Pak. J. Bot., 47(3): 1075-1086, 2015.

IN SILICO ANALYSIS, MAPPING OF REGULATORY ELEMENTS AND CORRESPONDING PROTEIN-DNA INTERACTION IN *atpβ* GENE PROMOTER FROM DIFFERENT TOMATO VARIETIES

MAHAM HASSAN[1], ZABTA KHAN SHINWARI[2] AND TARIQ MAHMOOD[1]*

[1]*Department of Plant Sciences, Quaid-i-Azam University, Islamabad-45320, Pakistan*
[2]*Department of Biotechnology, Quaid-i-Azam University, Islamabad-45320, Pakistan*
**Corresponding author: tmahmood@qau.edu.pk; tmahmood.qau@gmail.com*

Abstract

Chloroplast *atp*B gene encodes the β-subunit of ATP synthase having a function in the synthesis of ATP. ATP synthase is usually present in the mitochondrial and chloroplast membranes as well as in prokaryotes with a highly conserved structure. With the aim to study the functional diversity, *atp*B gene promoter from *Solanum lycopersicum* varieties (VCT-1, 127, PETO-86, LBR-16, Continental, LBR-06, CLN-2498D, CLN-2777, LBR-09) was amplified, sequenced and analyzed. BLAST tool analyzed the sequences. *Cis* acting elements were predicted using Consite tool in the promoter region as these elements have importance in gene regulation and maps of the *cis*-acting elements were constructed with the help of DOG software for each variety. It was seen from the constructed maps that the distribution pattern of *trans*-acting elements was quite diverse in the *atp*B promoter of studied tomato varieties. Various *trans*-acting elements (HMG-IY, HFH-2, TBP, c-Fos and SOX17) were docked with respective DNA sequence using HADDOCK online software. Main focus was on the analysis of DNA-protein interactions and for this purpose; hydrogen bonds formed with either phosphate backbone or base of DNA were checked. Particularly interactions between amino acid and base pairing were checked. It was revealed that Arginine and Lysine had a greater probability to interact with Thymine and Adenine as compared to other bases and all bonds formed were feasible as their distances were less than 3.5 A^{o}. During the analysis of interactions, it was also found that Lysine, Arginine, Asparagine and Serine are well capable to bind with Thymine while Glycine and Lysine have made bonds mostly with Adenine.

Key word: Tomato, *atp*B promoter, Transcriptional factors, DNA-protein interactions.

Introduction

Tomato *(Solanum lycopersicum* L.) important member of family *Solanaceae,* is a fruit grown almost everywhere in the world. It is the second most consumed vegetable (USDA, 2013) having many important features regarding nutrition, economics and is a good model for examining the genetic studies of plants (Kalloo, 1991). In tomato, genetic research was made possible on the basis of available resources like marker collection, bacterial artificial chromosomes (BAC), tomato species (wild and mutant), permanent recombinant inbred (RI) mapping population, F2 synteny mapping population, virus induced gene silencing libraries (VIGS) and microarrays (Barone *et al.,* 2007).Tomato genome was sequenced in 2003 by the International Solanaceae Genomics Project (SOL) (Mueller *et al.,* 2005a).

Chloroplasts are semiautonomous, self-replicating organelle with its nuclear and mitochondrial genome which plays important functions (Finkemeier and Leister, 2010). Chloroplasts are highly conserved in nature, so its gene sequences are used to study evolution, plant phylogenies (Shinwari *et al*., 1994, 1994a) and systematics (Shinwari *et al*., 1994b). ATP synthase enzyme is responsible for the synthesis of ATP during photosynthesis process and in respiration. This enzyme is encoded by nine subunits, one is *atp*B subunit encoded in chloroplasts. Chloroplast *atp*B gene is mostly used for phylogenetic and comparative studies due to its conserved evolutionary rate (Savolainen *et al.,* 2000; Jamil *et al*., 2014).

Transcription process plays a vital role as it regulates expression of genes at specific developmental stages in particular tissues which is mainly controlled by a range of transcriptional factors (Latchman, 1997). Transcription regulation has great biological importance as any modifications/alterations during the action of transcription process could results in diseases (Maston *et al*., 2006). The most vital step during transcription initiation is the identification by a particular transcriptional factor and the bonding between amino acids and nucleotides of DNA site (Roeder, 1996; Maston *et al*., 2006).

When transcriptional factors bond with DNA various interactions occur like van der Waals, ionic bonds and hydrogen bonds (Angarica *et al*., 2008). As DNA-protein interactions have vital roles so to understand the binding specificity process within the interactions is also important. In plants transcription regulation has a very important role in the development and creating response against unfavourable environmental conditions. Tissue specific behaviour, growth and development processes (Grasser, 2007) particularly during seed maturation period, seed storage proteins expression, acquiring tolerance towards desiccation and entrance into dormant stage were found to be dependent upon transcription regulation (Vicente-Carbajosa & Carbonero, 2005; Narusaka *et al*., 2003).

In the present study, *cis*-acting elements in the *atp*B promoter region of the studied tomato varieties would be found and only the uncommon ones would be mapped for each variety separately. Furthermore, interactions would be observed between selected transcriptional factors and their regulatory elements. The main focus would be on hydrogen bonds as two third of interactions are hydrogen bonds (Angarica *et al.,* 2008) and they provide stability and specificity to protein-DNA complexes.

Material and Methods

Seeds of different tomato varieties, VCT-1, 127, PETO-86, LBR-16, Continental, LBR-06, CLN-2498D, CLN-2777 and LBR-09 were collected from National Agriculture Research Centre, Pakistan and were grown in pots. After few weeks fresh leaves were collected for the extraction of DNA by CTAB method (Richards, 1997).

Primer designing: Primer 3 (version 4.0) (http://primer3.sourseforage.net/) was used for the designing of primer pair via using gene of *atp*B promoter from *Nicotianatabacum* chloroplast DNA sequence obtainable from NCBI Genbank (www.ncbi.nih.gov). The size of primer was ~1000 bp. The primers made had following sequence in 5'-3' direction; *atp*BP Forward: 5' CCAGAAGTAGTAGGATTGATTCTCA 3' and *atp*BP Reverse 5' TCTTCAGGTGGAACTCCAGGTT 3'.

Amplification: PCR reaction was executed in a 25 µL reaction mixture containing 10 x PCR buffer, 2 mMdNTPs, 25 mM MgCl2, 25 pM forward and reverse primer, 30-50 ng/µL template DNA, nanopure water and finally 1.5 U *Taq* polymerase (Fermentas) by means of PCR MultiGene Thermal Cycler (Labnet). For PCR reaction its conditions were initial denaturation for 5 min at 94^{o}C, after that 35 cycles of 3 steps containing denaturation for 40 sec at 94^{o}C, annealing for 40 sec between temperature range of 49-50^{o}C (as different varieties annealed at different temperatures) and extension for 45 sec at 72^{o}C. Final extension for 20 min at 72^{o}C was the last step. PCR products were then verified by running it on 1.5% agarose gel in 0.5 x TAE buffer. Thermo Scientific GeneJET PCR Purification Kit (Thermo Scientific) was used for the purification of verified products of PCR amplification according to standard procedures.

Sequencing of amplified products: Sequences were sent to Macrogen, South Korea for sequencing. When the sequences were obtained, they were submitted to Genbank in order to get accession numbers.

Sequence analysis: BLAST tool was used for sequence analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Regulatory elements identification: ConSite a web based tool was used for the determination of regulatory elements and the binding sites of *cis*-acting elements in the promoter regionavailable at http:/www.phylofoot.org/consite (Sandelin *et al.,* 2004).

Mapping: For mapping of *cis*-acting elements DOG (Domain Graph) software was used. It is available at http://www.mybiosoftware.com/protein-sequence-analysis (Ren *et al.,* 1997). *Cis*-acting elements obtained from *atp*B promoter region were mapped for each variety only the uncommon ones were mapped.

Modelling of regulatory elements: Regulatory elements were modelled by using an online available software 3D-DART (HADDOCK) (http://haddock.science.uu.nl/ services/ 3DDART) that helps us providing DNA structure in PDB format. On the other hand, PDB (Protein Data Bank) was used to get the structural data of TFs available at http://www.rcsb.org/pdb (Berman *et al.,* 2000). PDB consists of information related to the 3D structure of large bio-molecules like proteins, nucleic acids. The best PDB's were selected and used for further analysis of DNA protein docking.

Protein-DNA docking: For the purpose of protein-DNA docking HADDOCK web server was used freely available at http://haddock.chem.uu.nl. Easy interface was performed which required structures of both biological molecules as well as their active residues (interacting residues) (De Vries *et al.,* 2010). Active residues of TFs were recognised with the help of an additional tool DISPLAR (http://pipe.scs.fsu.edu/displar.html) (Tjong & Zhou, 2009) which serves to provide DNA binding sites on the surface of proteins. Active residues and the structures of both molecules were submitted to be docked.

