGTT		
aTGTGTTggg		
ttagTGTGTC		
atAGTGTGtt	A-G-T-G-T-[GT](24.0506)	FP(GTGTT)
atAGTGTGtt		
gtacAGTGTT		
gtagAGTGTT		
ttAGTGTGtc		
TGTTGGctcc	T-[CG]-T-T-G-G (24.1729)	FP (CTTG)
aTCTTGGgct		
atgTGTTGGg		
aTCTTGGgct		
atgTGTTGGg		
aTAGTGTgtt	T-A-G-T-[CG]-T (24.1729)	ND
aTAGTGTgtt		
TAGTCTcact		
tTAGTCTgc		
tTAGTGTgtc		
atAGTGTGtt	A-G-T-[CG]-T-G (24.1729)	ND
atAGTGTGtt		
ctaAGTCTGt		
ttAGTCTGc		
ttAGTGTGtc		
ATGGGCgtt	A-T-G-G-G-[CG](24.1729)	ND
ATGGGCgtt		
cATGGGGccg		
ggATGGGGt		
ggATGGGGtg		
GAACCCatta	G-A-A-C-[CT]-C (24.0745)	ND
cgGAACTCg		
cgGAACTCg		
GAACCCaagt		
GAACCCaagt		
tGAACTCttg		
aTACTTTtt	T-A-C-T-[CT]-T 24.0745	ND
aTACTTTtt		
gTACTTTtg		
gTACTTTtt		
TACTCTcact		

Table 2. Contd.

Clones	Nucleotide sequence	Target sites
atACTTTTt	A-C-T-T-[GT]-T (24.0506)	ND
atACTTTTt		
gtACTTTTg		
gtACTTTTt		
tACTTGTtcc		
aTACTTTtt	T-A-C-T-T-[GT](24.0506]	ND
aTACTTTtt		
gTACTTTtg		
gTACTTTtt		
TACTTGttcc		
GTGTGTatc	G-[AT]-G-T-G-T (24.0483)	ND
ataGTGTGTt		
ataGTGTGTt		
gtaGAGTGTt		
tgaGAGTGTc		
ttaGTGTGTc		
ttaTTGGGt	T-T-G-G-G (20.8503)	Nkx 2.5
atcTTGGGct		
atgtgTTGGG		
atcTTGGGct		
atgtgTTGGG		
GTATCctagt	G-T-A-T-C (20.8503)	GATA
GTATCtgct		
tgtgtGTATC		
tgaaGTATCc		
tgacGTATCc		
ttTTCTTTc	T-[AT]-C-T-T-T (24.0483)	NFAT-1
ttTTCTTTca		
aTACTTTtt		
aTACTTTtt		
gTACTTTtg		
gTACTTTtt		
attAAGGGGt	A-[AT]-G-G-G-G (24.0483)	FP (TAAGGGG)
attAAGGGGt		
cATGGGGccg		
ggATGGGGt		
ggATGGGGtg		
atTAAGGGgt	T-A-A-G-[AG]-G(24.0204)	FP(TAAGGGG)
atTAAGGGgt		
TAAGAGctag		
TAAGAGcttg		
TAAGAGtcg		

brain). The proteins were electro blotted to nitrocellulose membrane and probed by radiolabelled novel target site. Southwestern data showed the heart enriched binding activity of target site, ATAGTGTGTT (Figure 6). Numbers of complexes formed in gel (identified by a thick arrow in

Table 2. Contd.

Clones	Nucleotide sequence	Target sites
ATAGTGtgtt	A-[CGT]-A-G-T-G 23.5388	ND
ATAGTGtgtt		
ctACAGTGtc		
gtACAGTGtt		
gtAGAGTGtt		
tgAGAGTGtc		
tGTTGGCtcc	G-T-T-G-G-[CGT](23.5388)	NF-1
atgtGTTGGG		
attcGTTGGT		
atgtGTTGGG		
attcGTTGGT		
atgtGTTGGG	G-[ACT]-T-G-G-G (23.5351)	ND
atgtGTTGGG		
gGATGGGgt		
gGATGGGgtg		
tGCTGGGca		
tgtTAGGGT	T-A-G-[AGT]-G-T (23.4591)	ND
aTAGTGTgtt		
aTAGTGTgtt		
gTAGAGTgtt		
tTAGTGTgtc		
atctTGGGCT	T-G-G-G-C-[AGT](23.4591)	FP(TTGGGCT)
aTGGGCGtt		
atctTGGGCT		
aTGGGCGtt		
tgcTGGGCA		
ctccCTATTT	C-T-[AGT]-T-T-T (23.4591)	ND
ataCTTTTT		
ataCTTTTT		
CTGTTTcggc		
gtaCTTTTT		
atagtGTGTT	G-T-G-T-T (20.8503)	FP (GTGTT)
atGTGTTggg		
atagtGTGTT		
atGTGTTggg		
gtacaGTGTT		
gtagaGTGTT		

