World Research Journal of Computer-Aided Drug Design
ISSN: 2320-5687 & E-ISSN: 2320-5695, Volume 1, Issue 1, 2012, pp.-01-08.
Available online at http://bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000049

NETWORKING OF SHANK AND TSC1 PROTEIN IN AUTISTIC DISORDER

RAMASWAMY V.* AND THAMBI M.K.

International Centre for Intellectual Training and Empowerment (INCITE), Sri Rams, N.H. Road, Kazhakuttom, Trivandrum-695 582, Kerala, India.
*Corresponding Author: Email- vidhya.ramaswamy@gmail.com

Received: April 18, 2012; Accepted: November 01, 2012

Abstract- Protein-Protein interactions occur when two or more protein binds together, often to carry out their biological functions. At least in yeast cells, protein-protein interactions are not random but, well organized. Protein networks in the protein-protein interactions helps within the context of all other cellular proteins. In protein networks or protein-protein interactions, some proteins are very busy "talking" to many other proteins, but some highly connected proteins are unlikely to "talk" to each other. Two proteins called as SHANK and TSC1 are involved in very different autism related syndrome. Both the proteins are not thought to be related, proved to be connected by 21 other proteins. The analysis of the networks against the DNA of patients, reports that the abnormalities involve three of the network genes. These findings suggest that different type of autism may share a common pathway even when they occur in distinct syndrome or alone. Using Bioinformatics tool String 9.0 the present study analyzed the protein networking between SHANK and TSC1. A study of networking of SHANK and TSC1 was initiated to understand the interaction of proteins within the network of proteins of Autistic disorders.

Keywords- Autism, SHANK, TSC1, Protein Networks

Citation: Ramaswamy V. and Thambi M.K. (2012) Networking of Shank and TSC1 Protein in Autistic Disorder. World Research Journal of Computer-Aided Drug Design, ISSN: 2320-5687 & E-ISSN: 2320-5695, Volume 1, Issue 1, pp.-01-08.

Introduction

Inside the cell the majority of proteins can be found in highly interactive networks. The architecture of the cellular network and how it changes in response to the cellular environment is critical for the normal physiological functioning of the cell. In fact, these networks are responsible for the robustness and adaptability of living cells and perturbations to these networks result in pathological conditions such as cancer and neurological disorders [1]. The network patterns revealed the results suggested common mechanisms in the cancer biology and system should provide a foundation for a network or pathway-based analysis platform for cancer and other diseases [2]. The response of the network to a stimuli's is largely driven by the formation of new, transitory interactions, which are typically regulated by post translational modifications. Protein-protein interactions occur when two or more proteins bind together, often to carry out their biological function. Protein-protein interactions are at the core of the entire interactomics system of any living cell. Interactions between proteins are important for the majority of biological functions. Strong and Eisenberg (2007) and others reported that protein networks help define individual proteins within the context of all other cellular proteins and suggested methods for the identification and analysis of genome-wide protein networks and how protein networks can be used to aid the identification of novel drug targets [1,3-8].

Autism is a disorder of neural development characterized by impaired social interaction and communication and by restricted and repetitive behavior. These signs all begin before a child is three years old. It affects information processing in the brain by altering how nerve cells and their synapses connect and organize. Autism has a strong genetic basis, although the genetics of autism are complex and it is unclear whether ASD is explained more by rare mutations or by rare combinations of common genetic variants. In rare cases, autism is strongly associated with agents that cause birth defects. It is highly variable neurodevelopment disorder that first appears during infancy or childhood and generally follows a steady course without remission. Autism has a strong genetic basis, although the genetics of autism are complex and it is unclear whether ASD is explained more by rare mutations with major effects or by rare multigene interactions of common genetic variants. At India's current population, there are more than 2 million autistic persons in the country. While the disorder is not rare, the majority of autistic people in India has not been diagnosed and do not receive the services they need. The needs of autistic children in India are not being met in either the regular or special education sys-

tems. In India, tremendous lack of awareness and misunderstanding about autism among the medical professionals, who may either misdiagnose or under diagnose the condition.