After docking the resultant model was assessed on the basis of HADDOCK score by the HADDOCK Web server. Energies like van der Waals, Electrostatic, De-solvation and restraint violation with the buried surface area constitute the score. Cluster size gives the quantity of good structures and as models with lowest energy are assumed to be good structures, so they are favoured (De Vries *et al.,* 2010).

Visualization of docked molecules: PyMOL is an open source molecular graphics tool accessible at http://www.pymol.org. It excels at 3D visualization of small molecules and other important biological macromolecules like proteins as well as density, trajectories and surfaces. So for the visualization of docked molecules PyMOL was used.

Results and Discussions

"nblast" (http://blast.ncbi.nlm.nih.gov/Blast.cgi) result showed 99 % resemblance with already reported complete chloroplast genome of *Solanum lycopersicum*, cultivar IPA-6 (Accession # AM087200.3).
Sequences from all studied tomato varieties were allotted Accession numbers (Table 1)

Table 1. Accession numbers of*atp*B promoter sequences sequenced from studied tomato varieties.

Tomato varieties	Accession numbers
Solanumlycopersicum var. VCT-1	KM008701
Solanumlycopersicum var. 127	KJ958555
Solanumlycopersicum var. PETO-86	KM008702
Solanumlycopersicum var. LBR-09	KM008703
Solanumlycopersicum var. LBR-16	KM008704
Solanumlycopersicum var. Continental	KM008705
Solanumlycopersicum var. LBR-06	KM008706
Solanumlycopersicum var. CLN-2498D	KM008707
Solanumlycopersicum var. CLN-2777	KM008708

Mapping of elements: *Cis*-acting elements determined through ConSite were then mapped by another software DOG. A pie chart was constructed to show the percentage of common *trans*-acting elements occurring in all the studied varieties (Fig. 1). On the contrary, the ones which

were inserted or deleted in a particular variety were made on different maps each for a respective variety (Figs. 2-10). Various elements were present in a particular variety showing its insertion in it and indicating that it may have a role in that plant. Similarly, deletion of some factors showed that it might not have a role or their presence was not required.

Distribution of *trans*-acting elements: The pie chart as well as the maps show the *trans*-acting elements which appeared in *atp*B promoter region in tomato plants used. HFH-2 showed highest percentage followed by Hunchback and HMG-IY. The *trans*-acting elements which appeared only at 1 position are CFI-USF, E74-A, SAP-1, RXR-VDR, Myc-Max, CREB, HNF-1, TEF-1, dorsal_1 and so. Whereas, those which occupied at 2 distinct positions are bZIP910, HLF, p65, NF-Y and Chop-cEBP.

Interaction between transcription factors and regulatory elements (Docking): The interactions which occur when DNA binds with protein are van der Waals, hydrogen bonds and or ionic interactions (Angarica *et al.,* 2008). DNA-protein interactions have significant role, so it becomes important to recognize them. These interactions have basic role by which one could understand the molecular mechanism of gene regulation (Nakashima *et al*., 2000; Wang *et al*., 2009; Si *et al*., 2011). Only five transcriptional factors HMG-IY, HFH-2, TBP, c-Fos and SOX17, were selected to dock as they were present in abundance or had a role in plants. The main focus is on hydrogen bonds formed between protein and DNA, as these bonds provide stability and specificity to the complexes (Coulocheri *et al.,* 2007). The results detail given by HADDOCK web server is given in Table 2.

HMG-IY: The High mobility group (HMG) are chromosomal proteins present in almost all eukaryotes (Bustin & Reeves, 1996). The basic function played by this particular factor is the promotion of activation of gene (Thanos & Maniatis, 1995; Kidokoro *et al*., 2009) and also involved in various cellular processes. These proteins bind with DNA through planar, crescent shape region known as "AT hook" which specifically binds with minor groove of short AT rich stretches of DNA (Evans *et al*., 1995; Huth *et al*., 1997). The hook of protein has a palindromic peptide sequence Pro-Arg-Gly-Arg-Pro (P-R-G-R-P) bordered by several other positively charged residues (Reeves & Nissen, 1990) which binds with B-form DNA.

Hydrogen bonds between DNA and protein occurred at 19 (Table 3). Arg 10 makes direct hydrogen bond to the base Thy at 8 and 26 position while rest of residues made bond with phosphate backbone (Fig. 11).

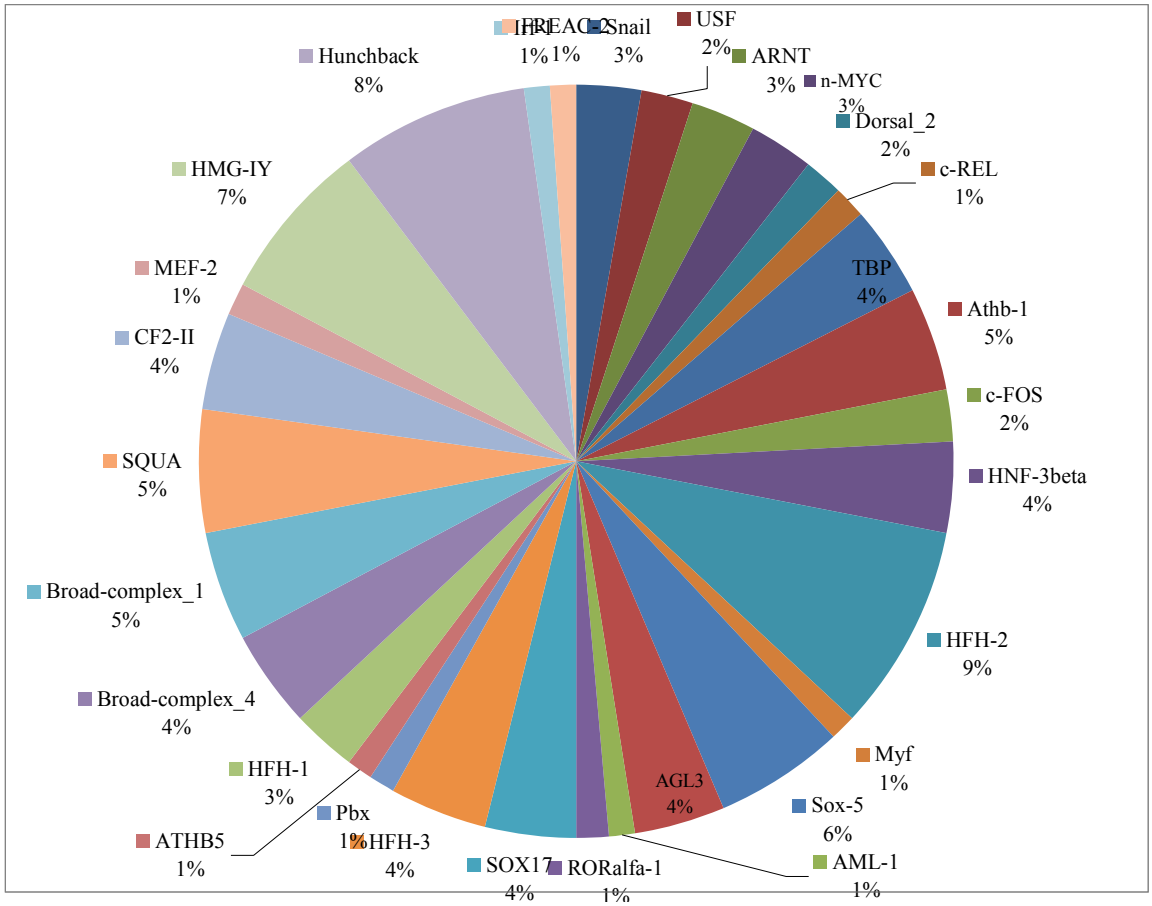

Fig. 1. Pie chart showing percentage of *trans*-acting elements which appeared in all 9 tomato varieties.

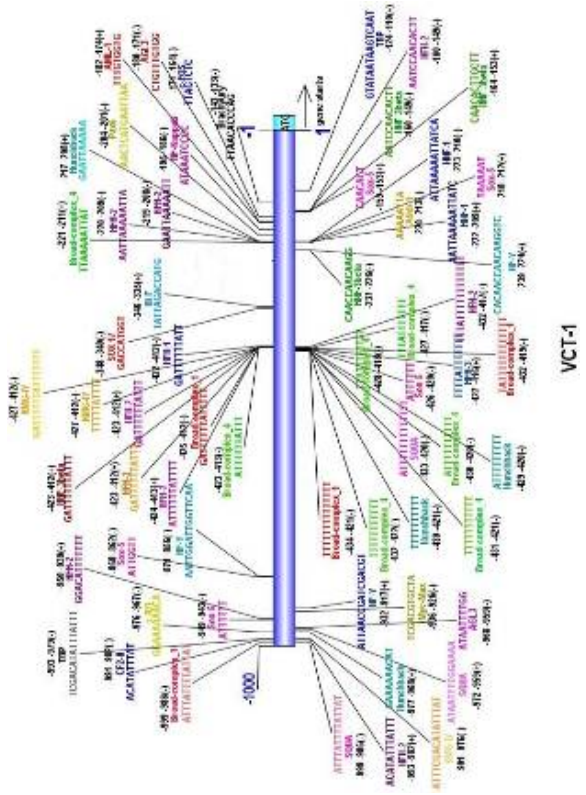

Fig. 2. Map of *cis*-acting elements exclusively present in *Solanum lycopersicum* var. VCT-1.

