Homology Modeling and Characterization of Novel Stress Inducing Genes and their Split Sites Recognition in the Genome of Zebra Fish (*Danio Rario*) by Employing Special Consensus Pattern Matching Algorithms- using an in-Silico Method.

J.Venkateshwara Rao

Abstract: ***Zebra fish has long been considered to be as a strong animal model in biology and modern genetics; however now a days its gaining lot of importance in environmental studies as well. The readily availability of entire genome sequences made to permit carrying out in silico studies at Genomic level. As everyone is known that stress is much more complex and complicated process that involves so much of gene regulations known as up regulation and down regulation, the corresponding stress proteins, broadly known as heat shock proteins. In the current study, the potential transcription factor binding sites were traced out by using bioinformatics tools and about 50 heat shock protein genes were predicted by using special alogorithms using pattern matching and position weight matrices. The 3D structure of DNA-binding domain of HSTF-1 (Heat Shock Transcription factor-1) which is crucial for regulating heat shot proteins was traced out and builted by using homology modelling methods. The 3D structure of the heat shock transcription factor-1 and together with predicted transcription factor binding sites may be validated in future experimental works which would help us in understanding the complex responsive stress mechanisms lying in Zebra fish.***

Keywords : DBD, HSP Proteins, Motif, Transcription Factors

I. INTRODUCTION

Since the exploration of genetic material and fast development of molecular genetic tools over the last 75 years has remarkably witnessed the tremenderous growth in the biological knowledge base and exploitation of unknown genetic information and understanding the under laying mechanisms of how an organism undergoes various adaptive changes. All the living organisms undergoes a rapid molecular response to adverse environmental factors, which we well known as Heat Shock Responding mechanisms (Lindquist, 1986). The enhanced Heat Shock Reaction is characterized by means of an increase in expression of certain group of specific proteins known as Heat Shock Proteins (Hsp), which have been deeply conserved in nature (Lindquist, 1988). In Eukaryotic Organisms, these enhanced levels of gene expression has been shown to be regulated by heat shock transcription mechanisms. Knowing of hsp sensing genes provide us very enhanced model system for knowing of gene regulation at transcription stage itself. Characterization of these stress genes is very much crucial and fundamental step in understanding the insights into various levels of adaptation and acclimatisation of various organisms. Stress is a very serious situation of condition produced by several environmental or other factors which creates disturbances in the homeostasis of an organism and balance is threatened as a result of imposed external stimuli which can be referred as stressors. Fish are generally known to more exposed to stressor both in natural and artificial conditions and undergo a continuous neuroendocrine, physiological and behavioural changes in a defensive way of attempt to resist changes and compensates the challenges imposed on them, there by challenging the stress (Feder and Hoffman, 1999). These adaptations are considered to be compensatory or habitutative mechanisms that allow the organism to cope up with stressor in order to maintain its balance and ultimately help for its survival. *In-Silico*, experimental methods are familiar with to be performed on computer or by means of using computer simulation methods. The first such kind of experiments was carried out by Mirmontes (1992) in the workshop conducted with title “Cellular Automata: Theory and Applications” to illustrate life science experiments carried out by using wholly in a computer. The rate at which new sequences are being added up to the databases had a considerable effect in the fields of computational bioinformatics. A major amount of hardship and exertion being spend on how to effectively and resourcefully hold and right to use these data and as well as incorporating new methods intended at mining these data mines to line up to support new biological discoveries. Novel databases are required to reduce the burden of the existing databases and analyzing for a specific requirements of end users and also at the same time user friendly to cater specific needs of all the users.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Dr.J.Venkateshwara Rao*, Asst.Professor, Department of Zoology, Osmania University, Telangana . Email Id:venbio@gmail.com

Past few decades have witnessed the usage of Zebra fish as a model vertebrate organism primarily to understand molecular genetics, and increasingly for toxicological experiments and environmental monitoring and
other studies as well. In 2001 Year, Sanger Institue has started the Zebra fish genome sequencing project and released several genome assemblies, the latest being Zv11 in the public database. The awareness and easy access of genomic data has resulted in large scale explosion of in silico approaches to extract biological information from the genomic sequences. Now a days, in silico approaches are extensively used to carry out genome level identification and characterization of important stress genes. Studying the mechanisms of stress response and its regulation in model organisms like Zebra fish at genome level would help us in better understanding of stress factors, the knowledge which could be further applied in rearing of important stress free domesticated animals and crops.