Numbers of proteins are involved in autism disorder. Royler et al., (2008) reveled, in the case of protein interaction networks, their topology has been explored through the clustering of proteins into groups that share the same biological function, are similarly localized in the cell, or are part of a complex [9]. SH3 and multiple ankyrin repeat domains protein (SHANK) is a protein that in humans is encoded by the SHANK gene. SHANK1 has been shown to interact with SPTAN1, BAIAP2, ARHGEF7, DNM2 and Somatostatin receptor 2. Shank1, Shank2 and Shank3 constitute a family of proteins that may function as molecular scaffolds in the postsynaptic density (PSD). Tuberous Sclerosis protein1, also known as TSC1 or hamartin, is a human protein and gene. This peripheral membrane protein was implicated as a tumor suppressor. It may be also involved in vesicular transport and docking, in complex with TSC2. Defects in this gene may cause tuberous sclerosis due to a functional impairment of the hamartin-tuberin complex.

The proteins, SHANK and TSC1 are involved in very different autism related syndrome. Both the proteins are not thought to be related, proved to be connected by 21 other proteins. The interaction between the SHANK and TSC1 proteins gives information about the coordinated functions of each and every protein in the network, its help to produce new therapeutic approaches for autism. From these interactions, functional similarity of interacting protein need to be known and these results are helpful to specify the protein responsible for the autism and finally lead to produce more effective drugs for autism.

Materials and Methods

Shank and TSC1 are the proteins which are involved in very different autism related syndrome. Using different bioinformatics tools, it is possible to generate the protein networking in between these proteins and to analyze the properties of proteins which are involved in the interacting network of TSC1 and Shank family of proteins. The interaction network between TSC1 and Shank family of proteins are implicated in molecular pathway for autistic disorder. The functional similarity of interacting proteins within the network helps to focus the specific proteins which are responsible for autism.

Proteins involved in Autistic Disorder

SHANK

SH3 and multiple ankyrin repeat domains protein is a protein that in humans is encoded by the SHANK gene. Shank proteins are scaffold proteins identified through their interaction with a variety of membrane and cytoplasmic proteins. Shank polypeptides contain multiple sites for protein-protein interaction, including ankyrin repeats, an SH3 domain, a PDZ domain, a long proline-rich region, and a SAM domain. The specific localization of Shank proteins at postsynaptic sites of brain excitatory synapses suggests a role for this family of proteins in the organization of cytoskeletal or signaling complexes at specialized cell junctions. SHANK1 has been shown to interact with SPTAN1, BAIAP2, ARHGEF7, DNM2 and Somatostatin receptor 2. Shank1, Shank2 and Shank3 constitute a family of proteins that may function as molecular scaffolds in the postsynaptic density (PSD).

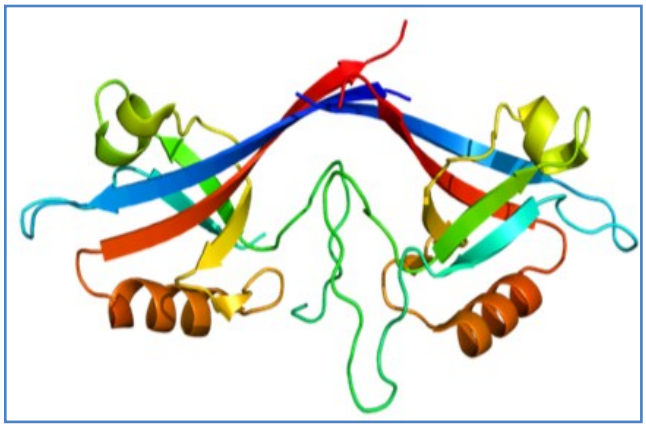

Fig. 1- Structure of SHANK1 protein

TSC1

Tuberous sclerosis protein 1, also known as TSC1 or hamartin, is a human protein and gene. This peripheral membrane protein was implicated as a tumor suppressor. It may be also involved in vesicular transport and docking, in complex with TSC2. Defects in this gene may cause tuberous sclerosis, due to a functional impairment of the hamartin-tuberin (TSC1-TSC2) complex. Defects in TSC1 may also be a cause of focal cortical dysplasia. Tuberous sclerosis complex (TSC) is an autosomal dominant disorder that results from mutations in the TSC1 or TSC2 genes and is associated with hamartoma formation in multiple organ systems. The neurological manifestations of TSC are particularly challenging and include infantile spasms, intractable epilepsy, cognitive disabilities and autism. TSC1 protein was expressed in the nervous system and in many endocrine tissues, including pancreatic islets, the parathyroids, testis and ovary. TSC1 was also detected in the many epithelial tissues of organs, such as kidney, uterus, small and large intestine and liver.