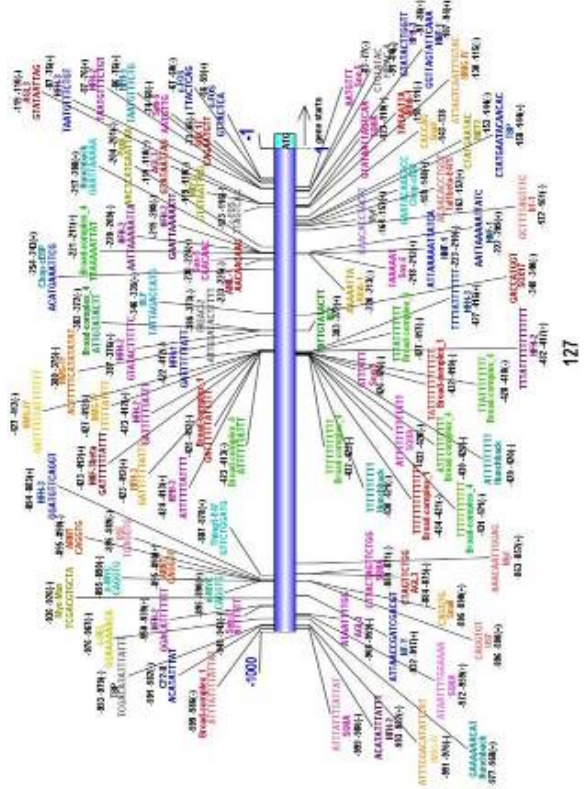

Fig. 3. Map of *cis*-acting elements exclusively present in *Solanum lycopersicum* var. 127.

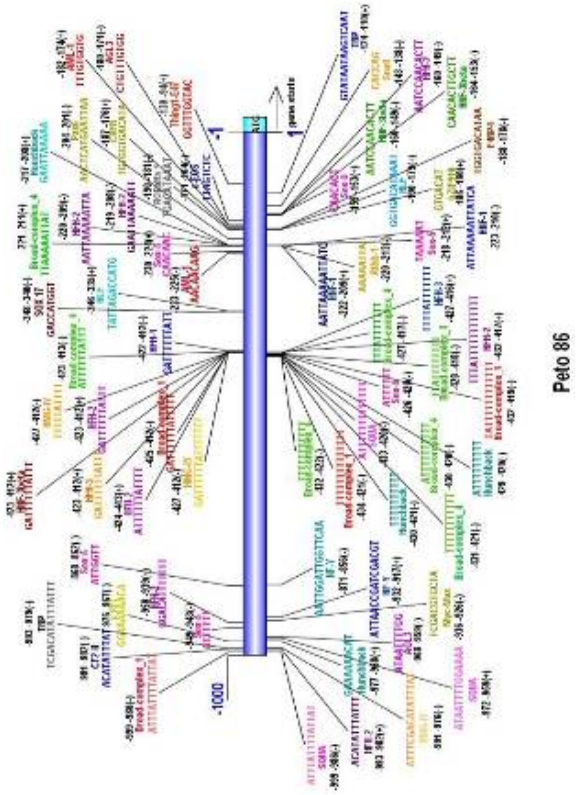

Fig. 4. Map of *cis*-acting elements exclusively present in *Solanum lycopersicum* var. PETO-86.

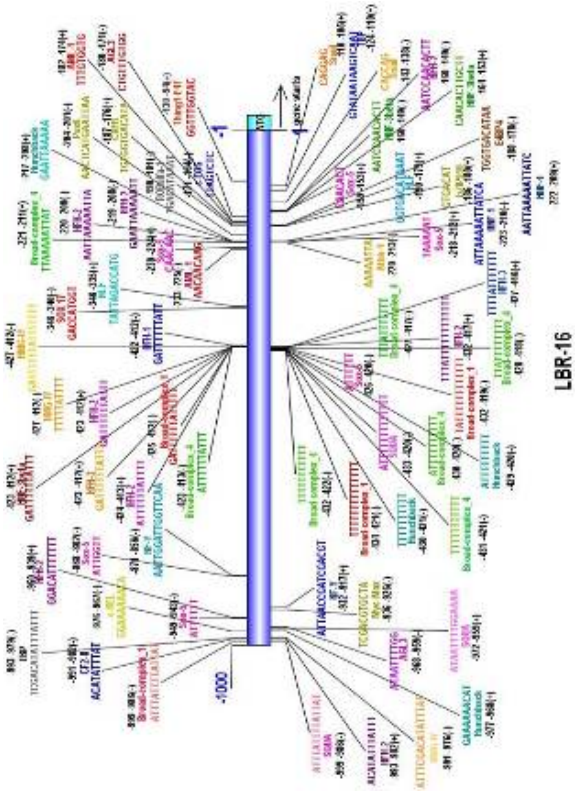

Fig. 5. Map of *cis*-acting elements exclusively present in *Solanum lycopersicum* var. LBR-16.

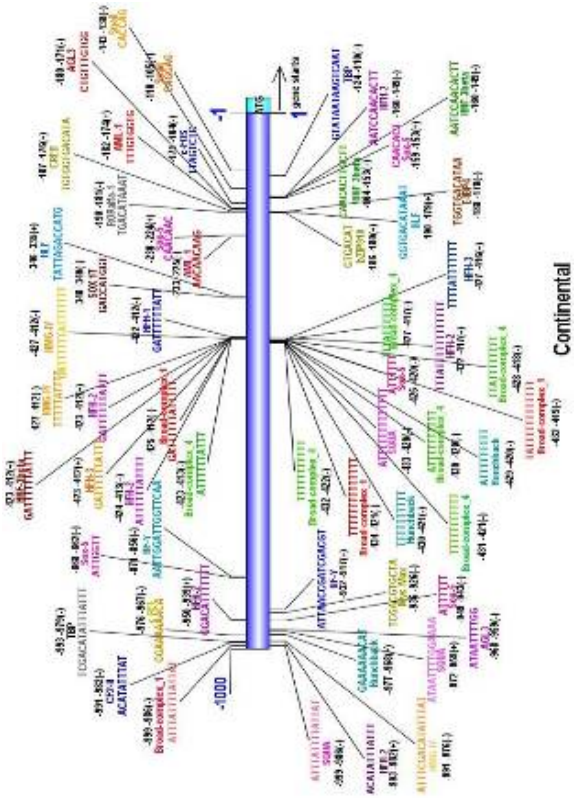

Fig. 6. Map of *cis*-acting elements exclusively present in *Solanum lycopersicum* var. Continental.

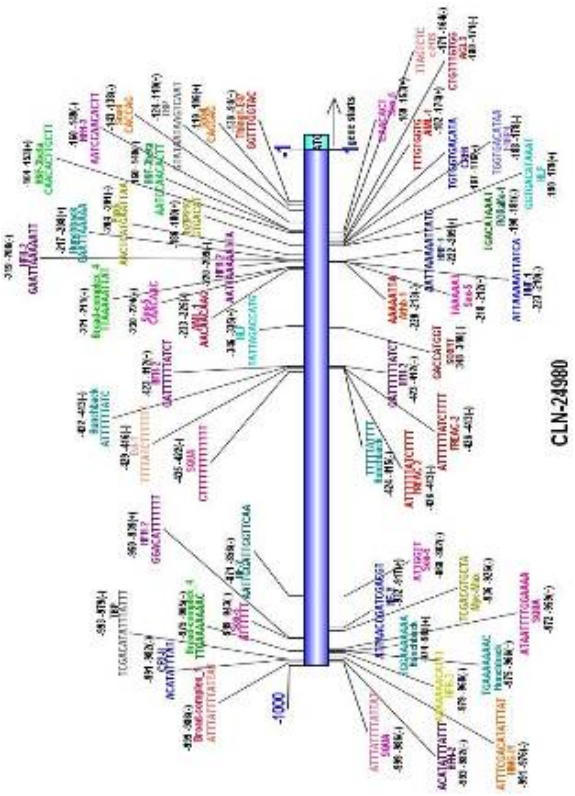

Fig. 8. Map of *cis*-acting elements exclusively present in *Solanum lycopersicum* var. CLN-2498D.

