ORIGINAL ARTICLE

An *in Silico* Analysis of Physicochemical Characterization and Protein-Protein Interaction Network Analysis of Human Anti-apoptotic Proteins

V. G. Vidhya[1], Y. Swarnalatha[2], Anusha Bhaskar[3]

[1]Department of Biotechnology, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Kancheepuram District, Tamil Nadu, India, [2]Department of Biotechnology, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India, [3]Abhi Diagnostic and Research Lab, Perambalur, Tamil Nadu, India

Abstract

Introduction: Apoptosis is a physiological mechanism, playing an essential role in regulating development, homeostasis, and immune defense by removing abnormal cells in organisms. A balance between pro-apoptotic and anti-apoptotic mechanisms determines cell death signal where the pro-apoptotic proteins promote apoptosis and anti-apoptotic proteins inhibit apoptosis. In general, the inhibitors of apoptosis proteins (IAPs) inhibit the caspase activation pathways and play an important roles in regulating apoptosis in many species. **Methodology:** A total of 50 different human anti-apoptotic proteins retrieved from UniProt Database were analyzed and characterized using *in silico* tools. A parsimonious phylogenetic tree for these proteins was constructed using the Poisson correction model. Genomic and proteomic data are often combined with protein-protein interaction networks (PPIN) whose structure is routinely analyzed by tools to characterize hubs for treating cancer. **Results and Discussion:** Primary structure analysis shows that most of the anti-apoptotic proteins are hydrophilic in nature due to the high content of glutamate and serine residues. The presence of disulfide bonding 27 proteins infers that these proteins may form disulfide bonds, which are regarded as a positive factor for stability. The aliphatic index computed by Expasy's ProtParm infers that these proteins may be stable for a wide range of temperature. Secondary structure analysis shows that most of the human anti-apoptotic proteins have predominant coiled structures due to the rich content of more flexible glycine and proline amino acids. Top 10 hub proteins were identified using PPIN analysis. **Conclusion:** Thus, the characterization of human anti-apoptotic proteins provides additional targets and new therapeutic approaches for treating cancer.

Key words: Anti-apoptotic proteins, inhibitors of apoptosis protein, protein-protein interaction networks

INTRODUCTION

Apoptosis is the most common form of cell death in all organisms. However, the evasion of apoptosis is the reason for tumor establishment and growth and is said to be the hallmark of cancer.[1] Cancer, generally, occurs due to deregulated cellular proliferation. As tumor growth progresses, the surrounding cellular environment becomes deficient in growth factors and lacks adequate oxygen supply. In normal cells, the above conditions would trigger apoptosis, but cancer cells proliferate in these adverse conditions.[2]

The inhibitors of apoptosis inhibitors of apoptosis proteins (IAPs) are a group of anti-apoptotic proteins that were first identified in baculoviruses. These proteins are evolutionarily conserved and found to contain a Zn^{2+} ion coordinating protein-protein interaction motif called baculovirus IAP repeat domain and a RING finger domain at their carboxyl terminus.[3] These proteins render cancer cells

Address for correspondence:
Vidhya VG. Department of Biotechnology, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur – 603 203, Kancheepuram District, Tamil Nadu, India. Phone: +91-9500031501.
E-mail: vidhyavg@gmail.com

Received: 09-10-2018
Revised: 04-12-2018
Accepted: 15-12-2018

insensitive to apoptotic stimulation through direct caspase and pro caspase inhibition with the help of transcription factor NF-Kappa β.[4] Thus, when the IAPs are overexpressed, cells are no longer able to die in a physiologically programmed fashion and become increasingly resistant to standard chemo and radiation therapies.[5] With the recognition of apoptosis suppression as a fundamental aspect of human cancer, the IAPs and other anti-apoptotic proteins are now acknowledged as outstanding therapeutic targets.

The major drawback of experimental methods that are used to characterize proteins is the time frame involved and of the high cost. Hence, *in silico* approach provide a reliable and ready solution to these problems. The computational tools also help us in understanding physicochemical, structural, and functional properties of proteins. The building blocks of proteins called amino acids provide information required for determining the molecules function, physical, chemical, and structural properties. Therefore, in the light of the above, a complete *in silico* analysis of 50 human anti-apoptotic proteins was carried out to better understand and emphasize the need for novel therapeutic approaches for cancer therapy.

METHODOLOGY

Sequence retrieval

A total of 50 human anti-apoptotic protein sequences were retrieved from UniProt, a public domain database to analyze their physiochemical, structural, and functional properties.

Amino acid composition

The amino acid composition of chosen proteins was computed using CLC workbench.[6]

Primary structure analysis

Counts of hydrophilic and hydrophobic residues of anti-apoptotic proteins were calculated from the primary structure of proteins using CLC workbench.

Physiochemical properties

The computed amino acids of anti-apoptotic protein sequences contain various information such as isoelectric point (pI), molecular weight (Mw), extinction coefficient (Ec), instability index (II), aliphatic index (AI), and Grand average of hydropathicity (GRAVY). As these parameters are very essential for studying their physiochemical properties, they were computed using Expasy's ProtParm tool.[7]

Conserved motif identification

Motifs were identified in profiles using Multiple EM for Motif Elicitation (MEME) server.[8]

Functional analysis

To determine the functional linkage and stability of the proteins, the presence and absence of cysteine bond (disulfide bond) and their bonding pattern were predicted with the help of CYS_REC tool.[9]

Secondary structure prediction

SOPMA tool was used for predicting the secondary structure of proteins. The method works by making consensus prediction from multiple alignments. Using this tool, the positional possibility of four states - alpha helix, β strands, turns, and random coils was assessed using default parameters with a window width of 17 and similarity threshold -8.

Multiple sequence alignment and phylogenetic analysis

The evolutionary alignment of the protein sequences was inferred using maximum parsimony (MP) method and the evolutionary divergence between the protein sequences were calculated using the Poisson correction method. All positions containing gaps and missing data were eliminated. The final evolutionary parsimonious tree was formed using MEGA.[10]

Protein-protein interaction

For studying the protein-protein interaction network (PPIN) of anti-apoptotic proteins they were converted to seed sequences to mine PPI data from STRING tool. Text mining, experiments, databases, co-expression, neighborhood, gene fusion, and cooccurance were the interaction sources selected for constructing the PPI network. PPIs that possessed only high confidence score of 0.7 were considered for network generation. Network construction and visualization were done by Cytoscape v 3.5.1. The topological parameters of PPI network, that is, number of nodes, number of edges, average node degree, average clustering coefficients, topological coefficients, and shortest path lengths were noted through network analyzer, by treating the network as a directed graph.[11] In addition, functional enrichment of input seed sequences of anti-apoptotic protein sequences was carried out by STRING to identify significantly enriched gene ontology (GO) biological processes and molecular function. Cytohubba, a java plugin for Cytoscape software was employed to determine the hub proteins of the PPI network of seed proteins.[12] In this study, two centrality measurements, i.e., betweenness and radiality were applied to mine the top 10% hub proteins of the network.