Tools Used to Predict the Protein Networking

STRING

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; which are derived from four sources. They are

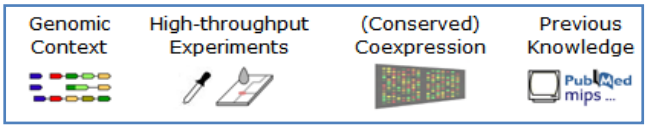

Fig. 2- STRING sources

STRING quantitatively integrates interaction data from these sources for a large number of organisms and transfers information between these organisms was applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

The network view summarizes the network of predicted associations for a particular group of proteins. The network nodes are proteins. Hovering over a node will display its annotation, clicking on a node gives several details about the protein. The edges represent the predicted functional associations. An edge may be drawn with up to 7 differently colored lines, these lines represent the existence of the seven types of evidence used in predicting the associations.

A red line indicates the presence of fusion evidence; a green line-neighborhood evidence; a blue line-cooccurrence evidence; a purple line-experimental evidence; a yellow line-text mining evidence; a light blue line-database evidence; a black line-coexpression evidence. Hovering over an edge will display the combined association score.

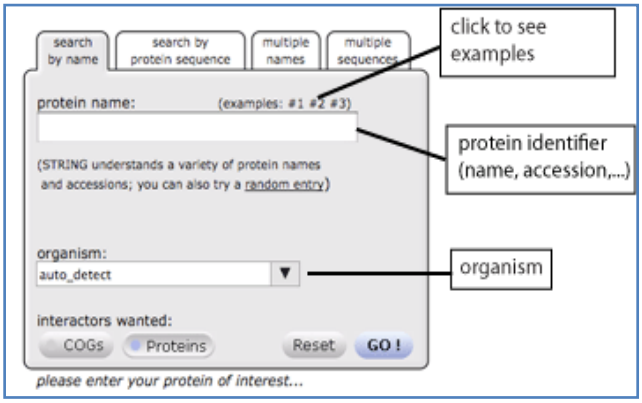

Fig. 3- STRING home page

T-COFFEE

T-COFFE is the Tree based Consistency Objective Function For Alignment Evaluation. T-Coffee is a novel method for multiple sequence alignments. The main characteristic of T-Coffee is that it will allow combining results obtained with several alignment methods (Clustal, Mafft, Probcons, Muscle...) into one unique alignment (M-Coffee). By default, T-Coffee will compare all sequences two by two, producing a global alignment and a series of local alignments (using lalign). The program will then combine all these alignments into a multiple alignment. T-Coffee can align Protein, DNA and RNA sequences. It is also able to combine sequence information with protein structural information (3D-Coffee/Expresso), profile information (PSI-Coffee) or RNA secondary structures (R-Coffee). This presentation gives an overview of the T-Coffee algorithm and of the original implementation and validation of the package.

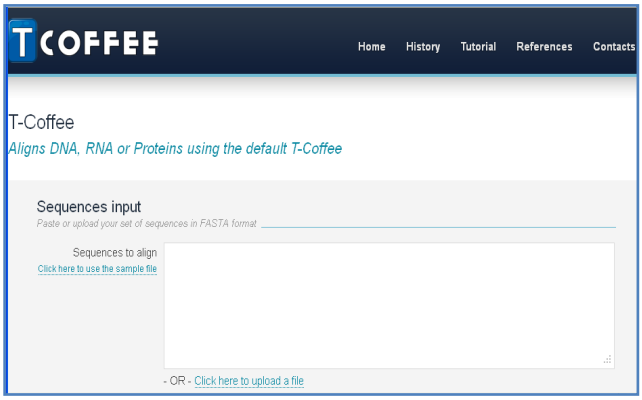

Fig. 4- T-COFFEE input page

MEGA

The Molecular Evolutionary Genetics Analysis (MEGA) software was developed with the goal of providing a biologist centric, integrated suite of tools for statistical analyses of DNA and protein sequence data from an evolutionary standpoint. Over the years, it has grown to include tools for sequence. MEGA is an integrated tool for conducting automatic and manual sequence alignment, inferring phylogenetic trees, mining web-based databases, estimating rates of molecular evolution, inferring ancestral sequences and testing evolutionary hypotheses. Comparative analysis of molecular sequence data is essential for reconstructing the evolutionary histories of species and inferring the nature and extent of selective forces shaping the evolution of genes and species. MEGA is user-friendly software for mining online databases, building sequence alignments and phylogenetic trees and using methods of evolutionary bioinformatics in basic biology, biomedicine and evolution.