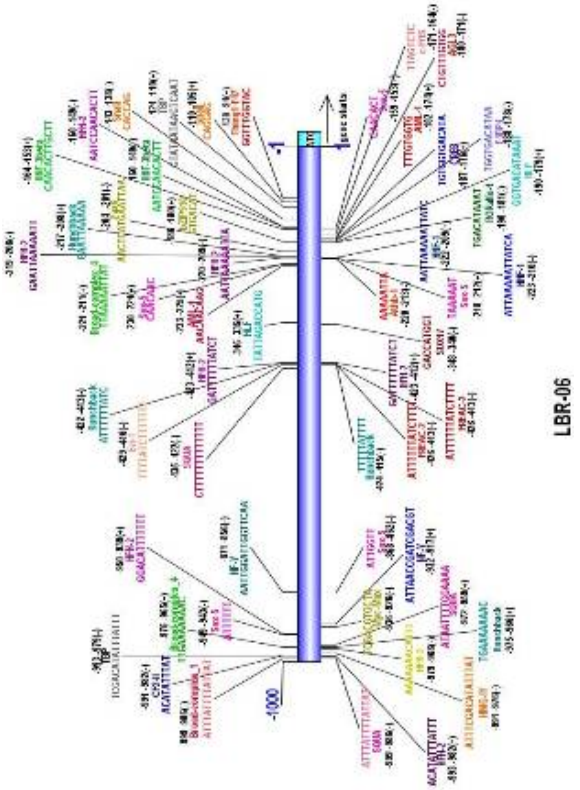

Fig. 7. Map of *cis*-acting elements exclusively present in *Solanum lycopersicum* var. LBR-06.

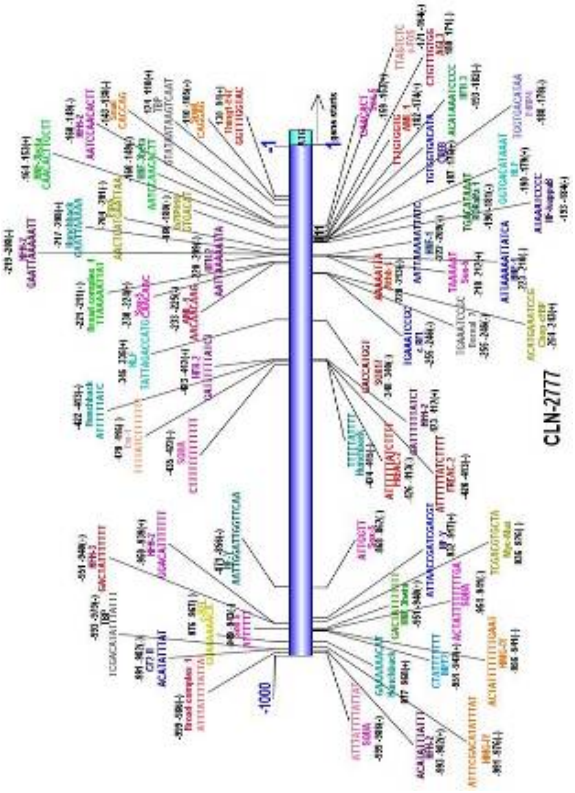

Fig. 9. Map of *cis*-acting elements exclusively present in *Solanum lycopersicum* var. CLN-2777.

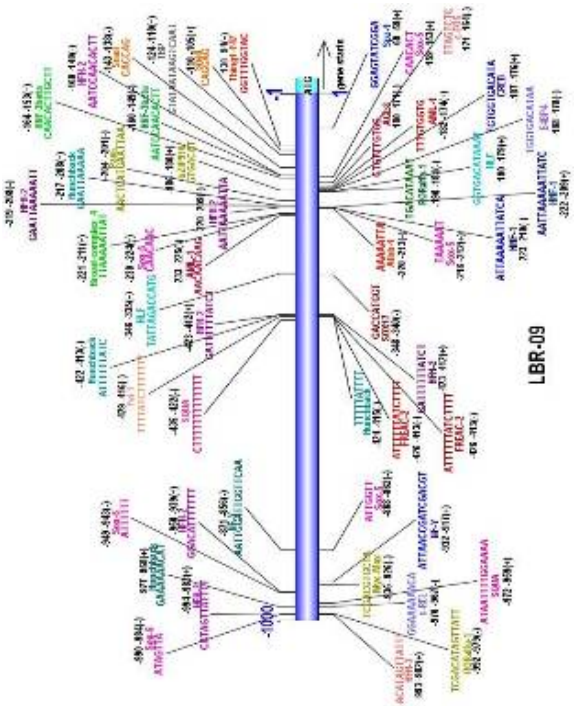

Fig. 10. Map of *cis*-acting elements exclusively present in *Solanum lycopersicum* var. LBR-09.

HFH-2: HFH proteins belong to the HTH (helix turn helix) superfamily, having vital role in immune response and development particularly in animals but their role in plants is yet to be known. HFH proteins binds to DNA with help of its two wings, W1 makes connection with major groove whereas W2 undergoes a motion indicating that it contribute toward DNA-protein interactions (Jin *et al*., 1999). The portion of DNA binding in winged helix protein is H3 recognition helix which overlies minor groove where by using its lysine side chain a single contact with DNA is made (Gajiwala & Burley, 2000).

Hydrogen bonds were formed between DNA and protein occurring at 16 positions (Table 4). Lys 3, Lys 63 and Ser 50 made direct hydrogen bond to the base Thymine at 11 and 19 and Adenine 2 position (Fig. 12) whereas the rest of amino acids interacted with phosphate backbone.

TBP: TBP (TATA box binding protein), a universal transcription factor which is mandatory for the initiation of transcription by nuclear RNA polymerase (Davidson, 2003). The concave side of the saddle shaped C-terminal binds with the DNA resulting in unwinding and bending of DNA towards major groove. The bent DNA structure was termed as TA-DNA(Guerstein & Shakked, 1996). The bending of DNA may be a significant part that help TBP to recognize specific site (Starr *et al*., 1995).

Twenty-one hydrogen bonds were formed when interaction occurred between TBP and DNA (Table 5). Direct hydrogen bond made by Arg 79, Thr 112 and Asp 165 with the bases Thy 2, 4 and Thy 17 whereas rest were hydrogen bonds made with phosphate group of DNA (Fig. 13). It can be seen from above study that Arginine and Lysine have more probability to bind with Adenine and Thymine rather than the other two bases.

Table 2. HADDOCK web server results of interaction between transcriptional factors and *cis*-acting elements.

Protein/DNA	Haddock score	Cluster size	RMSD Å	Van der Walls energy (Kcal/Mol)	Electrostatic energy (Kcal/Mol)	Desolvation energy (Kcal/Mol)	Restrains violation energy (Kcal/Mol)	Buried surface area
HMG-IY-ATTTCGACATATTTAT	-142.0 +/- 5.0	11	13.1 +/- 0.5	-85.4 +/- 3.2	-661.6 +/- 25.0	28.9 +/- 1.5	468.1 +/- 39.32	1883.3 +/- 84.9
HMG-IY-ATTTTGATTTTTTATC	-159.5 +/- 4.2	12	9.7 +/- 0.7	-85.8 +/- 3.7	-615.1 +/- 16.2	26.8 +/- 2.4	226.0 +/- 54.93	1857.7 +/- 38.2
HMG-IY-AGACAAATGAAAGACT	-165.6 +/- 6.6	17	4.6 +/- 1.5	-84.8 +/- 4.3	-646.8 +/- 21.4	28.2 +/- 1.7	203.6 +/- 46.39	1852.1 +/- 40.2
HFH-2-ATACTAATATTA	-85.7 +/- 4.5	14	2.0 +/- 1.2	-62.4 +/- 5.5	-411.7 +/- 75.0	-7.0 +/- 6.1	660.7 +/- 58.30	1684.7 +/- 52.2
HFH-2-TCTTTTTTTTTT	-85.8 +/- 11.9	16	2.5 +/- 1.5	-63.2 +/- 4.1	-557.6 +/- 29.7	18.5 +/- 4.7	704.7 +/- 113.96	1777.0 +/- 40.4
HFH-2-AAGTAAACAATT	-114.6 +/- 4.6	21	15.0 +/- 0.3	-65.0 +/- 8.6	-426.6 +/- 43.9	-7.5 +/- 2.6	432.5 +/- 38.35	1722.2 +/- 30.9
TBP-GTATAAAAAAGAAAG	-175.1 +/- 5.7	6	5.8 +/- 0.4	-71.9 +/- 5.0	-803.8 +/- 32.7	41.0 +/- 4.5	166.3 +/- 58.91	2178.1 +/- 83.3
TBP-GTATAATAAGTCAAT	-151.6 +/- 8.6	14	15.3 +/- 0.2	-57.6 +/- 6.4	-744.8 +/- 38.3	37.4 +/- 6.1	176.0 +/- 51.79	2158.8 +/- 201.9
c-Fos- TTTGTCAC	-148.7 +/- 7.9	65	3.6 +/- 3.1	-57.9 +/- 1.1	-664.3 +/- 35.2	29.8 +/- 3.4	122.5 +/- 28.71	1476.9 +/- 25.5
c-Fos ATAAGTCA	-144.5 +/- 16.1	80	2.9 +/- 2.5	-55.1 +/- 4.2	-628.6 +/- 72.2	25.1 +/- 5.8	112.1 +/- 39.79	1494.3 +/- 147.4
c-Fos TTGATTCA	-144.7 +/- 10.4	53	3.7 +/- 2.8	-54.6 +/- 4.8	-645.5 +/- 48.3	30.8 +/- 6.3	81.4 +/- 6.14	1460.7 +/- 101.7
SOX-17-GACCATGGT	-172.8 +/- 4.3	70	1.6 +/- 1.0	-90.6 +/- 7.1	-565.7 +/- 31.8	11.5 +/- 1.4	194.6 +/- 38.63	2225.1 +/- 20.9
SOX-17-GAAAATATA	-182.6 +/- 3.5	77	1.0 +/- 0.6	-93.3 +/- 6.1	-606.5 +/- 34.9	11.6 +/- 5.7	204.4 +/- 76.67	2296.5 +/- 104.1
SOX-17-AACAATTGG	-144.0 +/- 5.2	89	6.8 +/- 0.5	-84.3 +/- 2.6	-598.5 +/- 44.4	8.0 +/- 4.9	520.4 +/- 97.85	2194.0 +/- 71.8