* FP: target site was independently confirmed by foot print analysis. ND: Foot printing analysis has not been done. GTGTT the sequence for characterization.

both panels) are due to the binding of proteins to the vector sequences adjoining the core randomized sequences. The southwestern assay in stage specific and species specific manner was also performed which also confirms the fact that the novel regulatory target site is essentially expressed in chick embryonic heart (Data not shown).

Characterization of binding activity of novel target site by gel mobility shift assay

We further confirmed the binding of our novel target site (ATAGTGTGTT) by gel mobility shift assay by using nuclear extract from embryonic chicken heart. Gel mobility shift assay was done with probe prepared from a double stranded synthetic oligonucleotide harboring only the core sequence (ATAGTGTGTT) of novel target sites (with no flanking sequence). As shown in Figure 7 binding activity is specific for the novel target sites as 50 ng of nonspecific competitor is not able to compete in the assay.

Developmental regulation of novel target site

Finally, the developmental expression of the binding activity of novel target site (ATAGTGTGTT) was tested. For that, the nuclear extract from various stages of chick embryonic heart forming regions was prepared and probed with the radiolabeled novel target site. The expression was found to be developmentally regulated. It was found that its expression was moderate in early stages of chick development (4 - 8th day), high in middle stages (12 – 16th day) and very less in later stages of development (Figure 8).

DISCUSSION

Eukaryotes are characterized by enormously complex genome organization which is decoded by highly co-ordinated molecular interactions involving multiple regulatory modules such as cell signaling cascades, a large repertoire of sequence specific DNA binding proteins (transcription factors and other related proteins), chromatin modifiers and numerous small DNA elements located in the regulatory regions of various genes. Although tools of cellular and molecular biology have enhanced our understanding about various cell signaling pathways and mechanisms of transcription, our knowledge about the regulatory DNA elements specifying the spatial and temporal pattern of gene expression is inadequate. A complete catalogue of eukaryotic DNA elements (with positional weight matrices) targeted by transcriptional and other sequence specific regulators are a major challenge to the genome biologists.

The present study involves primary characterization of a library of regulatory DNA elements (<10 base pairs) constructed in our laboratory recognized by nuclear proteins from embryonic chick heart. Genetic cascades regulating cardiac development are highly conserved

Table 3 Compilation of promoters which contain the novel target site GTGTT at least three times or more within the first 500 base pairs from the transcription start site.

Name of the gene and species	Number of sites in the Proximal promoter (-500)
a. Ribosomal protein S15, Human	5
b. Von Willebrand factor, Human	4
c. Aldolase B, Human	3
d. TOPK, a mitotic kinase, Human	3
e. MBLL39, Factor that co-localize with nuclear	3
f. foci of expanded-repeat transcripts, Human	3
g. SLC38A2, Interacts with melanin-concentrating hormone, Human	3
i. SF3B1, Splicing factor, Human	3
j. BNIP3L, Bcl-2 family protein, Human	3
k. TGM2, transglutaminase 2, Human	3
l. PRG1, Human	3
m. Human PSMD1, Human	3
n. DLK1, death associated protein kinase homologue, Human	3
o. TSFM, Mitochondrial Translation factor, Human	5
p. Human POLR2C	4
q. Tenascin, Chicken	3
r. Vitellogenin II, Chicken	3
s. SnRATA U2, Mouse	3
t. SnRATA U1a, Mouse	3
u. SnRATA U2, Mouse	3
v. FSH b, Rat	5

from *Drosophila* to human. Therefore, identification and characterization of targeted DNA elements involved in cardiac development is also relevant in the context of understanding the general mechanisms of transcription.