II. REVIEW OF LITERATURE

Heat shock proteins were extracted from large number of variety of fishes and subsequently were cloned. They were cloned from Tilapia (*Oreochromis mossambicus*), Zebra Fish (*Danio rerio*), Pufferfish (*Fugu rubripes*) and rainbow trout (*Oncorhynchus mykiss*). Pan et al. (2000) characterized an hsp 90 nucleotides from Atlantic salmon fish (Salmo salar) that corresponds to Zebrafish hsp90b with a 92% identical amino acid characteristics.Heat Shock Proteins Genome Organization: We know very modest information about the genomic sequence, genomic structures or composition of the genes that encoding heat shock proteins in various fishes. Hsp genes were cloned from few number of fishes so far. HSP analysis and comparisons between of their promoters from a vast range of organisms leads to recognition of a palindormic heat shock elements (HSE), with a sequence of CnnGAAnnTTCnnG reported by (Bienz and Pelham, 1987). It has been confirmed that hsp induction proceeds primarily from the firm binding of an activated hsp transcriptor factor (HSF with HSE's upstream Hsp genes (Morimoto, 1992). in view of the fact that, majority of the inducible hsp genes do not possess portions of intron regions, the mRNA is swiftly converted into promising functional protein with in no time following exposure to a stress bursters (Basu, 2002).First main obligatory movement for specific task regulation of the gene expression is interactions between regulated key proteins and specific targeted DNA sequences. The binding of cis-acting sequences with regulatory proteins determined by the consensus recognition motif of DNA molecule. A key transcription factor which is known as Heat Shock Factor (HSF) regulates transcription level, where it binds to conserved region found in the upstream portion of heat shock protein called as the heat shock element (Heiss and Duncan, 1956). The upstream portion of (5') regulatory elements of various hsp genes are very much comparable with those of other genes mutually with the existence of TATA box, CCAAT box and transcription start regions. Thakurta *et al.* (2002) have recognized cis based regulatory elements concerned in the hsp reaction process in *Caenorhabditis elegans* where upstream regions of these genes were thoroughly scanned for heat shock factors using computational DNA pattern recognition methods. A continuous appearances of inverted repeats of the pentamer nGAAn (Enokia et al.,2011) represents heat shock elements. The DNA based binding domain known as (DBD), a hydrophobic nature repeated region (HR-A or B) is needed in establishing of homo-trimer formation and also for execution of transcription based activation domain at C- terminus, which represents complete functional domains of HSF proteins. Westerheide and Morimoto (2005), The binding of DBD domain in each monomer with 5-bp sequence corresponding to 5'-nGAAn-3' represents homotrimerization. This starts with HSE begin with either of 5'-nGAAn-3' or 5'-nTTCn-3' repeats. The main factor which involved in stress and enhancement of hsp genes is HSFS1 (Sistonen,1994) while HSF2 and HSF3 shows similar activity in tissue dependent activation processes (Goodson et al., 1995).

III. MATERIAL AND METHODS

1. Data Collection:

Zebrafish genome: The Zebra fish genome consists of 25 chromosomes and has an estimated haploid size of 1.14 GB ande 11623 scaffolds. The Zebrafish, ZrcZ11 assembly (danRer11) was downloaded from the UCSC genome browser during May 2019 from its ftp siteAnd downloaded from ftp://hgdownload.soe.ucsc.edu/goldenPath/danRer11/.

Assembly statistics: A new kind of library was generated from haploid of Zebra fish, the WGC assembly was based on 20,541, 433 reads comprising 14, 160, 626, 498 base paris with a coverage of 6.5X. The programs like Phusion was utilized to cluster the reads and Phrap was emploed for cluster assembly formation and consensus generation.

The Zebra fish gene and Finding of Gene Prediction Routes: The zebrafish gene & gene prediction routes was obtained from the UCSC browser found at the URL (http://genome.ucsc.edu/cgi-bin/hgTables) for ZrcW11 assembly which gives the gene information coordinates of all the refSeq genes in zebrafish.