RESULTS AND DISCUSSION

Sequence retrieval

A total of 50 human anti-apoptotic proteins were searched and retrieved from UniProt protein database and listed in Table 1 along with their accession number. The identified protein sequences were downloaded in FASTA format and used for studying their primary, secondary structures, physicochemical, properties, and their interaction with other proteins along with their function by using various computational tools and servers.

Amino acid composition

The amino acid composition of human anti-apoptotic protein sequences was computed using CLC Workbench. From the results, it was observed that leucine is the most abundant amino acid present in all these proteins and the next abundant amino acids present predominantly were serine, glutamate, and alanine. The composition of tryptophan was present in the least quantity compared to all other amino acids [Table 2]. The presence of aspartic acid in some proteins is necessary to make contact with solvent due to their ability to form hydrogen bonds. Since these interactions are often crucial for the stabilization of the protein three-dimensional structure, they are normally conserved.

Primary structure analysis

The primary sequence analysis results shown in Table 3 revealed that among 50 proteins considered 46 proteins are hydrophilic in nature due to the presence of high content of glutamate and serine whereas the other four proteins Bcl-2 - apoptosis regulator (P10415), HTRA2 - serine protease (O43464), Protein lifeguard 2 and 4, Q9BWQ8 and Q9HC24, respectively, are hydrophobic in nature due to the presence of high non-polar amino acids.

Physiochemical properties

Physiochemical characterization is very important to characterize specific proteins. The average Mw of human anti-apoptotic proteins calculated is 3,877,450 dalton. Expasy's ProtParam tool was used to characterize proteins. The results are shown in Table 4. Protein pI is calculated using the pKa values of amino acids. The pKa value of amino acids is important in defining the pH-dependent characteristics of a protein. The computed pI value of proteins Q05655, Q99933, O60313, O14746, O43521, Q9NZM5, Q96TA2, Q9UK96, O43464, Q9P286, Q01851, and P49842 is >7 which indicates that these anti-apoptotic proteins are basic, and the pI of all other proteins is <7 which reveals that these are acidic in character. The Ec predicts the amount of light absorbed by a protein at certain wavelength. Although Expasy's ProtParam

Table 1: List of human anti-apoptotic proteins

Accession	Protein name
Q6UX06	Olfactomedin-4
Q05655	Protein kinase C delta type
O00571	ATP-dependent RNA helicase DDX3X
Q96J02	E3 ubiquitin-protein ligase Itchy h...
P49756	RNA-binding protein 25
Q9NR09	Baculoviral IAP repeat-containing p...
O14746	Telomerase reverse transcriptase
Q969V6	MKL/myocardin-like protein 1
P08069	Insulin-like growth factor 1 recept...
Q8IUQ4	E3 ubiquitin-protein ligase SIAH1
P98170	E3 ubiquitin-protein ligase XIAP
Q9UKV3	Apoptotic chromatin condensation in...
O43521	Bcl-2-like protein 11
P42574	Caspase-3
P49747	Cartilage oligomeric matrix protein
Q9NZM5	Ribosome biogenesis protein NOP53
P31751	RAC-beta serine/threonine-protein k...
Q99933	BAG family molecular chaperone regu...
P55210	Caspase-7
P63167	Dynein light chain 1, cytoplasmic
Q96TA2	ATP-dependent zinc metalloprotease...
P10415	Apoptosis regulator Bcl-2
O15392	Baculoviral IAP repeat-containing p...
O43255	E3 ubiquitin-protein ligase SIAH2
Q13077	TNF receptor-associated factor 1
Q9UK96	F-box only protein 10
Q14457	Beclin-1
P49407	Beta-arrestin-1
Q6FI81	Anamorsin
O43464	Serine protease HTRA2, mitochondria...
P31749	RAC-alpha serine/threonine-protein...
Q9P286	Serine/threonine-protein kinase PAK...
P55212	Caspase-6
Q01851	POU domain, class 4, transcription ...
P55957	BH3-interacting domain death agonist...
P52333	Tyrosine-protein kinase JAK3
Q96RI1	Bile acid receptor
Q9Y2G2	Caspase recruitment domain-containi.
O60346	PH domain leucine-rich repeat-conta.
Q6ZVD8	PH domain leucine-rich repeat-conta.
P49842	Serine/threonine-protein kinase 19
Q9Y371	Endophilin-B1
Q8NFZ5	TNFAIP3-interacting protein 2
O60313	Dynamin-like 120 kDa protein, mitoc...

(Contd...)

Table 1: (*Continued*)

Accession	Protein name
Q86WB0	Nuclear-interacting partner of ALK
Q9NR28	Diablo homolog, mitochondrial
Q9BWQ8	Protein lifeguard 2
Q07820	Induced myeloid leukemia cell diffe...
Q8IWZ3	Ankyrin repeat and KH domain-contai...
Q9HC24	Protein lifeguard 4

computes the Ec for a range of 276, 278, 279, 280, and 282 nm wavelength, 280 nm is favored because proteins absorb strongly there. The extinction coefficient of anti-apoptotic proteins at 280 nm ranges at a maximum of 1.619 M–1 cm–1 with respect to the concentration of Cys, Trp, and Tyr. The computed EC values will help in the quantitative study of protein-protein and protein-ligand interactions in solution. Stability of anti-apoptotic proteins was studied by analyzing the values for II, AI, and GRAVY index. The II provides an estimate of the stability of the proteins in a test tube. A protein

Table 2: Amino acid composition of human anti-apoptotic proteins computed using CLC workbench