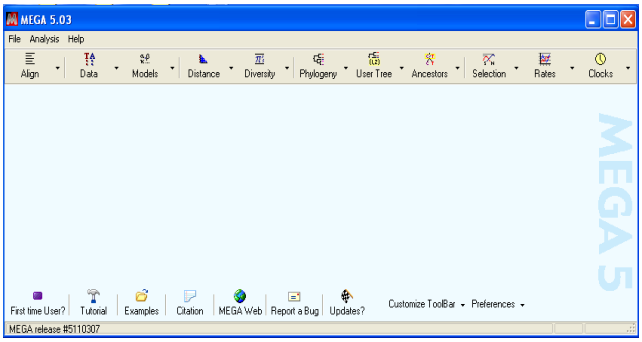

Fig. 5- MEGA home page

ScanProsite

The ScanProsite tool allows scanning protein sequence against the PROSITE database. The user can provide either a UniProt Knowledgebase or Protein Data Bank (PDB) sequence identifier (AC or/and ID) or a sequence in fasta or UniProtKB format. By default the motifs to search for the occurrence are PROSITE patterns and profiles. The ScanProsite tool also allows searching for hits by specific motif in protein sequence database. The motif to search for hits may either be PROSITE pattern and/or profile or provided information. By default, the protein sequence database to be scanned is UniProtKB/Swiss-Prot, including splice variants. Others protein sequence databases may also be scanned, such as UniProtKB/TrEMBL and/or PDB. Adjust search limits by specifying filter and pattern option

It is often useful to be able to search a pattern against a random database in order to evaluate its specificity. It is desirable that the database be not completely random, but comparable to the databases which are to be scanned in terms of amino acid frequency and local compositional bias

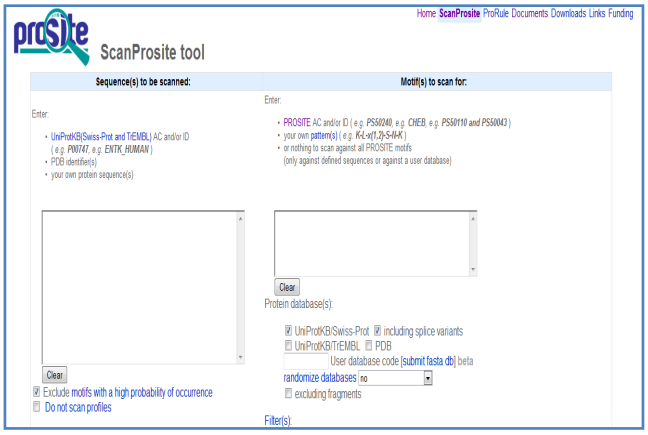

Fig. 6- Home page of scanprosite

The interaction study between TSC1 and Shank family of proteins are possible by different bioinformatics tools. The generated protein networks between TSC1 and Shank family of proteins gives information about the evaloutionary relationships and functional similarities of proteins within the network. This results also helps to understand the coordinated functions of each proteins in the network and it leads to produce new theraputic approaches for autism.

Results

Generating the protein network of TSC1 and Shank family of proteins to specify the exact protein which is responsible for autism involves greater challenges; with the help of bioinformatics tools. The proteins TSC1 and Shank family of are involved in very different autism related syndrome. This helps to known the functional similarity and evolutionary relationships of proteins within the networks and all these findings add knowledge to develop more effective drugs to prevent or cure the autism.