Heat shock protein (HSP) genes: The heat shock protein genes were collected by a combination of literature search and database searching. The heat shock protein genes were downloaded from NCBI using the query heat shock proteins under taxa '*Danio rerio'*. Accordingly, a total of 142 heat shock reference sequences were downloaded and stored in FASTA format.

The Homology searching: The homologues of 10 small HSPs were obtained from other species by performing similarity search BLASTP against the NR databases with the parameters, expect threshold as 1 and BLOSUM matrix 80.

2. Computational Programs and Software Used:

NCBI BLAST: A stand-alone application of NCBI BLAST tool version 2.2.22 was downloaded as BLAST Tool executables during the November 2010 from the wbsite NCBI URL ("ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/").

ClustalX2: ClustalX version 2 was downloaded from its official website (www.clustal.org) to perform multiple sequence alignment analysis of nucleic acids and protein sequences.

Match Tool: The Match tool version 1 was utilized online for locating potential binding sites for transcription factors and transcription biding sites in nucleotide sequences. It makes use of a library of (mono nucleotide bp) weight matrices from the database ofTRANSFAC6.0 (http://www.gene-regulation.com/pub/programs.html#match).

DNA Pattern Find: DNA Pattern Finding tool Sequence Manipulation Suite version of 2 was used as an online tool for pattern-matching.
""http://www.bioinformatics.org/sms2/dna_pattern.html".

Prot Param: The sequences of functional domains of small heat shock proteins were subjected to physicochemical properties analysis using the web tool Prot Param (http://www.expasy.ch/tools/protparam.html).

Modeller 9v8: Modeller 9v8 version for protein structure determination by homology modelling was downloaded as a stand-alone utility to predict protein structure by homology modelling
(http://salilab.org/modeller/download_installation.html).

Accelrys DS Visualizer: DS-Accelrys Visualizer version of 2 was downloaded for screening the 3D structures of the proteins ("http://accelrys.com/products/discovery-studio/").

XAMPP: XAMPP version 2.5.8 a multi cross-platform web server providing solutions using Apache HTTP Server, MySQL database, and interpreters for scripts written in PHP and Perl programming languages was used for database designing ("http://www.apachefriends.org/en/xampp.html").

Prediction of Potential Transcription Factor Binding Site:

DNA Pattern Find tool of the Sequence manipulation suite 2 was used to scan the 1000 base pairs upstream regulatory regions of the HSP genes for the heat shock element consensus region by the patterns **n**GAA**nn**TTC**n** and **n**GAA**nnnnnn**TTC**n** where '**n**' can be any nucleotide.

Splice Site Extraction: Splice Site extraction has been done using Perl programming. The Zebrafish gene prediction track file contains all the information needed to extract splice sites which was downloaded from UCSC Table browser. This file contains the name of all the genes, their orientation and chromosome number location along with information of start and end coordinates of the strand, exon and coding sequences (cds). By employing perl programs, the above file was split into 25 files based on the chromosome number so that data handling could be simplified.

Classification of splice sites: Splice sites were primarily classified into four splice site subtypes "U2 type GT--AG, GC--AG and AT—AC" and "U12 type GT—AG" on the basis of conserved region of exon-intron boundaries with the help of a perl script.

IV. RESULTS

Identification of Potential HSP Transcription Factor Binding Sites

HSP transcription factor binding sites: Identification of potential transcription factor binding sites like CAT box, MyoD Box and TATA box were obtained for Hsp genes. These predictions were then narrowed down to ten genes based on the positions where these transcription factor binding sites normally lie through literature search. Out of the ten genes four accords to the small Hsp protein family, five corresponds to the HSP 40 family and one to HSP 70 family (Table 1).

Heat shock element prediction: Heat shock elements were predicted for 24 genes in the 1 Kb upstream regions. The predicted heat shock consensus elements were classified into two categories, one with the consensus motif **n**GAA**nn**TTC**n** and the other with **n**GAA**nnnnnn**TTC**n** where '**n**' can be any nucleotide. Majority of the genes identified (eleven) to possess these heat shock elements belong to the members of the HSP 40 protein family. Eleven genes belonged to the first type and the remaining thirteen genes to the second type (Table 2 and 3).