Acc No	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
Q6UX06	4.3	1	5.1	5.7	3.7	6.7	1	4.3	5.3	10.6	2.2	6.9	4.1	3.9	4.1	9.2	6.9	8.2	1.2	5.7
Q05655	5.8	3.1	5.9	7.2	6.8	6.7	2.8	5.8	8.7	8.4	2.7	4.3	3.6	3.8	4.9	5.5	4.4	5.3	1.3	3
O00571	5.6	1.1	7.3	6.2	4.7	11.5	2.1	4.7	4.8	7.1	2.1	3.6	3.6	3	8.3	11.3	3.6	4.7	1.1	3.6
Q96J02	3.5	1.7	5.1	7.3	4.8	7.4	1.8	4.5	5.2	8.6	1.9	5.4	6.9	5.3	6	7.6	6.2	5.3	2.2	3.2
P49756	5.1	0.7	7	17.6	1.8	3.4	1.3	4.7	10.3	6.5	2	2.4	7.1	2.8	13.2	5.2	2.8	3.7	0.7	1.5
Q9NR09	7.5	2.3	5	5.9	2.7	5.7	3	4.6	4.7	12.8	2.2	3.9	5.6	5	3.8	10.1	6.3	6.4	1	1.8
O14746	8.7	2.6	3	4	4.2	6.6	3	2	3.5	13	1.1	1.9	7.7	4.2	11	6.6	5.1	7.8	1.6	2.5
Q969V6	8.4	0.8	4.4	6.4	1.7	6	1.9	2.7	5.5	12.1	1.8	1.7	12.9	7.4	3.4	12	5.3	4.4	0.2	0.9
P08069	5.3	3.2	4.5	8.3	3.4	6.7	1.7	5.3	5	8.5	2.9	6.2	5.9	2.6	5.9	7.1	5	6.4	1.5	4.6
Q8IUQ4	7.4	6.4	3.5	5.3	4.6	6	3.5	5	3.5	9.6	3.2	4.3	6.4	5	4.3	7.1	7.1	5	1.1	1.8
P98170	6	4	5.2	7.4	5.4	6	2.4	4.8	5.8	6.6	2	5.4	4.2	4.6	5.8	7.8	5.6	5	1.6	3.8
Q9UKV3	6.1	0.5	4.8	14.8	0.9	4.8	2.2	2.3	7.2	6.9	1	1.4	8.1	5.4	11.2	11.3	5.3	4.5	0.6	0.6
O43521	8.6	2	4.5	5.6	4	6.1	2.5	3	1	7.1	4	3	12.1	6.1	10.1	10.6	3	2.5	1	3
P42574	4.3	2.9	7.2	7.2	5.4	5.8	2.9	6.9	7.9	7.2	3.6	5.4	2.5	1.4	5.1	9.4	5.8	4.7	0.7	3.6
P49747	6.2	6.1	12.9	4.5	3.4	9.4	1.8	2.5	2.6	5	1.1	5.7	6.2	6.7	6.2	4.8	4.6	7.4	1.3	1.5
Q9NZM5	9.8	0.4	4.4	9.6	2.9	6.1	1.3	1.3	9	11.9	0.4	1.9	6.3	6.7	12.6	4.8	4.2	5	0.6	0.8
P31751	5.4	1.5	6.2	8.9	5.4	5.6	2.3	4.4	6.4	9.4	3.3	2.5	5.2	3.1	7.3	5.4	6	6.2	1.5	4
Q99933	7.5	1.2	3.5	13.3	2	6.1	1.4	3.2	7.2	9.3	2	2	5.2	6.7	9.9	7.2	6.7	5.2	0.3	0
P55210	5.9	3.6	8.9	6.3	5.6	5.9	2.3	5.6	8.3	6.6	2.3	4.6	4	3.6	5	6.9	5	5.9	0.7	3
P63167	7.9	3.4	5.6	7.9	5.6	4.5	4.5	7.9	11.2	4.5	3.4	4.5	1.1	4.5	2.2	4.5	4.5	5.6	1.1	5.6
Q96TA2	6.5	0.6	4.8	7	5	6.6	2.6	6.1	7.4	10.1	2.7	4.1	4.7	3.8	5.4	7.8	6.2	6.1	0.6	1.9
P10415	11.3	0.8	5	4.2	5	9.2	3.8	2.9	1.7	9.2	2.9	2.5	8	2.5	7.1	6.3	5	6.7	2.5	3.3
O15392	9.2	4.2	4.9	13.4	7.7	4.2	2.8	3.5	11.3	7.7	2.1	3.5	7	2.8	4.2	2.8	4.9	1.4	2.1	0
O43255	10.8	5.9	2.8	4.6	3.4	7.4	4.6	4.6	3.4	8	2.5	3.4	9	4.6	3.4	7.7	5.9	5.2	0.9	1.9
Q13077	7.7	4.1	4.1	8.4	4.6	6.3	2.4	2.6	5	12	1.9	3.1	5.8	6	5.5	8.7	4.3	5	0.7	1.7
Q9UK96	5.6	2.8	4	5.9	3.2	9.6	2.9	7	4.5	8.9	1.4	6.5	4.4	4	6.3	8.7	4	6.6	1.4	2.4
Q14457	4.7	2	5.3	11.3	4.7	5.3	1.3	3.1	6.4	10.9	2.9	5.3	2.9	7.1	4.9	6.4	6.7	4.7	1.6	2.4
P49407	5.5	1.7	7.4	8.4	3.8	5.5	2.2	3.6	8.4	10	1.2	4.5	7.2	2.4	5.7	4.1	6.5	8.9	0	3.1
Q6FI81	8.7	3.2	6.1	8.3	2.6	6.4	1.6	3.2	9	12.2	1.6	3.2	5.4	3.2	3.2	10.9	3.5	6.4	0.6	0.6
O43464	10.7	0.2	4.1	4.6	2	9.8	1.3	5	1.3	9.8	1.3	2.4	7.4	3.5	9.6	7.2	7	9.6	1.3	1.7
P31749	5.2	1.5	5.8	10.2	5.6	6.3	2.7	4.2	7.5	8.5	3.3	2.7	4.6	3.5	6.3	4.8	6.3	5.8	1.5	3.8
Q9P286	4.7	1.1	5.4	5.4	3.1	6.1	3.8	3.9	5.7	8.2	2.9	2.2	7.8	4.9	5.6	13.2	5	5	1	5
P55212	6.5	3.4	6.8	6.8	6.1	6.5	4.1	4.4	6.8	8.9	2.4	3.8	3.4	2.4	5.8	6.1	5.5	6.1	0.7	3.4
Q01851	16.2	1.2	2.6	4.5	2.4	13.4	6.4	2.9	5	8.6	2.9	2.6	7.6	3.3	3.8	8.1	3.3	3.6	0.5	1

(Contd...)