Protein Networking Result Using STRING 9.0:

Networking of TSC1

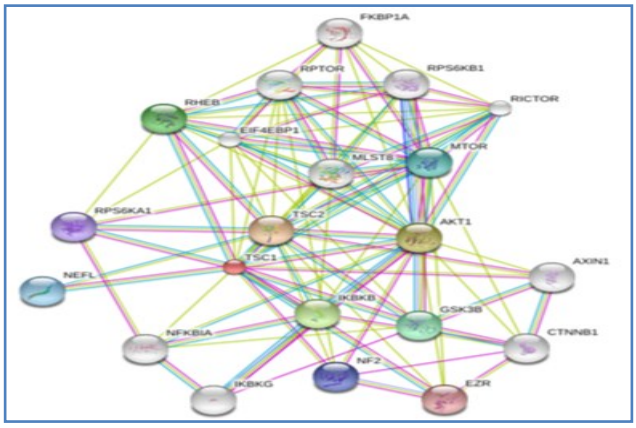

Fig. 7- Interacting proteins of TSC1

In this protein network, the TSC1 protein primarily interacts with other proteins such as TSC2, IKBKB and RHEB.

Networking of SHANK family of proteins

Fig. 8- Interacting proteins of SHANK1 protein

The Shank1 proteins directly interact with LPHN1, DLG4, HOMER1, SSTR2, LPHN2 and BAIAP2 proteins.

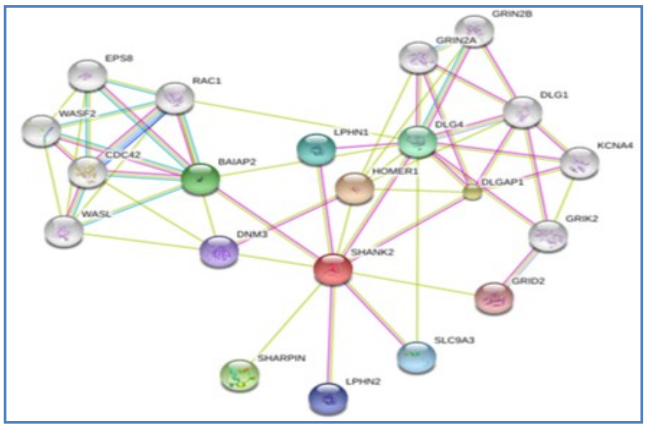

Fig. 9- Interacting proteins of SHANK2 protein

DLG4, LPHN1, HOMER, SLC9A3 and BAIAP2 are the primarily interacting proteins of Shank 2 protein.

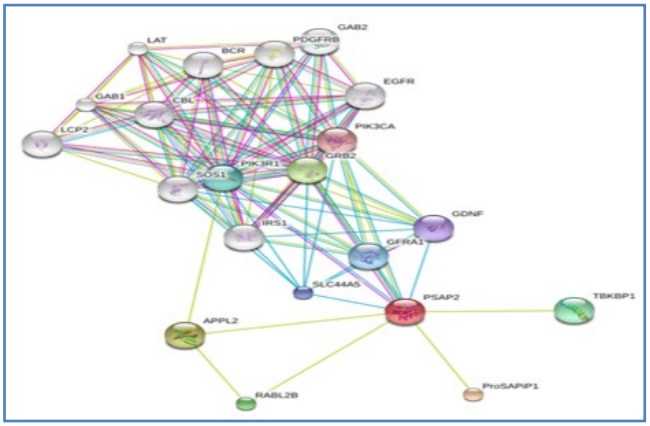

Fig. 10- Interacting proteins of SHANK3 protein

Networking of TSC1 and SHANK family of proteins

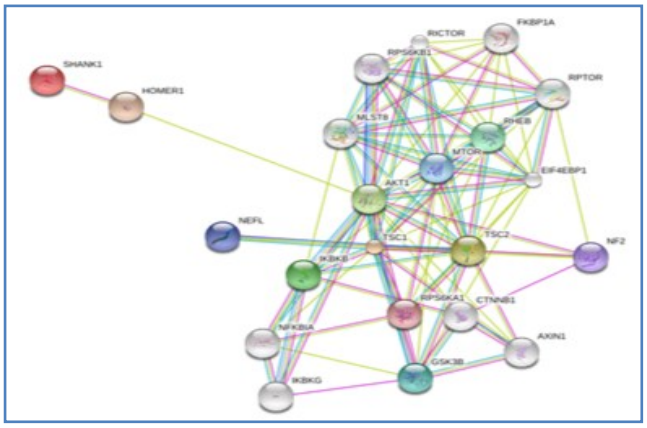

Fig. 11- Networking of TSC1 and SHANK1

The main interacting proteins between TSC1 and Shank1 protein are AKT1 and HOMER1.