Genome Stretch Identification of HSP Transcription Factor Family: The peptide sequence of Hsf -DBD domain (Pfam ID:PF00447) was extracted as a query in BLASTP searches for all likely homologues determined in the zebrafish Genome and 26 hits were generated, which were described in the NCBI database as heat shock transcription factors.

3D structure Modelling of DNA binding domain Protein of Zebrafish Hsp transcription factor 1

The predicted 3D protein structure modelling of the DDB (DNA binding domain) of zebrafish heat shock factor has 106 amino acids starting with alanine (A) and ending with valine (V). The structure consists of an anti-parallel structures of four-stranded beta-sheets (β1to β4) packed against a bundle of four alpha helices. The first helix is right at the beginning of the structure stretching from the second amino acid to the tenth amino acid followed by two antiparallel beta strands. After a small distance, the second and the third helices are present while the fourth helix is at the end of the domain ranging from the 93rd position to the 100th position. The backbone of the predicted structure fits the template selected as seen by the superimposition of the two structures carried out in swiss pdb viewer. The RMS value of the predicted model was found to be 0.58 A°. The structure Ramachandran plot revealed that 94.6 percent of residues were found in the most favoured region and 5.4 percent in the allowed region with no residues in the outlier region (Fig 1 and 2).

Small Heat Shock Protein Sequence Analysis

Functional Elements Extraction

Extraction of mRNA, Introns, Exons and UTR sequences

The genomic coordinates of functional elements like exons, introns, mRNA, 3' UTR and 5' UTR were obtained from modifying the gene prediction track by using a perl script developed by us. A total number of 136252 exon sequences, 56261 mRNA sequences, 56261 5' and 3' UTR sequences and the entire intronic sequences were extracted and stored in FASTA sequence format.

Extraction of splice sites: Based on the genomic coordinates of exons start and end positions, the terminal end sequences at exon/intron and intron-exon junctions were identified by, 38nt long at 5' splice site (8nt lengtbh within the region of exon, 30 nucleotide bases length within the region of intron area) and 43 nucleotide length at the 3' splice site (8 nucleotide length of the exon region and 35 bases length within region the intron). In this manner, a total of 114,950 splice sites were identified and extracted by Perl programming (Fig 5).

Classification of splice sites: The splice sites are classified into following types of categories. U2 variety GT splice AG, GC Splice AG and AT Splice AC and U12 variety GT Splice AG on the basis of their conserved sequences and terminal dinucleotides by employing a perl script.

U2 type splice sites:U2-dependent splice sites have loosely conserved sequences at the branch point and highly conserved at terminal regions of dinucleotides at 5' end of splice site and 3' end of splice site.

A total of 93761 number of U2 type splice sites were successfully extracted out of which 92826 number of splice sites were GT-AG type, 892 splice sites were GC-AG type and 43 number of splice sites AT-AC type splice sites (Fig 4 to 5). It was found that in the small HSP proteins belongs to zebrafish, entire alpha crystalline domain at the C-terminal end was conserved. By aligning only the amino acids between the catalytic residues, it was observed that the catalytic residues GLV (Glycine, Leucine and Valine) were conserved in all the small heat shock proteins under study.

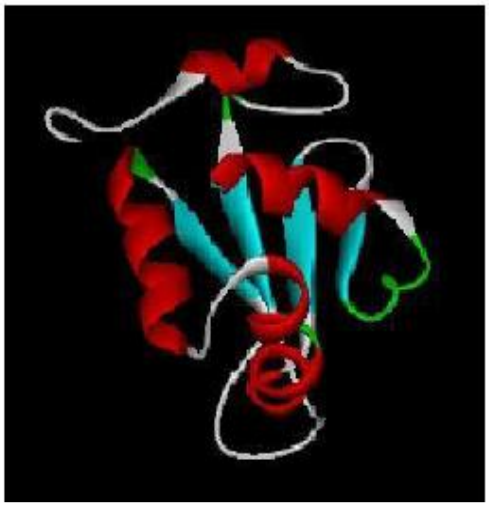


Fig1: 3D Model of Heat Shock Factor of Zebra Fish.