The first part of this study involved sequence characterization of 200 candidate target sites (for sequence specific DNA binding proteins) (Table 1 and Figure 1). About 50% target sites were represented binding sites for known transcription factors, thereby indicating that the remaining unique sequences are likely targets of yet unknown transcriptional regulators. Due to the non-availability of databases for other sequence specific DNA binding proteins, verification could not be made whether any of those novel sequences represent such activities. It thus likely that the present collection contains thousands of novel binding sites (but not necessarily an equal number of novel binding proteins) present in the pool.

Since the long term objective of this study is to create a comprehensive catalogue of regulatory target sites, a number of criteria were subsequently imposed for assessing the authenticity of those novel sequences. As required, majority of the tested sequences showed specific binding activities (study done by other researchers in the laboratory) thereby establishing that the novel sequences are the bonafide targets for gene regulatory proteins.

Since the candidate target sites were selected from a pool of decanucleotide sequences, the protein binding sequences they harbored were likely to be restricted to 4 -10 bases (the minimum length of DNA sequences targeted by sequence specific DNA binding proteins are four). A number of such sequences were subsequently identified by DNAse-I foot printing analysis and as expected, their lengths varied from four to eight nucleotides (except only two binding sites that spanned into the adjoining vector sequence). Foot printing analysis performed established the diversity as well as the authenticity of these candidate target sequences. In a pool of random sequences, the frequency of occurrence of any particular sequences would be decreasing with increase in length. A tetra nucleotide sequence will thus occur once in every 256 bases while a penta-nucleotide sequence will occur once in every 1016 bases. It is thus expected that among the selected sequences, smaller binding sites would occur more frequently than the larger sequences (that is, in proportion to their frequency of occurrence in the pre-selection pool). In agreement to such proposition, we observed that the smaller binding sites such as GTGTT had number of homologues (similar sequences with variations in one or two nucleotide positions) as identified by sequence comparison. Functional relatedness among these (similar) sequences was thereby