The interacting proteins between TSC1 and shank2 are mTOR, NF2, IKBKB, AKT1 and HOMER.

There is no direct interaction between TSC1 and Shank3 (PSAP2) proteins.

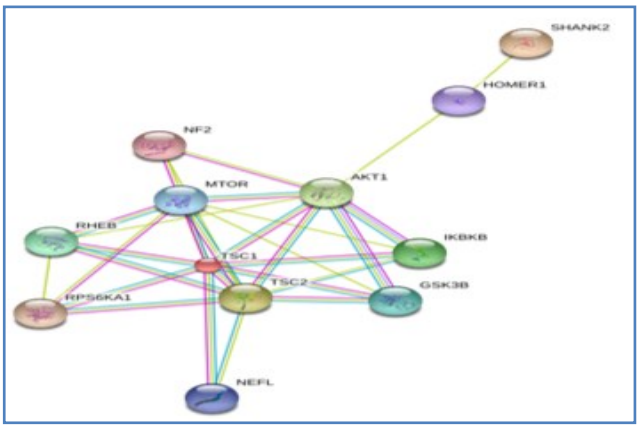

Fig. 12- Networking of TSC1 and SHANK2

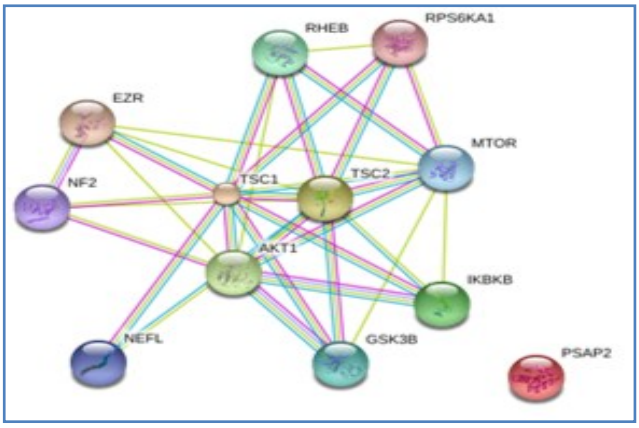

Fig. 13- Networking of TSC1 and SHANK3

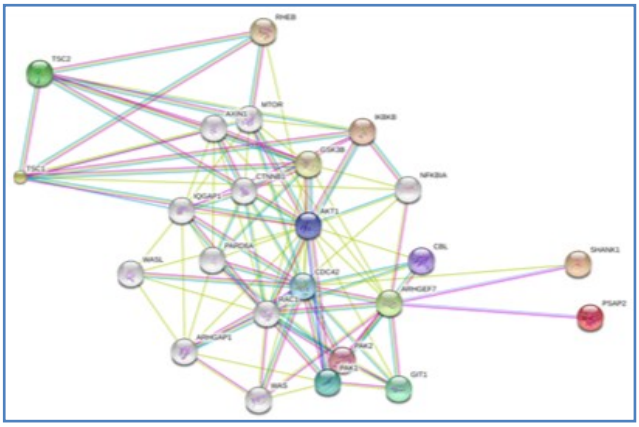

Fig. 14- Networking of TSC1, SHANK1 and SHANK2

The interaction protein between TSC1, Shank1 and Shank2 are TSC2, GSK38, AKT1, CDC42, PAK1 and ARHGEF7.

Complete networking of TSC1 and Shank family of proteins

PSAP2, SHANK1, TSC1, SHANK2 and ARHGEF7 are the input proteins to get complete networking of TSC1 and Shank family of proteins. There are 21 interacting proteins between TSC1 and Shank family of proteins.

The interacting proteins between TSC1 and Shank family of proteins are TSC1, Shank1 and Shank2 are TSC2, GSK38, RHEB, CBL, AKT1, CDC42, PAK1, PAK2 and ARHGEF7.

Using STRING 9.0 analyzed the protein networking between SHANK and TSC1. SHANK and TSC1 are involved in very different autism related syndrome. Both the proteins are not thought to be related, proved to be connected by 21 other proteins. By using STRING 9.0 networking of 21 protein in between the SHANK and TSC1. The main proteins in this networking of TSC1 and SHANK family of proteins are TSC2, RHEB, AXIN1, MTOR, IKBKB, GSK3B, CTNNB1, IQGAP1, WASL, PARD6A, AKT1, NFKBIA, CBL, CDC42, RAC1, ARHGAP1, WAS, PAK2, PAK1, GIT1 and ARHGEF7.