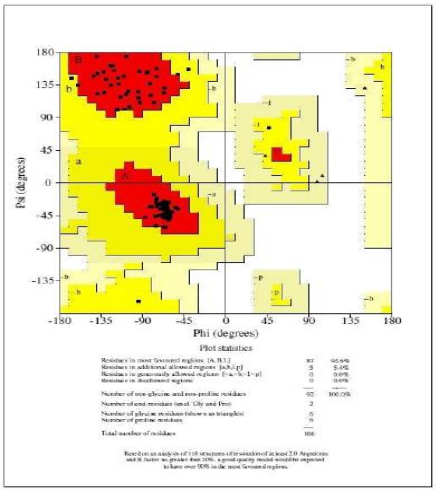


Fig 2: Ramachandran Plot of Heat Shock Factor of Zebra Fish

Table 1. Prediction of Transcription Binding Factors

SL NO	GENE NAME	GENE ID	TF NAME	TF BINDING SITES	START	END	STRAND
1	HSP, alpha-crystallin-related, 1	NM_001008615	TATA	ctaTAAAAg	-30	-28	POSITIVE
2	HSP, alpha-crystallin-related, b3	NM_001098452	TATA	actcTAAAAa	-70	-66	NEGATIVE
3	Heat shock protein, crystallins, alpha B, a(hspb5)	NM_131157	TATA	gcaaaAAACActtta	-97	-87	NEGATIVE
4	Heat Shock Protein family 40, DNA J homologue, Sub-Family B, member 11	NM_199021	TATA STATx	cTTTTAtaac TTACCagaa	-66 -536	-76 -526	NEGATIVE
5	HSP, alpha-crystallin-related, b11	NM_001098427	TATA CCAAT box MyoD	tTTTTAaata agcggCCACTct tccaCATGaaac	-45 -190 -351	-35 -179 -339	POSITIVE

Table 2. Prediction of Heat Shock Elements of the Consensus Motif nGAAnnTTCn Type

SL NO	GENE NAME	GENE ID	START POSITION (from 1K upstream region)	END POSITION (from 1K upstream region)
1	Heat shock protein crystallin crystallin, alpha B, b(5b)	NM_001002670	222	230
2	HSP, alpha-crystallin-related, b11	NM_001098427	660	668
3	Heat Shock Protein family 40, DNA J homologue, member 12	NM_212659	565	575
4	Heat Shock Protein family 40, DNA J homologue, member 1a	NM_001003571	155	163
5	HSP Protein 90, α (cytosolic), class variant B member 1.	NM 131310	169	177

Table 3. Prediction of Heat Shock Elements of the Consensus Motif nGAAnnnnnnTTCn Type

SL NO	GENE NAME	GENE ID	START POSITION (from 1K upstream region)	END POSITION (from 1K upstream region)
1	HSP, alpha-crystallin-related, 1	NM_001008615	869	872
2	HSP, crystallin, alpha A	NM_152950	250	263
3	HSP, crystallin, alpha B, a(hspb5)	NM_131157	850	863