Table 3. Possible hydrogen bonds between HMG-IY and DNA sequence ATTTTGATTTTTTATC.

Protein residue	Protein atoms	Amino acid position	DNA residue	DNA atoms	Nucleotide position	Distance in A°
LYS	HZ3	7	T	O1P	26	2.2
LYS	HZ3	7	T	O2P	27	2.1
ARG	HN11	8	A	O2P	25	2.9
ARG	HN11	8	C	O1P	24	2.1
ARG	HE	8	T	O2P	26	2.1
ARG	HH21	10	T	O2	8	2.4
ARG	HE	10	T	O2	26	2.1
ASN	HD22	18	A	O1P	14	2.3
LYS	HN	19	T	O1P	23	2.1
LYS	HZ3	19	C	O1P	24	2.1
LYS	HZ2	17	G	O1P	13	2.1
GLY	HN	20	T	O1P	22	2.1
GLY	HN	20	T	O1P	23	2.2
ARG	HH21	12	G	O4	13	2.5
ARG	HN	12	T	O1P	23	2.1
SER	HG	16	C	O3	24	2.6
GLY	HN	15	A	O1P	12	2.1
GLY	HN	11	A	O4	25	2.4
GLY	HN	11	A	O4	24	2.4

Table 4. Possible hydrogen bonds between HFH-2 and DNA sequence AAGTAAACAATT.

Protein residue	Protein atoms	Amino acid position	DNA residue	DNA atoms	Nucleotide position	Distance in A°
LYS	HZ2	3	T	O5	11	2.3
LYS	HZ3	3	T	O1P	12	2.0
LYS	HZ1	43	A	O1P	9	2.0
LYS	HZ1	43	A	O5'	10	2.3
LYS	HZ2	63	A	O1P	1	2.0
LYS	HZ1	63	A	H5"	2	2.8
ARG	HN	52	A	O1P	2	1.9
HIS	HE2	53	G	O6	3	2.3
SER	HN	7	G	O1P	17	2.4
SER	HN	7	T	O2P	18	1.9
TYR	HH	40	T	O1P	18	2.1
ASN	HD2	49	T	O2P	19	2.1
TRP	O	47	T	O2P	18	3.6
TYR	HH	8	T	O2P	18	2.0
SER	HE2	50	T	O6	19	2.3
SER	HG	50	T	O2P	18	2.1

c-Fos: c-Fos belongs to the bZIP (basic leucine zipper) proteins family mainly involved in cellular events. It comprises a coiled coil dimerization interface formed by leucine zipper and α helices formed by basic regions which actually contacts the bases present in DNA recognition site in major groove (Konig & Richmond, 1994; Glover & Harrison, 1995).

The possible hydrogen bonds formed when c-Fos binds with DNA are 17 (Table 6). Arg 140, Arg 147, Asn 147 and Arg 143 made hydrogen bond with the base at Thy 1, Gua 4, Thy 5 and Gua 11 (Fig. 14). In this interaction, the highest number of hydrogen bonds were formed between Arginine and Thymine also with guanine but to a lesser extent. This indicates that a particular protein interacts differently with the DNA.

SOX17: SOX (SRY-related HMG box) transcriptional factors have important role in various developmental processes. SOX17 proteins have HMG domain of ~80 amino acids, highly conserved (Kamachi *et al*., 2000) which helps in binding to the minor groove of DNA and causes a topological deformation from the standard B-DNA form (Werner *et al*., 1995, Hassan *et al*., 1998). As the binding occurs, the minor groove is widened, whereas the width of major groove decreases at interaction site core.

Fourteen possible hydrogen bonds were observed in this complex (Table 7). Arg 140, Arg 83, Lys 104, Arg 70 and Met 76 made hydrogen bonds with the bases Gua 15, Cyt 18, Ade 2, Ade 5, Gua 16, Cyt 18 and Cyt 4 respectively (Fig. 15). In this complex, Arginine and Lysine have made bonds mostly with guanine while this combination is not limited as Arginine and Lysine can also form bonds with other bases (adenine and thymine).

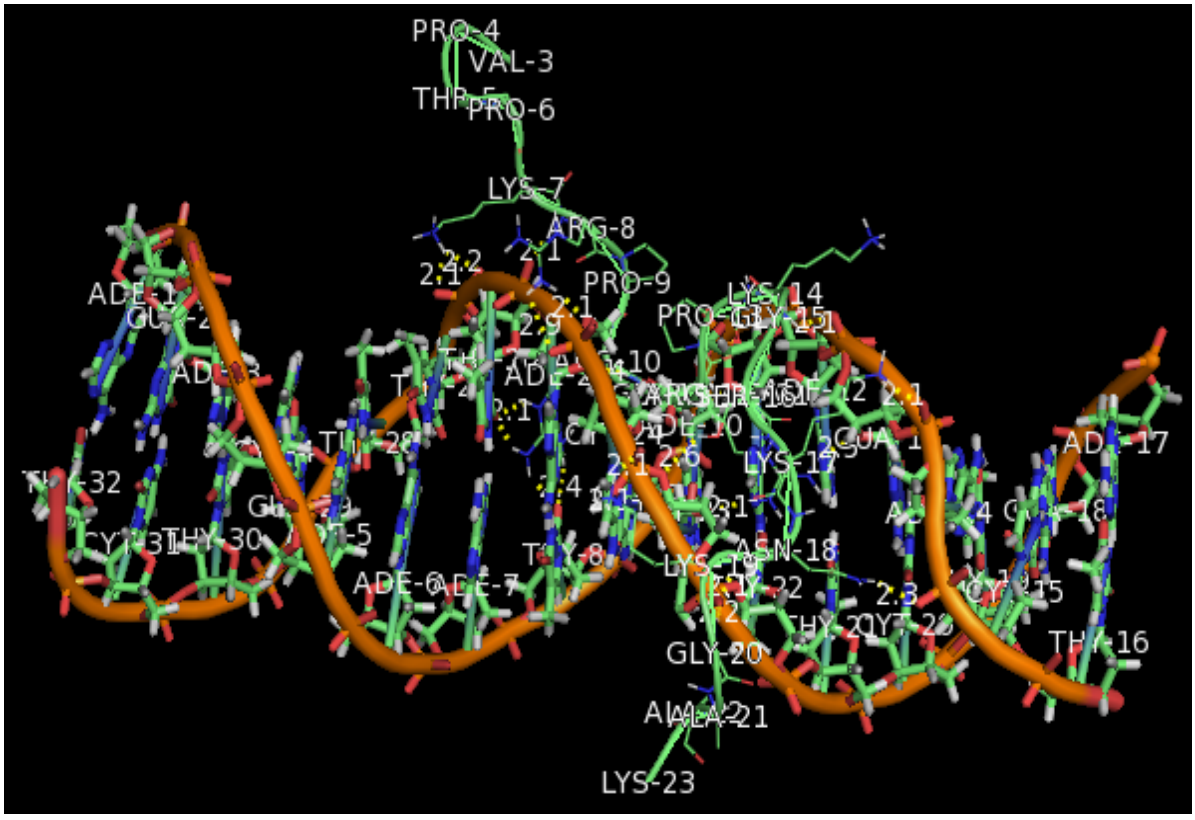

Fig. 11. Interaction between HMG-IY protein with DNA sequence ATTTTGATTTTTTATC.