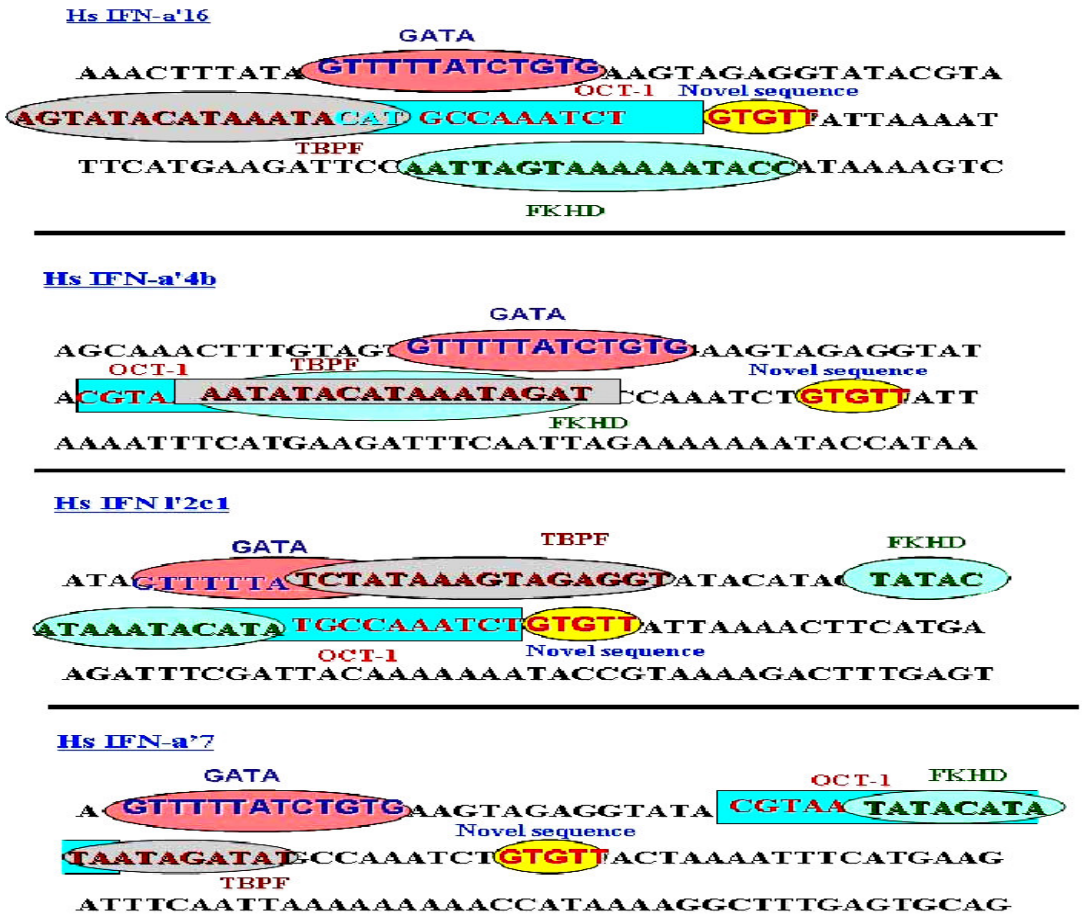

Figure 5. Clustering of GTGTT and other known *cis*-regulatory elements in interlukin genes.

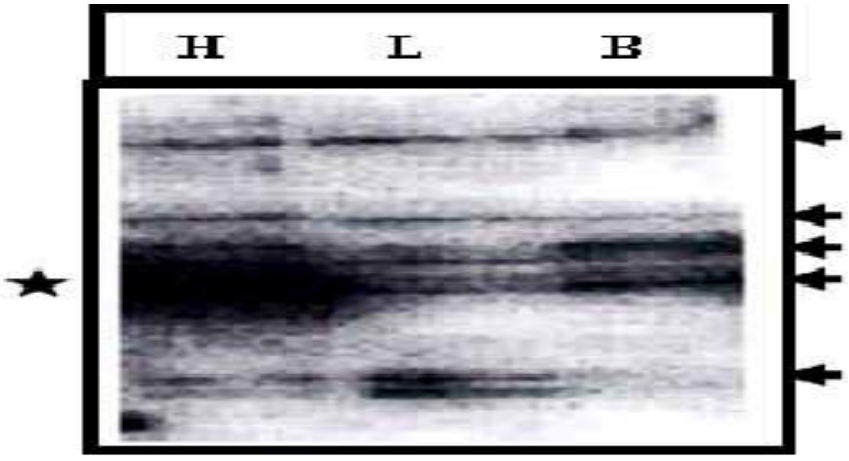

Figure 6. Characterization of binding activity of novel (GTGTT) target site by Southwestern analysis. South western blotting was carried out by resolving nuclear extracts prepared from chick embryonic heart (H), brain (B) and liver (L) on SDS-PAGE. The proteins were electro blotted to nitrocellulose membrane and probed. Specific complexes in figure are identified as ★. Number of complexes formed in gel (identified by a thick arrow) is due to the binding of proteins to the vector sequences adjoining the core sequences as observed in foot print analysis.