Table 2: (***Continued***)

Acc No	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
P55957	6.7	1.5	7.7	7.2	2.6	5.1	2.6	3.6	2.1	14.4	2.6	6.7	3.1	4.6	9.2	8.7	4.6	5.1	0.5	1.5
P52333	7.8	2.8	5.2	5.9	4.1	7	3.3	3.3	3.7	13.4	1.8	1.3	6.9	4.7	6.9	7.9	3.4	6.4	1.3	2.7
Q96RI1	4.5	3.5	3.5	9.3	3.9	5.4	2.7	4.5	6.8	9.5	4.7	3.9	4.9	6.6	5.1	6.4	5.6	4.9	0.6	3.7
Q9Y2G2	5.6	0.7	5.1	8.8	4.9	5.6	3.5	4.4	4.6	11.6	2.1	3.5	5.6	4.6	4.2	8.4	3.9	7.9	1.6	3.5
O60346	9.6	2.4	4.3	6.9	2.4	7.2	2.5	3.1	3.9	11.4	1.6	4	8.7	4.7	5.6	9.3	4	6.3	0.5	1.7
Q6ZVD8	4.9	3	5	7	2.6	5.5	2.9	3.9	4	14.5	2	5.3	5	4.4	4.9	9.1	7	6	0.5	2.2
P49842	7.6	1.9	5.7	4.6	4.6	9.8	2.2	4.1	3.5	9	1.4	0.8	6.8	4.6	10.9	6	5.2	7.9	1.6	1.9
Q9Y371	8.5	1.4	4.9	7.9	3	5.2	1.1	4.9	6.8	12.1	2.5	6	3.3	4.9	4.9	7.4	7.1	4.1	0.5	3.3
Q8NFZ5	12.1	2.3	4.9	11.9	0.7	5.6	3.7	2.1	3.3	11.4	2.1	1.6	3.5	9.3	11.4	5.1	2.8	4	0.9	1.2
O60313	5.5	1	5.9	9.2	3.8	3.4	2.8	5.3	8.9	10.5	2.1	3.9	3.3	4.9	6.5	6.9	5.3	6.6	1.9	2.4
Q86WB0	6.8	4	5.6	7.6	4	5.4	1.4	3	5	9.4	2.2	0.8	7.8	4.4	6.2	13.3	6.2	4.6	2.2	0.4
Q9NR28	10.5	1.7	1.7	11.7	2.5	2.9	2.1	4.2	5.4	9.6	2.5	1.3	1.7	6.3	5.9	8.8	9.2	6.7	1.7	3.8
Q9BWQ8	8.5	1.9	3.2	3.2	8.2	6.3	1.9	3.5	2.8	13.3	1.9	2.5	6	3.8	2.5	7.3	8.2	7.9	1.9	5.1
Q07820	10.6	0.6	4.6	8.9	3.1	11.7	1.1	4.6	3.7	10	1.7	2.6	6.6	1.7	8	6	5.7	6.6	0.9	1.4
Q8IWZ3	9.4	1.4	5	6.3	2.4	7.7	2.9	3.5	4.9	9.1	1.9	4.9	7	5	3.5	10.1	7.1	6.6	0.4	1
Q9HC24	6.7	0.6	3.3	3.3	11.1	4.4	1.7	6.1	3.3	17.8	1.7	1.7	2.8	1.7	3.3	8.3	7.8	8.3	0.6	5.6

whose II is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. This study results showed that the II value of seven proteins Q6UX06, P31751, P55210, Q14457, P49407, P31749, and P55212 are in the range of 30.03–39.29 (II < 40) indicating that these proteins are stable while the rest are unstable (II > 40). The AI is a parameter for estimating thermal stability of a protein directly associating with the mole fraction of aliphatic side chains (Alanine, isoleucine, leucine, and valine) in the protein. In this study, high AI values of proteins (55.3–129.87) imply high thermostability of these proteins. The GRAVY value for a protein or a peptide is calculated by adding the hydropathy values (Kyte and Doolittle, 1982) of each amino acid residues and dividing by the number of residues in the sequence or length of the sequence. GRAVY index indicates the solubility of proteins, increasing positive score indicates a greater hydrophobicity. A low GRAVY value deciphers that there is better interaction between protein and water. The GRAVY index of anti-apoptotic proteins is ranging from -1.481 to 0.853.

Conserved motif identification

A motif occurrence is defined as a position in the sequence whose match to the motif has position $P < 0.0001$. For finding the conserved motifs present in these anti-apoptotic proteins MEME tool was used. MEME discovers novel, Ungapped motifs (recurring, fixed-length patterns) in these sequences. Weak hits are not displayed. Therefore, in this study, five significant conserved motifs were identified [Figure 1]. All the motif matches shown here have a position $P < 0.0001$ and sequences with an $E < 10$ are shown in the result.

Table 3: Primary structure analysis of human anti-apoptotic proteins

Accession	% of hydrophobic residues	% of hydrophilic residues
Q6UX06	40.20	48.04
Q05655	39.05	47.63
O00571	33.54	50.30
Q96J02	34.11	49.95
P49756	26.10	62.63
Q9NR09	38.91	47.56
O14746	40.81	42.32
Q969V6	32.22	48.12
P08069	37.89	46.31
Q8IUQ4	37.59	43.62
P98170	35.41	50.30
Q9UKV3	22.97	63.61
O43521	33.33	46.46
P42574	36.46	52.35
P49747	28.40	49.93
Q9NZM5	32.85	54.39
P31751	39.50	48.23
Q99933	29.56	57.97
P55210	35.65	50.82
P63167	41.57	49.44
Q96TA2	39.07	49.03
P10415	43.93	38.08
O15392	33.80	50.71
O43255	37.35	40.43

(Contd...)