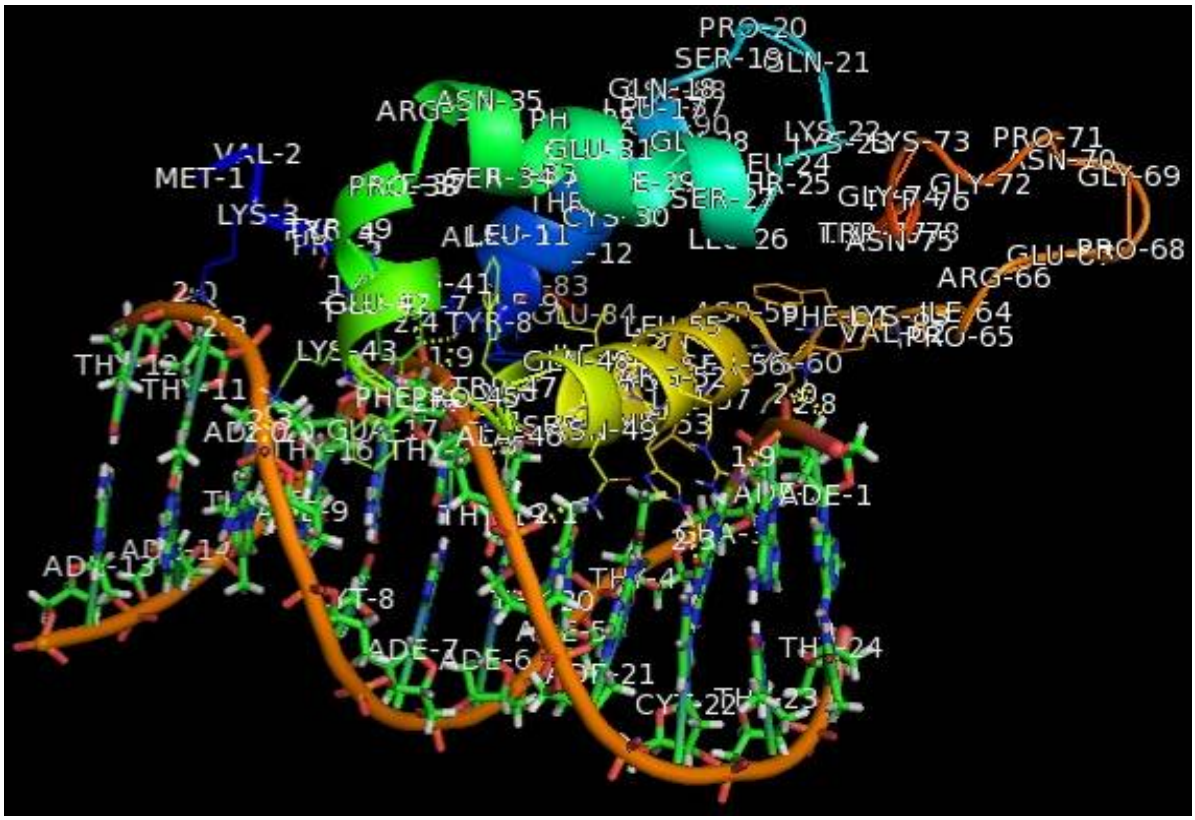

Fig. 12. Interaction between HFH-2 protein with DNA sequence AAGTAAACAATT.

Table 5. Possible hydrogen bonds between TBP and DNA sequence GTATAAAAAAGAAAG.

Protein residue	Protein atoms	Amino acid position	DNA residue	DNA atoms	Nucleotide position	Distance in A°
ARG	HH12	79	A	O1P	3	2.0
ARG	HH22	79	T	O5'	4	2.3
SER	HG	118	T	O2P	4	2.0
LYS	HZ3	120	A	O2'P	5	2.2
LYS	HZ3	120	A	O1P	6	2.3
THR	HG1	73	A	O2'P	6	1.9
LYS	HZ1	218	A	O1P	7	2.7
GLN	HE21	158	A	O1P	7	1.9
ARG	HH22	196	A	O1P	9	2.2
ASP	HZ2	165	T	O1P	16	2.1
ASP	HZ3	165	T	O5'	17	2.1
ARG	HH12	105	A	O1P	13	2.3
ARG	HH21	105	T	O1P	22	2.0
ASN	HD21	95	T	O2P	22	3.4
ASN	HD22	95	T	O1P	23	2.8
ARG	HH21	98	T	O3'	23	2.6
ARG	HH21	98	T	O1P	24	2.4
ASN	HH22	159	A	O1P	7	2.2
ASN	HH22	159	A	O2P	8	2.6
LYS	HZ3	127	T	O1P	18	2.0
THR	HG1	112	T	O2P	2	2.1

Table 6. Possible hydrogen bonds between c-Fos and DNA sequence TTTGTCAC.

Protein residue	Protein atoms	Amino acid position	DNA residue	DNA atoms	Nucleotide position	Distance in A°
ARG	HH21	140	T	O3'	1	3.4
ARG	HH21	140	T	O1P	2	2.4
ARG	HE	144	G	O1P	11	2.3
LYS	HZ1	148	T	O1P	10	2.0
ARG	HH22	155	G	O1P	9	2.2
ARG	HH22	158	C	O2P	6	2.5
ARG	HH11	158	T	O5'	5	2.7
SER	HG	154	T	O2P	5	2.5
ASN	HD21	147	G	O6	4	2.7
ASN	HD21	147	T	O4	5	2.8
ASN	HH21	147	T	O1P	2	2.5
LYS	HZ2	153	G	O2P	4	2.3
ARG	HH11	146	T	O1P	3	2.0
ARG	HH21	157	T	O1P	10	2.1
ARG	HH21	157	C	O1P	6	2.2
ARG	HH21	157	T	O2P	5	2.3
ARG	HH12	143	G	O2P	11	2.0

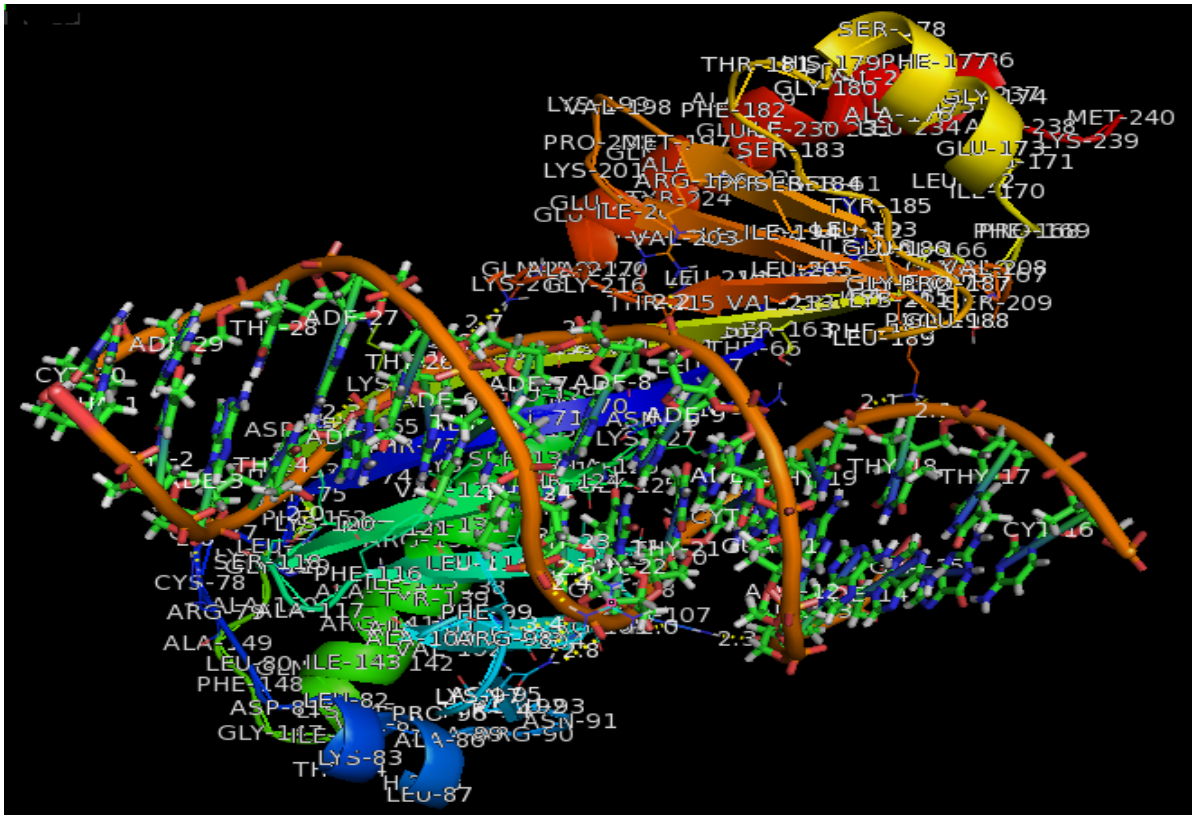

Fig. 13. Interaction between TBP protein with DNA sequence GTATAAAAAAGAAAG.