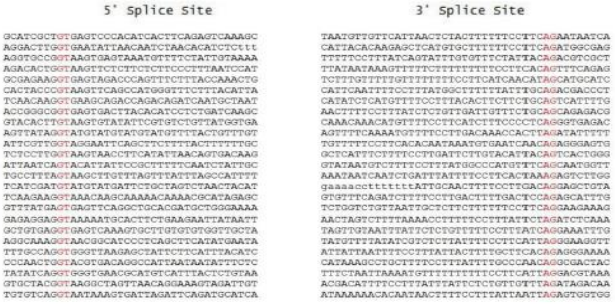


Fig 3. U12 GT-AG-BPS 10 Type Splice Site Motif

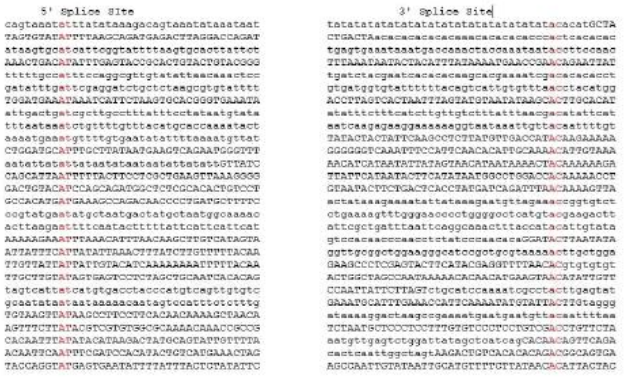
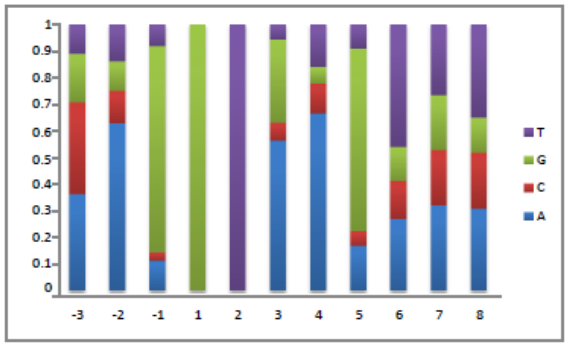


Fig 4: U2 : AT-GC Splice Site Motif

FIG 5: SPLICE SITE DIPICTION AT CHROMOSOME NUMBER 7



Graphical Representation of Position Weight Matrix for 5' U2 GT-AG Splice Site

V. DISCUSSION

Prediction of Heat Shock Transcription Factor Binding Sites: Stress response is a multifarious biological process that incorporates the up regulation and way down regulation of several genes. Heat shock proteins (HSPs) are also generally represented as stress proteins are primarily involved in stress response (Hashimoto *et al.*, 2004). In addition, these genes also represent the activities involved in normal development and growth. This complicates the process of understanding stress responsive genes. The HSEs of two types were obtained by pattern matching, the consensus pattern being, **n**GAA**nn**TTC**n** and **n**GAA**nnnnnn**TTC**n,** where '**n**' can be any nucleotide. Eleven genes were found to possess the first type HSE while 13 genes were found to have the second type HSE. These HSEs were found to be at different positions in the 1 Kb upstream regions and therefore need to be experimentally validated. General transcription factors like the TATA box and CCAAT box were predicted for 10 genes on both the strands using the transcription factor vertebrate. matrix of TransFac database in HSP40 and small HSPs. However, these computationally predicted results need to be validated experimentally.

Heat shock transcription factor: HSP response in eukaryotes is transcriptionally controlled by conserved HSP transcription factor 1 (Hsf 1), Wang *et al.*, 2007. A genome-wide identification of HSP transcription factor family was undertaken for zebrafish studied by Lin *et al.*, 2011, as the complete genome of zebrafish is available. A local BLASTp search was performed with the residues Hsf-type DNA Binding domain of PF00447 for searching possible homologues encoded amini acids in the genome. Twenty six hits were generated, which were already described in the NCBI database as heat shock transcription factors. Hence no new homologues of HSP transcription factor family could be identified. This indicates that the present level of protein database information in the zebrafish genome is not sufficient to identify new homologues hsp transcription factor family. However, tertiary structure of zebrafish heat shock factor was not available in the public database due to which the 3 Dimensional X-ray based structure of the DBD was predicted in study by means of homology based modeling. The DBD was attached to the hsp element represent of upstream of Hsp genes and regulates its activity. Further studies on the 3D structure of hsp transcription factor across the species would help in the better understanding of its function.