Table 3:(*Continued*)

Accession	% of hydrophobic residues	% of hydrophilic residues
Q13077	36.30	47.60
Q9UK96	36.51	46.65
Q14457	34.89	54.89
P49407	36.13	49.52
Q6FI81	35.90	49.04
O43464	41.49	41.05
P31749	37.92	49.79
Q9P286	33.80	51.18
P55212	38.57	48.12
Q01851	37.95	39.86
P55957	36.92	53.33
P52333	40.84	42.44
Q96RI1	36.42	49.79
Q9Y2G2	41.53	46.63
O60346	36.58	45.08
Q6ZVD8	36.74	49.74
P49842	38.04	43.48
Q9Y371	38.91	51.23
Q8NFZ5	34.50	54.08
O60313	38.02	54.17
Q86WB0	32.47	50.40
Q9NR28	41.42	52.30
Q9BWQ8	50.32	35.44
Q07820	38.86	42.28
Q8IWZ3	34.31	49.65
Q9HC24	57.78	34.44

Functional analysis

Cysteine residues are important for determining the thermostability of proteins. The disulfide bonds that connect these cysteine residues are significant in the protein folding and stability, which are generated between the thiol groups of cysteine residues by oxidative folding process. In this study, the cysteine residues in the proteins were determined using CYS_REC server. The results revealed that among 50 proteins considered, 27 proteins contain cysteine residues connected by ss bonds [Table 5]. The presence of these disulfide bridges is regarded as a positive factor for stability at the molecular level.

Secondary structure prediction

The secondary structure of human anti-apoptotic protein sequences was predicted using SOPMA server. This tool evaluates the percentage of alpha helices, extended strand, beta turn, and random coils with an output width of 70. From the computed percentage of each conformation,

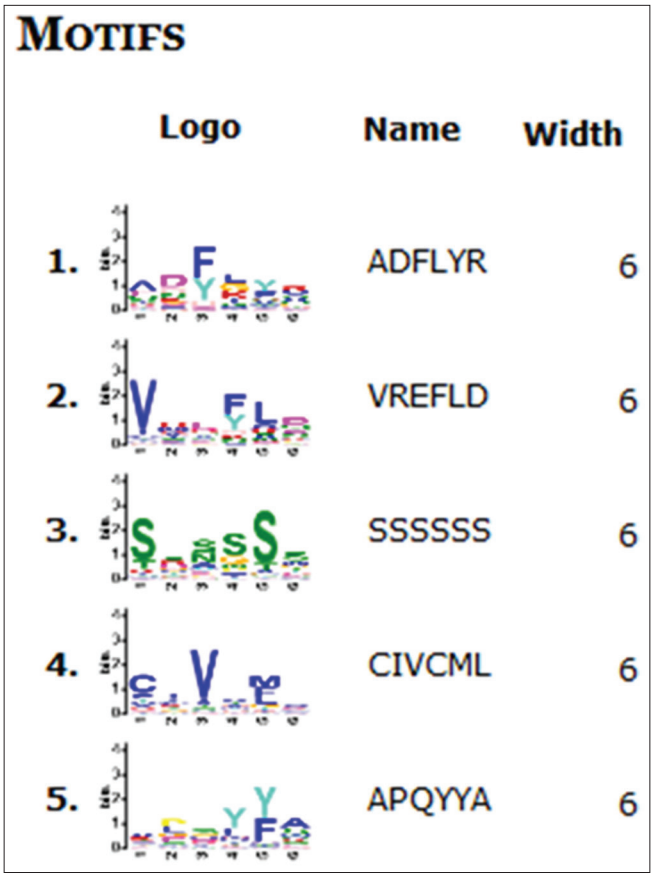

Figure 1: Conserved motifs identified using Multiple EM for Motif Elicitation tool

it is observed that in most proteins the percentage of the random coil was much greater than the percentage of other conformations such as helix, sheet, and turn. This high coiled structural content might be due to the presence of flexible glycine and hydrophobic proline amino acids. Proline has a special property of creating kinks in polypeptide chains and disrupting ordered secondary structure, thus results in coiling. No disordered protein binding sites were discovered. However, in few protein sequences (accession no: P49756, O14746, Q99933, P63167, Q96TA2, P10415, O15392, Q13077, Q14457, P55957, Q6ZVD8, Q9Y371, Q8NFZ5, O60313, Q9NR28, Q07820, and Q9HC24), the percentage of alpha helix was found to be higher than the percentage of random coils which might be due to the presence of high alanine content.

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignment is an important requisite for the evaluation of families of proteins and phylogenetic reestablishment. In this study, the evolutionary analysis was conducted in MEGA, and the results were inferred using the MP method. Analyzes were conducted using the Poisson correction model. The MP tree was obtained