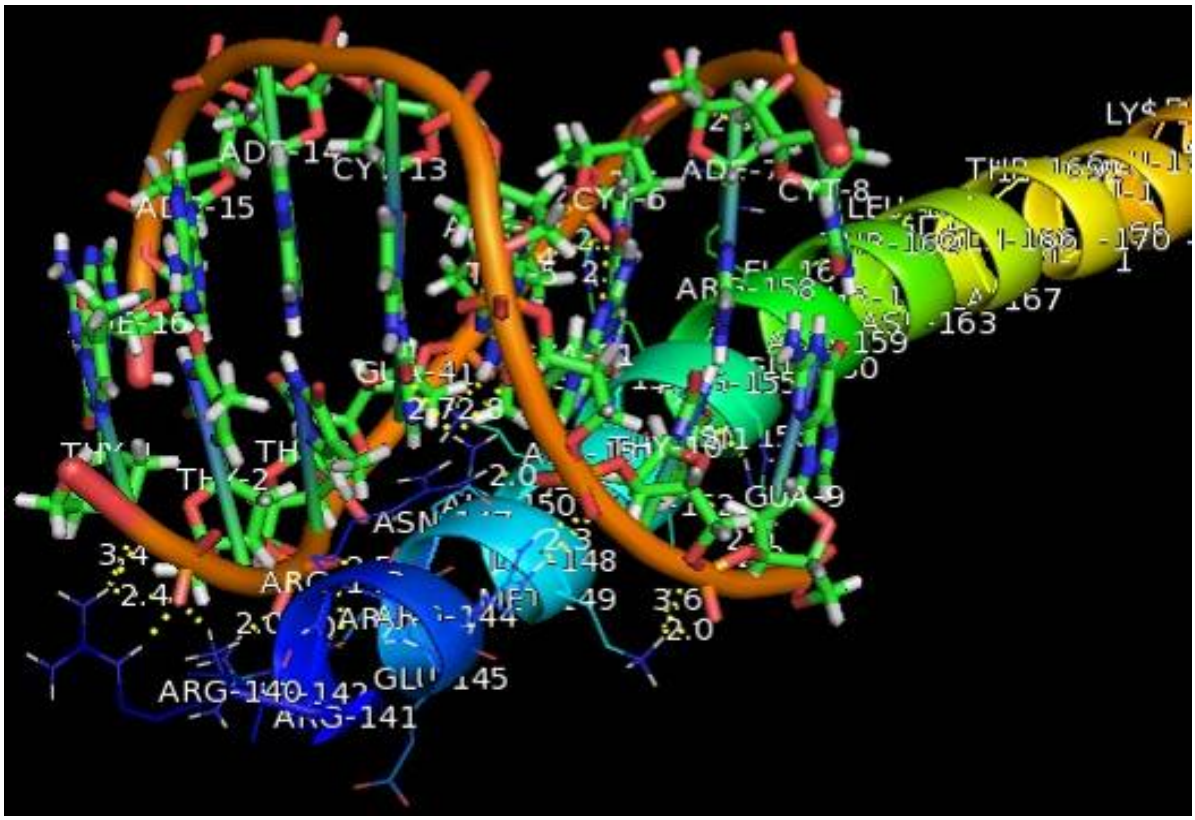

Fig. 14. Interaction between c-Fos protein with DNA sequence TTTGTCAC.

Table 7. Possible hydrogen bonds between SOX17 and DNA sequence GACCATGGT.

Protein residue	Protein atoms	Amino acid position	DNA residue	DNA atoms	Nucleotide position	Distance in A°
ARG	HH12	141	T	O1P	9	2.0
ARG	HN	140	G	O2	15	2.6
ARG	HE	140	G	O1P	16	2.5
ARG	HH21	69	T	O1P	17	2.3
ARG	HH21	69	C	O2P	18	2.1
ARG	HH22	83	C	O2	18	2.2
LYS	HZ1	104	G	O1P	1	2.3
LYS	HZ1	104	A	O5'	2	2.5
ARG	HH21	70	A	O3	5	2.2
ARG	HH11	70	G	O4'	16	2.2
ARG	HH21	70	C	O3	4	2.3
LYS	HZ1	136	T	O1P	6	2.0
MET	O	76	C	H3T	18	2.0
TRP	HE1	106	C	O1P	4	1.9

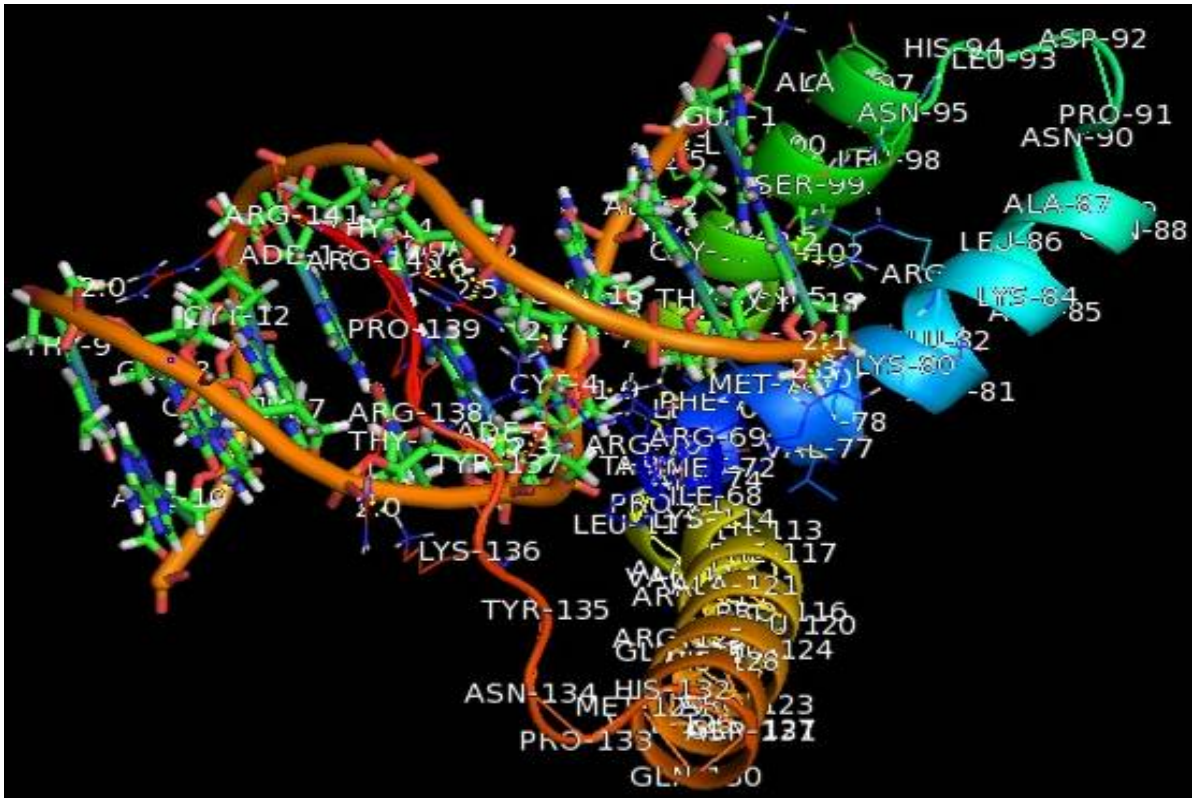

Fig. 15. Interaction between SOX17 protein with DNA sequence GACCATGGT.

Conclusion

It can be concluded from the above study that the distribution pattern of *trans*-acting elements was quite diverse in the promoter region of *atp*B gene in various tomato varieties. Few *trans*-acting elements (HMG-IY, HFH-2, TBP, c-Fos and SOX17) were docked with DNA to observe hydrogen bonds. Study of hydrogen bonds formation was done as these bonds provide specificity and stability to the DNA-protein complexes. Some were direct bonds formed with base of DNA while some were formed with phosphate backbone. Particularly interactions between amino acid and base pairing were checked. Our results revealed that Arginine and Lysine had a greater probability to interact with thymine and adenine as compared to other bases and all of the bonds formed were viable as their distances were less than 3.5 Å. During the analysis of DNA-protein interaction, it was displayed clearly that Lysine, Arginine, Asparagine and Serine were capable of binding with Thymine while Glycine and Lysine have made bonds mostly with Adenine.