Functional Elements Information System

Basic information on the functional elements is an important requirement to carry out gene characterization studies at the genome level in zebrafish. During present study this basic information was found to be limited. Though information on exon sequences and mRNA of Zebrafish can be obtained by accessing public databases like the NCBI and the UCSC Genome browser, obtaining user specific information like exon sequences and splice site motifs for all the genes on a particular chromosome is not easy. A database called UTRdb (http://www.ba.itb.cnr.it/UTR/) provides information for 21,276 zebrafish UTRs if the gene in question is known but the option of accessing the entire UTR collection for studying patterns in them is not available. Further, as bulk downloads are not provided it becomes cumbersome to download UTR sequences for multiple genes. In the present study 56,861 UTRs were extracted in comparison to the UTRdb database and in addition, the contents could be accessed by bulk downloads. Splice sites information for zebrafish is not available on any public database and the creation of zebrafish splice site compilations during the present study is the first of its kind. The findings in the present study concur with the results reported by earlier workers. The GC-AG type splice sites shows 0.77% similar with U2 type of introns. A previous study by Sheth *et al.* (2006) reported 0.82% in humans, 0.86% in mouse, 0.80% in ear cress plant (*Arabidopsis*), 0.44% in drosophila and 0.36% in the nematode, roundworm representing that more basic line of organisms tend to have fewer GC/AG splice sites which is supported by our current findings. The consensus sequence (AG|GTAAGT) for the U2 varient of GT-AG 5' splice site motif corresponds to an ideal base pairing with the 5' terminal end of U1 snRNA.

In spite of this, a high degree of unpredictability seems to be tolerated in majority of positions with the actual 5' splice sites, indicating a tolerance for mismatches in U1 base pairing.

In U2 GC/AG type, the replacement at position +2 of the 5' end splice site introduces a mis alignment in the U1: 5' end splice site helix. The PWMs shows that the 5' splice site sequence has a high degree of match to the consensus GT/AG 5' end splice sites, thereby fulfilment for the mismatch at position at +2. Non canonical structures of U2 variant splice sites with AT/AC dinucleotide terminal intron ends have been known to occur and are active. A total of 43 AT/AC U2 splice sites were observed in Zebrafish which is the highest number recorded among all the model species till now. This high number could also partly be attributed to the sequencing error as 50% ends of the AT-AC type were found to have low quality reads at the terminal dinucleotides.

VI. CONCLUSION

The zebrafish (*Danio rerio*) has a long history of usage as an experimental animal model in biomedical research. These works predominantly genetic and developmental studies, but zebrafish is increasingly being used as a model in stress research. All living cells display a rapid molecular response to adverse environmental conditions, a phenomenon broadly termed as the heat shock (HS) response. Characterization of candidate stress genes is a critical step in obtaining insight

into adaptation and acclimation of organisms. Potential transcription factor binding sites were predicted by scanning the 1 Kb upstream regions using pattern matching and position weight matrix methods. Heat shock elements were identified in 24 genes which were grouped into two categories based on their consensus motifs. Eleven genes had the HSE of the consensus motif **n**GAA**nn**TTC**n** type and 13 genes had the Consensus Motif **n**GAA**nnnnnn**TTC**n** type. Other general transcription factor binding sites like TATA box and CCAAT box were predicted for 10 genes. The DNA-binding domain of heat shock transcription factor 1 binds to the heat shock element upstream of transcription start site and regulates the expression of heat shock protein. The zebrafish heat shock transcription factor 1 exhibits a high sequence similarity with the human heat shock transcription factor 1. Hence, a 3D structure of the DNA-binding domain of heat shock transcription factor 1 was predicted by homology modelling to gain a deeper insight into the mechanism of heat shock regulation at the protein level. Zebrafish information system was created by extracting mRNA sequences, exons sequences, 5' and 3' UTR sequences and splice sites. The information system consists of a total number of 136,252 exon sequences, 56,261 mRNA sequences, 56,261 UTRs and 114,950 splice sites. All the extracted functional elements conform to the FASTA sequence formats which are user friendly. This work entails the highest collection of the UTR sequences available in the public database till now. Splice sites extracted during the study were successfully classified into U2 and U12 type adding additional information to the database on public domain. The creation of zebrafish splice site compilations during the present study is the first of its kind. As part of further studies, the zebrafish information system may be useful to identify patterns in UTRs and analyze them. The splice site information and distribution of U2 and U12 type splice sites may be used to study the origin and evolution of splice sites by comparative splice site analysis of other fishes and organisms in future.