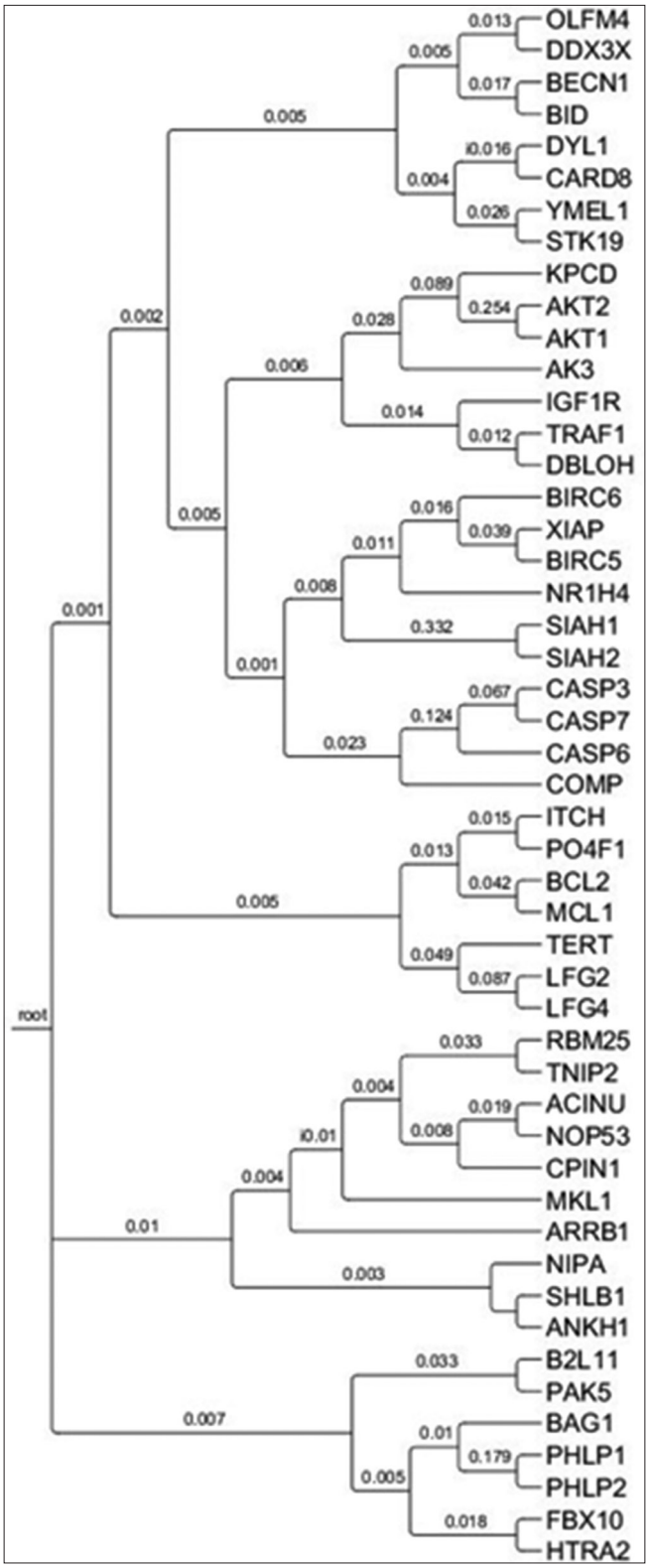

Figure 2: Bootstrap phylogenetic tree of human anti-apoptotic proteins using the maximum parsimony method

using the subtree-pruning-regrafting algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The analysis involved 50 amino acid sequences. All positions

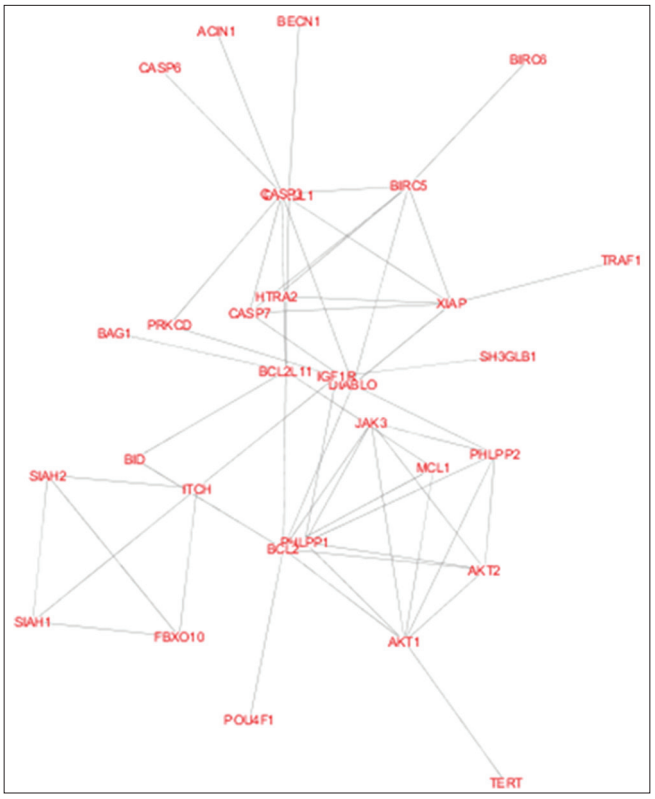

Figure 3: A map of protein-protein interaction networks constructed for human anti-apoptotic proteins using Cytoscape

containing gaps and missing data were eliminated. There were a total of 49 positions in the final dataset. The most parsimonious tree with length = 1567 is shown in Figure 2. The consistency index, the retention index, and the composite index were found to be 0.439694, 0.323575, and 0.142274 for all sites and parsimony informative sites.

Protein-protein interaction

Many proteins interact with each other and form a PPIN. PPIN helps in understanding their metabolic pathways, signaling cascades and also discover the function of newly identified proteins. In PPIN, a node indicates a protein and a connecting edge represents their interaction. Many proteins interact with very few proteins while some proteins have a very large number of connectivity. Proteins with large number of interactions are called hubs and their characterization is highly important for understanding cellular functions as these are the principal agents in the interaction network and affect its function and stability. In this study, the PPIN of human anti-apoptotic proteins was studied using Cytoscape tool and the top 10 hub proteins among them were identified using various network parameters such as gene proximity, gene fusion events, gene coexpression data, phylogenetic profiling, interacting protein domains, and GO. The complete PPIN

Table 4: Physicochemical characterization of anti-apoptotic proteins using ProtParm tool