References

Angarica, V.E., A.G. Pérez, A.T. Vasconcelos, J. Collado- Vides and B. Contreras-Moreira. 2008. Prediction of TF target sites based on atomistic models of protein-DNA complexes. *BMC Bioinformatics*, 1: 436.

Barone, A., M L. Chiusano, M.R. Ercolano, G. Giuliano, S. Grandillo and L. Frusciante. 2008. Structural and functional genomics of tomato. *Int. J. Plant Genom.,* 2008: 1-12.

Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.E. Bourne. 2000. The protein data bank. *Nucleic Acids Res*., 1: 235-242.

Bustin, M. and R. Reeves. 1996. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Progress in Nucleic Acid Research and Molecular Biology*, 54: 35-100.

Coulocheri, A.S., D.G. Pigis, K.A. Papavassiliou and A.G. Papavassiliou. 2007. Hydrogen bonds in protein-DNA complexes: where geometry meets plasticity. *Biochimie*, 89(11): pp. 1291-1303.

Davidson, I. 2003. The genetics of TBP and TBP-related factors. *Trends in Biochemical Sciences*, 28(7): 391-398.

De Vries, S.J., M.V. Dijk and A.M.J. JBonvin. 2010. The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols*, 5: 883-897.

Evans, J.N., J. Zajicek and M.S. Nissen. 1995. 1H and 13C NMR assignments and molecular modelling of a minor groove DNA binding peptide from the HMGI protein. *Int. J. Peptide & Protein Res.,* 45(6): 554-560.

Finkemeier, I. and D. Leister. 2010. Plant chloroplasts and other plastids. In: Encyclopedia of Life Sciences (ELS) (Chichester: John Wiley & Sons, Ltd).

Gajiwala, K.S. and S.K. Burley. 2000. Winged helix proteins. *Current Opinion in Structural Biology*, 10(1): 110-116.

Glover, J.M. and S.C. Harrison. 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA. *Nature*, 373: 257-261.

Grasser, K.D. 2007. *Annual Plant Reviews*, Regulation of Transcription in Plants. *Science*, 29: 368.

Guzikevich-Guerstein, G. and Z. Shakked. 1996. A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein. *Nature Structural and Molecular Biology,* 3(1): 32-37.

Hassan, El. M.A. and C.R. Calladine. 1998. Two distinct modes of protein-induced bending in DNA. *J. Mol. Bio*., 282(2): 331-343.

http://www.freshplaza.com/article/USDA-Tomatoes, potatoes most commonly consumed vegetables/ retrieved 2-5-2014.

Huth, J.R., C.A. Bewley, M.S. Nissen, J.N. Evans, R. Reeves, A.M. Gronenborn and G.M. Clore. 1997. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Structural Bio*., 4(8): 657-665.

Jamil, I, S. Qamarunnisa, A. Azhar, Z. K. Shinwari, S.I. Ali and M. Qaiser. 2014. Subfamilial relationships within Solanaceae as inferred from *Atpβ-rbc*L intergenic spacer. *Pak. J. Bot*., 46(2): 585-590.

Jin, C., I. Marsden, X. Chen and X. Liao. 1999. Dynamic DNA contacts observed in the NMR structure of winged helix protein-DNA complex. *J. Mol. Bio*., 289(4): 683-690.

Kalloo, G. 1991. Genetic improvement of tomato, Springerverlag, Berlin Heidelberg, Germany, 358.

Kamachi, Y., M. Uchikawa and H. Kondoh. 2000. Pairing SOX off: with partners in the regulation of embryonic development. *Trends in Genetics*, 16(4): 182-187.

Kidokoro, S., K. f, Z.K. Shinwari, K. Shinozaki and K.Y. Shinozaki. 2009. The Phytochrome-Interacting Factor PIF7 negatively regulates *DREB1* expression under circadian control in *Arabidopsis. Plant Physiol*., 151(4): 2046-2057.

Konig, P. and T.J. Richmond. 1993. The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Bio*., 233(1): 139-154.

Latchman, D.S. 1997. Transcription factors: an overview. *Int. J. Biochem.& Cell Bio*., 29(12): 1305-1312.

Maston, G.A., S.K. Evans and M.R. Green. 2006. Transcriptional regulatory elements in the human genome. *Annual Review of Genomics and Human Genetics*., 1: 29-59.

Mueller, L.A., S.D. Tanksley, J.J. Giovannoni, J. van Eck, S. Stack, D. Choi, B.D. Kim, M. Chen, Z. Cheng, C. Li, H. Ling, Y. Xue, G. Seymour, G. Bishop, G. Bryan, R. Sharma, J. Khurana, A. Tyagi, D. Chattopadhyay, N.K. Singh, W. Stiekema, P. Lindhout, T. Jesse, R.K. Lankhorst, M. Bouzayen, D. Shibata, S. Tabata, A. Granell, M.A. Botella, G. Giuliano, L. Frusciante, M. Causse and D. Zamir. 2005. The tomato sequencing project, the first cornerstone of the International Solanaceae Project (SOL). *Comparative and Functional Genomics*, 6(3): 153-158.

Nakashima, K., Z.K. Shinwari, S. Miura,Y. Sakuma, M. Seki, K.Y. Shinozaki and K. Shinozaki. 2000 Structural organization, expression and promoter activity of an Arabidopsis gene family encoding DRE/CRT binding proteins involved in dehydration- and high salinity-responsive gene expression. *Plant Mol. Bio*., 42(4): 657-665.

Narusaka, Y., K. Nakashima, Z.K. Shinwari, Y. Sakuma, T. Furihata, H. Abe, M. Narusaka, K. Shinozaki and K.Y. Shinozaki. 2003. Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high salinity stresses. *The Plant Journal*, 34(2): 137-149.

Reeves, R. and M.S. Nissen. 1990. The AT-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *Journal of Biological Chemistry*, 265(15): 8573-8582.

Ren, J., L. Wen, X. Gao, C. Jin, Y. Xueand and X. Yao. 1997. DOG 1.0: Illustrator of protein domain structures. *Cell Research*, 19: 271-273.

Richards, E.J. 1997. Preparation of plant DNA using CTAB. In: *Short Protocols in Molecular Biology*. (Eds): F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl. 3rd Ed. New York, John Wiley. 2: 10-11

Roeder, R.G. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences*, 2(9): 327-335.

Sandelin, A., W.W. Wasserman and B. Lenhard. 2004. ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Research*, 32.

Savolainen, V., R. Spichiger and J.F. Manen. 2000. Polyphyletism of celastrales deduced from a chloroplast noncoding DNA region. *Molecular Phylogenetics and Evolution,* 7(2): 145-157.

Shinwari, Z.K., R. Terauchi and S. Kawano. 1994. Molecular systematics of liliaceae-Asparagoideae-Polygonatae. 1. RFLP analysis of cpDNA in several species of Asiatic Disporum species. *Plant Species Bio*., 9: 11-18.

Shinwari, Z.K., R. Terauchi and S. Kawano. 1994b. Phylogenetic relationships among genera in the Liliaceae-Asparagoideae-Polygonatae sensu lato inferred from rbcL gene sequence data. *Plant Systematic & Evolution*, 192: 263-277.

Shinwari, Z.K., R. Terauchi, F.H. Utech and S. Kawano. 1994a. Recognition of the New World Disporum Section Prosartes as Prosartes (Liliaceae) based on the sequence data of the rbcL gene. *Taxon*, 43(3): 353-366.

Si, J., Z. Zhang, B. Lin, M. Schroeder and B. Huang. 2011. Meta DB Site: A meta approach to improve protein DNA- binding sites prediction. *BMC Systems Biology,* 5(Suppl 1): S7.

Starr, B.D., B.C. Hoopes and D.K. Hawley. 1995. DNA bending is an important component of site-specific recognition by the TATA binding protein. *J. Mol. Bio*., 250(4): 434-446.

Thanos, D. and T. Maniatis. 1995. Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell,* 83(7): 1091-1100.

Tjong, H. and H.X. Zhou. DISPLAR: an accurate method for predicting DNA-binding sites on protein surfaces. *Nucleic Acids Research*, 35: 1465-1477.

Vicente-Carbajosa, J. and P. Carbonero. 2005. Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int. J. Developmental Bio*., 49(5-6): 645-651.

Wang, L., M.Q. Yang and J.Y. Yang. 2009. Prediction of DNA-binding residues from protein sequence information using random forests. *BMC Genomics*, 10(1).

Werner, M.H., J.R. Huth, A.M. Gronenborn and G.M. Clore. 1995. Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, 81(5): 705-714.

(Received for publication 10 May 2014)