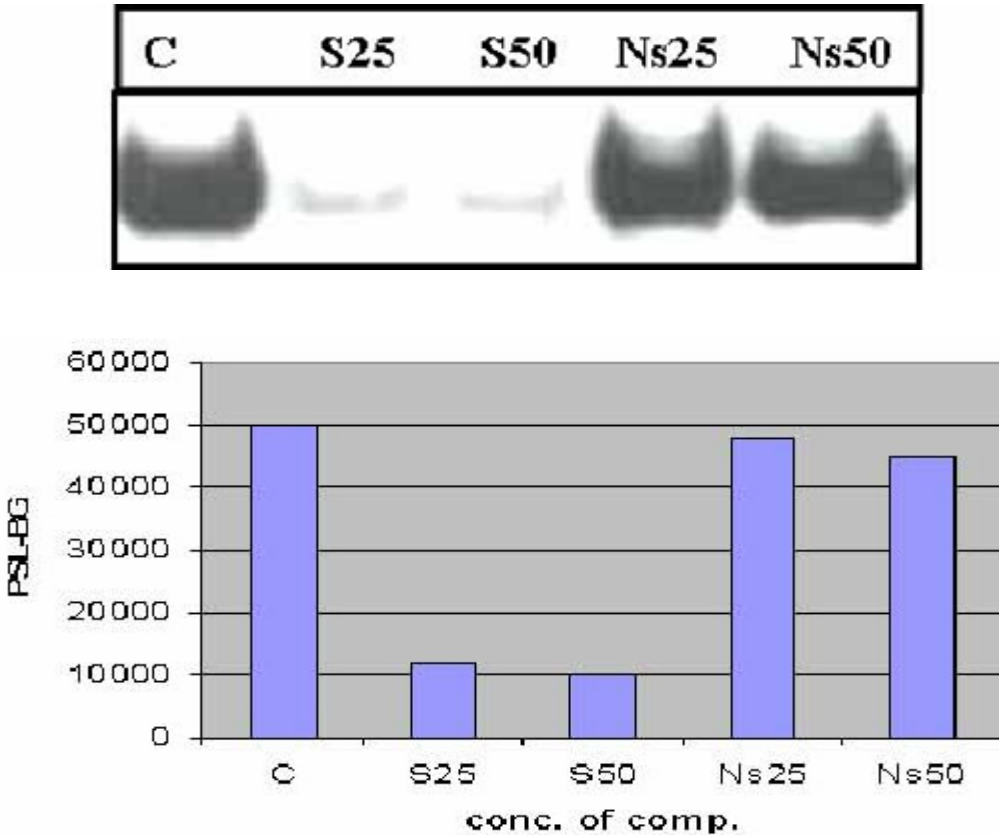

Figure 7. Characterization of binding activity of novel target site by gel mobility shift assay. Panel A-Gel mobility shift-competition assay was carried out with radiolabeled probe prepared from clone 2 monomer (ATAGTGTGTT, only core sequence) and incubated with embryonic chick hear extract. Lane 1: Control (no competitor was added.) Lane 2 and 3: 25 and 50 ng self competitor respectively (cold oligo) was added in the reaction. Lanes 4 and 5: 25 and 50 ng nonself competitor was added in the reaction, respectively; Panel B- quantitation analysis of the binding activities of various competitions is shown.

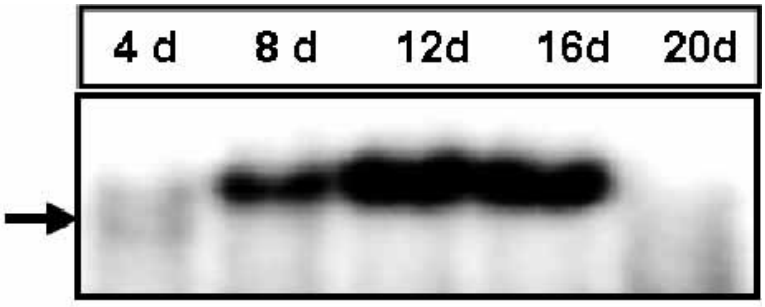

Figure 8. Developmental regulation of novel binding activities. Gel mobility shift assay was carried out on radiolabeled probe prepared from novel target site (ATAGTGTT). Five micrograms of nuclear extract prepared from heart forming regions of day 4, 8, 12 16 and 20 chick embryos were used. Specific complex is shown by arrow.

confirmed by gel mobility shift-competition assay. Thus a number of known transcription factor binding sites as well as unknown binding sequences (GTGTT) had multiple occurrences in the pool of sequences characterized.

The catalogue of regulatory target sites constructed is presumably too large for the experimental determination of each binding sites. A computational approach for identifying more target sequences (novel) was thus developed. Using the program “Pratt” we identified a number of frequently (at least five) occurring patterns and verified their authenticity by matching them with authentic binding sites (either experimentally identified or from TRANSFAC data base). Occurrence of a number of known as well as novel target sequences among these patterns demonstrates the usefulness of such analysis in predicting novel target sequences. We also used another parameter, that is, co-occurrence of a novel sequence in clusters with other known transcription factor binding sites in eukaryotic promoters. Analysis with the GTGTT sequence showed that it occurred in clusters with a set of known transcription factor binding sites in histone and interleukin promoters from multiple species thereby indicating its role in regulating interleukin and histone gene expression. We further characterize the binding activity of novel target site by gel mobility shift assay and south western assay which also confirms the heart enriched binding activity of candidate target sites. In the final study, we tested the concept of differential gene expression. For that, we tested the binding activity of our novel sequence by using the heart nuclear extract from various staged of chick embryonic development. Our data clearly shows the context dependent expression of novel target site.