Acc. No	Length	Mol. Wt	pI	Asp+Glu	Arg+Lys	EC	II	AI	GRAVY
Q6UX06	510	57279.8	5.5	55.0	48.0	1.3	30.0	86.3	−0.3
Q05655	676	77505.0	7.9	89.0	92.0	1.0	42.5	76.6	−0.4
O00571	662	73243.4	6.7	89.0	87.0	1.0	49.4	65.1	−0.6
Q96J02	903	102802.8	5.9	112.0	101.0	1.5	47.7	70.4	−0.6
P49756	843	100185.5	6.1	207.0	198.0	0.5	59.3	59.7	−1.5
Q9NR09	4857	530254.9	5.7	527.0	409.0	0.7	49.9	93.9	−0.1
O14746	1132	126996.7	10.5	79.0	165.0	1.1	54.8	89.9	−0.2
Q969V6	931	98919.0	5.6	101.0	83.0	0.2	74.8	79.0	−0.5
P08069	1367	154793.1	5.6	175.0	149.0	1.4	48.0	77.6	−0.4
Q8IUQ4	282	31122.9	6.4	25.0	22.0	0.8	45.9	78.6	−0.1
P98170	497	56684.9	6.2	63.0	58.0	1.3	42.6	65.4	−0.5
Q9UKV3	1341	151861.5	6.1	264.0	246.0	0.4	77.0	55.1	−1.3
O43521	198	22171.0	8.4	20.0	22.0	0.9	95.1	55.3	−0.7
P42574	277	31607.9	6.1	40.0	36.0	0.8	40.4	72.9	−0.5
P49747	757	82860.5	4.4	132.0	67.0	0.9	42.4	57.0	−0.7
Q9NZM5	478	54389.2	10.3	67.0	103.0	0.4	61.3	75.8	−1.0
P31751	481	55768.8	6.0	73.0	66.0	1.2	35.1	77.0	−0.5
Q99933	345	38778.8	7.7	58.0	59.0	0.1	69.4	71.3	−0.9
P55210	303	34276.8	5.7	46.0	40.0	0.7	31.2	70.8	−0.5
P63167	89	10365.9	6.9	12.0	12.0	1.2	42.7	72.4	−0.4
Q96TA2	773	86455.3	8.9	91.0	99.0	0.6	44.7	87.2	−0.3
P10415	239	26265.9	6.8	22.0	21.0	1.7	51.6	78.0	−0.1
O15392	142	16388.7	5.7	26.0	22.0	1.0	50.0	57.2	−0.7
O43255	324	34614.7	6.7	24.0	22.0	0.7	49.7	75.4	−0.1
Q13077	416	46163.6	5.8	52.0	44.0	0.6	51.1	79.5	−0.3
Q9UK96	956	105195.5	8.5	94.0	103.0	1.0	45.6	86.8	−0.3
Q14457	450	51896.3	4.8	75.0	51.0	1.1	39.3	72.8	−0.7
P49407	418	47065.7	5.8	66.0	59.0	0.4	37.1	84.4	−0.5
Q6FI81	312	33582.3	5.4	45.0	38.0	0.4	47.8	87.2	−0.3
O43464	458	48840.9	10.1	40.0	50.0	0.9	43.1	96.5	−0.1
P31749	480	55686.4	5.8	77.0	66.0	1.2	35.5	71.7	−0.6
Q9P286	719	80744.9	8.2	78.0	81.0	1.1	47.5	66.4	−0.6
P55212	293	33310.0	6.5	40.0	37.0	0.8	37.8	76.2	−0.4
Q01851	419	42697.3	9.2	30.0	37.0	0.4	49.2	71.3	−0.3
P55957	195	21994.7	5.3	29.0	22.0	0.5	60.7	91.5	−0.5
P52333	1124	125098.9	6.8	125.0	120.0	1.0	52.2	91.6	−0.1
Q96RI1	486	55914.3	6.4	62.0	58.0	0.8	55.3	73.4	−0.5
Q9Y2G2	431	48932.5	5.1	60.0	38.0	1.2	46.7	90.9	−0.3
O60346	1717	184672.4	5.9	192.0	163.0	0.5	65.6	84.4	−0.3
Q6ZVD8	1323	146751.1	5.5	159.0	118.0	0.6	53.1	94.1	−0.2
P49842	368	40915.9	9.8	38.0	53.0	1.1	49.5	81.3	−0.3
Q9Y371	365	40796.3	5.8	47.0	43.0	0.7	41.1	86.7	−0.4
Q8NFZ5	429	48699.6	6.0	72.0	63.0	0.6	60.4	76.3	−0.9
O60313	960	111630.7	7.9	145.0	147.0	1.2	44.3	86.3	−0.6

(Contd...)

Table 4:(*Continued*)

Acc. No	Length	Mol. Wt	pI	Asp+Glu	Arg+Lys	EC	II	AI	GRAVY
Q86WB0	502	55261.5	5.4	66.0	56.0	1.1	69.3	68.2	−0.4
Q9NR28	239	27130.8	5.7	32.0	27.0	1.3	49.4	83.7	−0.3
Q9BWQ8	316	35109.7	6.1	20.0	17.0	1.6	42.0	96.9	0.4
Q07820	350	37337.4	5.5	47.0	41.0	0.6	52.2	86.5	−0.2
Q8IWZ3	2542	269457.5	5.5	287.0	212.0	0.3	50.1	77.6	−0.4
Q9HC24	238	26971.0	6.6	17.0	16.0	0.9	48.7	129.9	0.9

pI: Isoelectric point, GRAVY: Grand average of hydropathicity, II: Instability index, AI: Aliphatic index

Table 5: Disulfide bond patterns predicted by CYS_REC

Accession	Protein name	Cys rec
Q6UX06	Olfactomedin-4	83–437, 85–246
Q05655	Protein kinase C delta type	172–344, 175–189, 192–261, 200–244, 208–272, 247–264
O00571	ATP-dependent RNA helicase DDX3X	175–468
Q96J02	E3 ubiquitin-protein ligase Itchy h...	57–871, 156–621, 164–181, 170–855, 178–539
P49756	RNA-binding protein 25	83–612
Q9NR09	Baculoviral IAP repeat-containing p...	
O14746	Telomerase reverse transcriptase	7–828, 171–199, 271–321, 517–842
Q969V6	MKL/myocardin-like protein 1	326–783, 508–612, 509–1019, 584–1004
P08069	Insulin-like growth factor 1 recept.	
Q8IUQ4	E3 ubiquitin-protein ligase SIAH1	16–98, 55–130, 62–65, 72–75, 105–121, 128–135
P98170	E3 ubiquitin-protein ligase XIAP	12–481, 63–90, 66–303, 200–213, 202–471, 203–450, 300–474
Q9UKV3	Apoptotic chromatin condensation in...	691–733, 1052–1083
O43521	Bcl-2-like protein 11	12–55, 110–121
P42574	Caspase-3	not SS-bounded
P49747	Cartilage oligomeric matrix protein	not SS-bounded
Q9NZM5	Ribosome biogenesis protein NOP53	not SS-bounded
P31751	RAC-beta serine/threonine-protein k...	not SS-bounded
Q99933	BAG family molecular chaperone regu.	322–330
P55210	Caspase-7	not SS-bounded
P63167	Dynein light chain 1, cytoplasmic	not SS-bounded
Q96TA2	ATP-dependent zinc metalloprotease.. .	not SS-bounded
P10415	Apoptosis regulator Bcl-2	not SS-bounded
O15392	Baculoviral IAP repeat-containing p...	not SS-bounded
O43255	E3 ubiquitin-protein ligase SIAH2	15–104, 94–161, 101–175, 110–170, 111–168, 114–322, 138–145
Q13077	TNF receptor-associated factor 1	22–27, 38–41, 54–57, 88–306, 95–132, 169–280
Q9UK96	F-box only protein 10	
Q14457	Beclin-1	18–159, 137–140, 165–353
P49407	Beta-arrestin-1	not SS-bounded
Q6FI81	Anamorsin	116–246, 237–277, 249–274, 251–285
O43464	Serine protease HTRA2, mitochondria.	not SS-bounded
P31749	RAC-alpha serine/threonine-protein.. .	60–296, 224–460
Q9P286	Serine/threonine-protein kinase PAK.	77–327, 245–590
P55212	Caspase-6	87–209
Q01851	POU domain, class 4, transcription.	90–310

(Contd...)