REFERENCES

1. Basu, N., Todgham, A. E., Ackerman, P. A., Bibeau, M. R., Nakano, K., Schulte, P. M. and Iwama, G. K., 2002. Heat shock protein genes and their functional significance in fish. *Gene.*, 295: 173-183.
2. Bienz, M. and Pelham, H. R., 1987. Mechanisms of heat shock gene activation in higher eukaryotes. *Adv. Genet*., 24: 31–72.
3. Enokia, Y. and Sakurai, H., 2011. Diversity in DNA recognition by heat shock transcription factors (HSFs) from model organisms. *FEBS lett.*, 585(9): 1293-1298.
4. Feder, M. E. and Hofmann, G. E., 1999. Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. *Annu. Rev. Physiol*., 61: 243–82.
5. Goodson, M. L., Parl-Sarge, K., and Sarge, K. D., 1995. Tissue dependent expression of heat shock factor 2 isoforms with distinct transcriptional activites. *Mol. Cell. Biol.*, 15: 5288-5293.
6. Hashimoto, K., Shibuno, T., Murayama-Kayano, E., Tanaka, H. and Kayano, T., 2004. Isolation and characterization of stress-responsive genes from the scleractinian coral *Pocillopora damicornis*. *Coral Reefs*., 23: 485–491.
7. Heiss, M. A., and Duncan, R. F., 1996. Sequence and structure determinants of Drosophila Hsp 70 mRNA translation: 5'-UTR secondary structurespecifically inhibits heat shock protein mRNA translation. *Nucl. Acid. Res.*, 24: 2441-2449.
8. Lindquist, S and Craig, E. A., 1988. The heat-shock proteins. *Annu.*
9. *Rev. Genet*., 22:1-77.
10. Lindquist, S., 1986. The heat-shock response. *Annu. Rev. Biochem.*, 55: 1151–91Lin, C., Mount, S. M., Jarmolowski, A and Makalowski, W. 2010. Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol Biol*., 10: 47-57.
11. Miramontes, P., 1992. A cellular automation model for the evaluation of nucleic acids (A cellular automation model for the evolution of nucleic acids). Ph D thesis. UNAM.
12. Morimoto, R. I., Sarge, K. D. and Abravaya, K., 1992. Transcriptional regulation of heat shock genes. *J. Biol. Chem.*, 267: 21987–21990.
13. Pan, F., Zarate, J. M., Tremblay, G. C. and Bradley, T. M., 2000. Cloning and characterization of salmon hsp90 cDNA: Upregulation by thermal and hyperosmotic stress. *J. Exp. Zool.*, 287: 199–212.
14. Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R. and Sachidanandam, R., 2006. Comprehensive splice-site analysis using comparative genomics. *Nucl. Acid. Res.*, 34(14): 3955–3967.
15. Sistonen, L., Sarge, K. D. and Morimoto, R. I., 1994. Human heat shock factors 1 and 2 are differentially activated and can synergistically induce hsp70 gene transcription. *Mol. Cell. Biol.*, 14: 2087-2099.
16. Thakurta, G. D., Palomar, L., Stormo, G. D., Tedesco, P., Johnson, T. E. and Walker, D. W., Lithgow, G., Kim, S. and Link, C. D., 2002. Identification of a novel cis-regulatory element involved in the heat shock response in *Caenorhabditis elegans* using microarray gene expression and computational methods. *Genome Res.*, 12: 701-712.
17. Wang, J., Wei, Y., Li, X., Cao, H., Xu, M. and Dai, J., 2007. The identification of heat shock protein genes in goldfish (*Carassius auratus*) and their expression in a complex environment in Gaobeidian Lake, Beijing, China. *Comp. Biochem. Physiol.*, 145: 350-362.
18. Westerheide, S. D., and Morimoto, R. I., 2005. Heat shock response modulators as therapeutic tools for diseases of protein conformation. *J. Biol. Chem.*, 280: 33097–33100.

AUTHORS PROFILE

Dr. J.Venkateshwara Rao, M.Sc. B.Ed, M.Tech, Ph.D, Post Doc (USA), BOYSCAST Fellow, is an Asst.Professor in Dept of Zoology, Osmania University, joined in Osmania University in 2007. In 2009, he got Young Scientist and Investigator award from DBT under DBT-RGYI Scheme and in 2010 he got most prestigious DST BOYSCAST fellowship to pursue high end research in advanced areas. He went to Virginia Common Wealth University, Richmond and conducted research in Signal transduction cascades in smooth muscle cells of Gastrointestinal Muscles.