In conclusion we propose a number of possible applications: (i) This novel target sites can be directly used for cloning the cognate protein; (ii) the cognate protein can be expressed and purified; (iii) the targeted protein can be extensively characterized; (iv) the information can be used for identifying the novel expression pathways by insilico analysis of the respective genome.

REFERENCES

Ahuja P, Sdek P, MacLellan WR (2007). Cardiac myocyte cell cycle control in development, disease, and regeneration. Physiol. Rev. 87: 521-544.

Barker DJ (2007). The origins of the developmental origins theory. J. Int. Med. 261: 412-417.

Birnbaum K, Benfey PN, Shasha DE (2001). Cis element/transcription factor analysis (cis/TF): a method for discovering transcription factor/cis element relationships. Genome Res. 11: 1583-1590.

Boyle S, De Caestecker M (2006). Role of transcriptional networks in coordinating early events during kidney development. Am. J. Physiol. Renal. Physiol. 291: 1-8.

Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. Anal. Biochem. 72: 248-254.

Bryan L, Tjian R (2000). Orchestrated response: a symphony of transcription factor for gene control. Genes Dev. 14: 2551-2569.

Corry GN, Underhill DA (2005). Subnuclear compartmentalization of sequence-specific transcription factors and regulation of eukaryotic gene expression. Biochem. Cell Biol. 83: 535-547.

Grondin B, Lefrancois M, Tremblay M, Saint-Denis M, Haman A, Waga K, Bedard A, Tenen DG, Hoang T (2007). c-Jun homodimers can function as a context-specific coactivator. Mol Cell Biol. 27: 2919-2933.

Gross P, Oelgeschläger T (2006). Core promoter-selective RNA polymerase II transcription. Biochem. Soc. Symp. 73: 225-236.

Hochheimer A, Tjian R (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev. 17: 1309-1320.

Ian BJ, Stephen FM, McGettigan PA, Guy P, Aedín CC, Desmond GH (2007). Integrating transcription factor binding site information with gene expression datasets. Bioinform. 23: 298-305.

Jonassen I, Collins JF, Higgins DG (1995). Finding flexible patterns in unaligned protein sequences. Protein Sci. 4: 1587-1595.

Jonassen I (1997). Efficient discovery of conserved patterns using a pattern graph. Cabios. 13: 509-522.

Keller BB, Liu LJ, Tinney JP, Tobita K (2007). Cardiovascular developmental insights from embryos. Ann. N.Y. Acad. Sci. 1101: 377-388.

Laemmili UK (1970). Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. Nat. 227: 680-685.

Nallur GN, Kulkarni P, Weissman SM (1996). Multiplex selection technique (MuST): An approach to clone transcription factor binding sites. Proc. Natl. Acad. Sci. USA. 93: 1184-1189.

Nemer G, Nemer M (2001). Regulation of heart development and function through combinatorial interactions of transcription factors. Ann. Med. 33: 604-610.

Saurabh S, Martin T (2002). Discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res. 30: 5549-5560.

Sindhu KV, Vibha R, Manveen KG, Debapriya C, Shyamal KG (2004). Isolation of a library of target-sites for sequence specific DNA binding proteins from chick embryonic heart: a potential tool for identifying novel transcriptional regulators involved in embryonic heart. Biochem. Biophys. Res. Commun. 323: 912-919.

Venter JC, Adams EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans RA, Holt RA, et al (2001). The sequence of the human genome. Science, 291: 1304-1351.

Wyeth WW, Albin S (2004). Applied bioinformatics for the identification of regulatory elements; Nat. Rev. Genet. 5: 276-287.

Yang E, Simcha D, Almon RR, Dubois DC, Jusko WJ, Androulakis IP (2007). Context specific transcription factor prediction. Ann. Biomed. Eng. 35:1053-1067.