Table 5: (*Continued*)

Accession	Protein name	Cys rec
P55957	BH3-interacting domain death agonist.	not SS-bounded
P52333	Tyrosine-protein kinase JAK3	54–475, 80–360, 115–1040, 162–839, 416–428, 474–695, 1024–1028
Q96RI1	Bile acid receptor	18–154, 20–179, 137–173, 140–197, 157–446, 189–217, 192–253
Q9Y2G2	Caspase recruitment domain-containi.	not SS-bounded
O60346	PH domain leucine-rich repeat-conta.	not SS-bounded
Q6ZVD8	PH domain leucine-rich repeat-conta.	not SS-bounded
P49842	Serine/threonine-protein kinase 19	54–223, 84–230
Q9Y371	Endophilin-B1	not SS-bounded
Q8NFZ5	TNFAIP3-interacting protein 2	21–165, 144–275, 171–405
O60313	Dynamin-like 120 kDa protein, mitoc.	11–786, 14–856, 375–853
Q86WB0	Nuclear-interacting partner of ALK	102–254, 117–156, 120–149, 125–500, 141–258, 249–272, 406–429
Q9NR28	Diablo homolog, mitochondrial	not SS-bounded
Q9BWQ8	Protein lifeguard 2	157–198, 158–223
Q07820	Induced myeloid leukemia cell diffe.	not SS-bounded
Q8IWZ3	Ankyrin repeat and KH domain-contai.	not SS-bounded
Q9HC24	Protein lifeguard 4	not SS-bounded

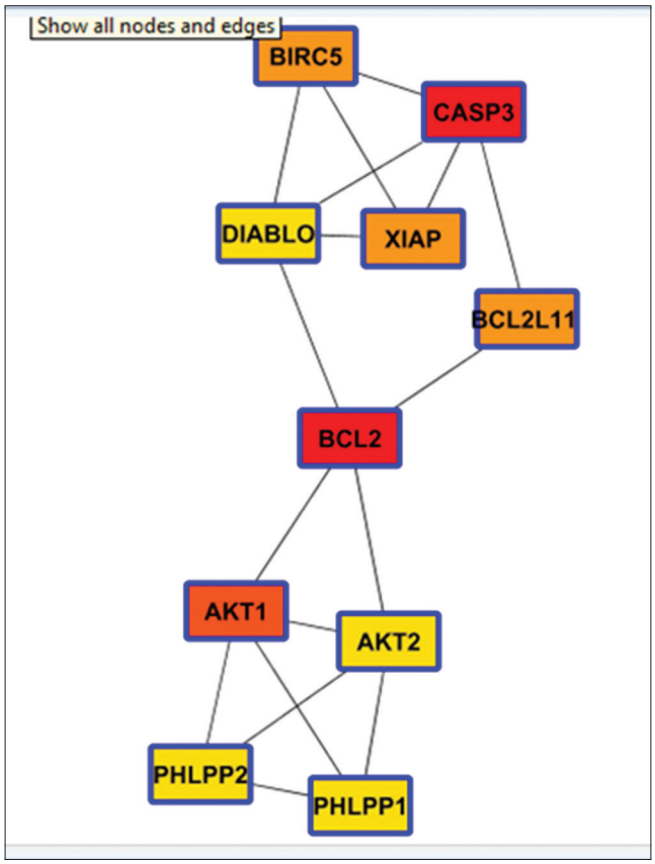

Figure 4: Top 10 hub proteins identified from protein-protein interaction networks using Cytohubba

of anti-apoptotic protein sequences and the hub proteins identified are shown in Figures 3 and 4.

REFERENCES

1. Wong RS. Apoptosis in cancer: From pathogenesis to treatment. J Exp Clin Cancer Res 2011;30:87.
2. Slatter TL, Hung N, Campbell H, Rubio C, Mehta R, Renshaw P, *et al.* Hyperproliferation, cancer, and inflammation in mice expressing a Δ133p53-like isoform. Blood 2011;117:5166-77.
3. Verhagen AM, Coulson EJ, Vaux DL. Inhibitor of apoptosis proteins and their relatives: IAPs and other BIRPs. Genome Biol 2001;2:REVIEWS3009.
4. Meng Z, Lou S, Tan J, Xu K, Jia Q, Zheng W, *et al.* Nuclear factor-kappa B inhibition can enhance apoptosis of differentiated thyroid cancer cells induced by 131I. PLoS One 2012;7:e33597.
5. Hunter AM, LaCasse EC, Korneluk RG. The inhibitors of apoptosis (IAPs) as cancer targets. Apoptosis 2007;12:1543-68.
6. Gupta S, Sarthi P, Buddhadev M, Kumar BA. *In silico* characterization of human tyrosinase using computational tools and servers. Int J Pharm Bio Sci 2013;4:181-93.
7. Goswami K, Varshitha J, Naresha S, Pattabhiramaiah M. Proteomic characterization of BRCA1 gene and its role in cancer progression. Res J Life Bioinform Pharm Chem Sci 2018. DOI: 10.26479/2018.0401.01.
8. Bose R, Arora S, Dwivedi VD, Pandey A. Amino acid sequence based *in silico* analysis of β-galactosidases. Int J Bioinform Biosci 2013;3:37-44.
9. Anusha PK, Mahesh P, Mandeep S, Reddy MS. Computational annotation of GSK -3 protein in homo sapiens. Int J Eng Sci Comput 2016;6:4720-33.
10. Asokan KV, Mundaganur DS, Mundaganur YD.

Catalase: Phylogenetic characterization to explore Protein cluster. J Res Bioinform 2010;1:1-8.
11. Vella D, Zoppis I, Mauri G, Mauri P, Di Silvestre D. From protein-protein interactions to protein co-expression networks: A new perspective to evaluate large-scale proteomic data. EURASIP J Bioinform Syst Biol 2017;2017:6.
12. Jin G, Zhang S, Zhang XS, Chen L. Hubs with network motifs organize modularity dynamically in the protein-protein interaction network of yeast. PLoS One 2007;2:e1207.

Source of Support: Nil. **Conflict of Interest:** None declared.