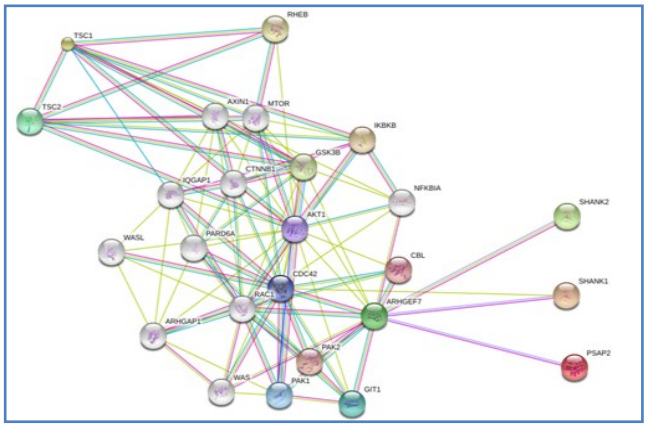

Fig. 15- Networking of TSC1, SHANK1, SHANK2 and SHANK3

Multiple Sequence Alignment Using T-COFFEE

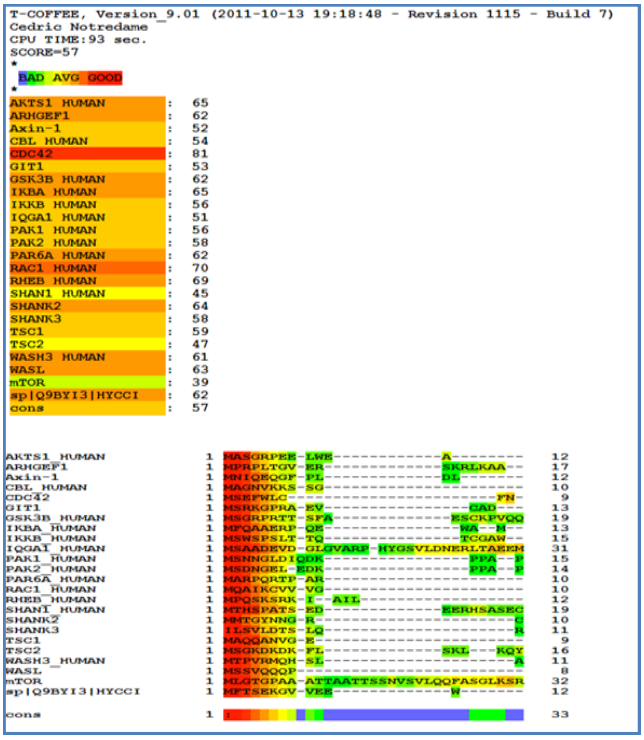

Fig. 16- Scores and conserved regions of interacting proteins

The T-COFFEE shows the conserved regions in the interacting proteins between the TSC1 and Shank family of proteins. The multiple alignments of the proteins show only first few residues shows the conserved region and these regions may responsible for the functional interaction between these proteins.

Evolutionary Tree Analysis using MEGA

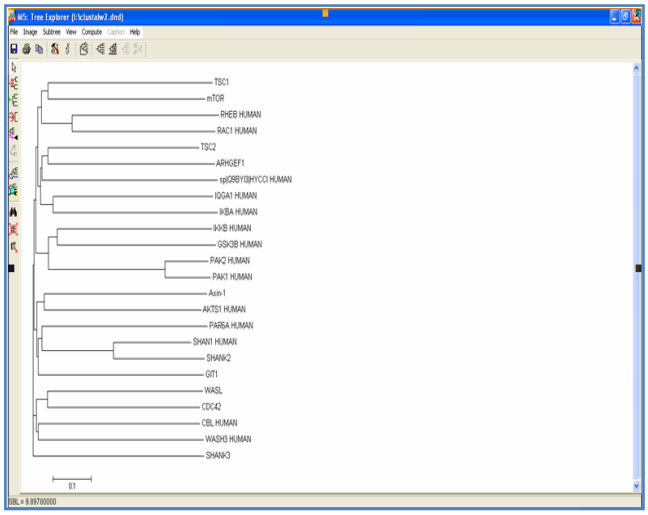

Fig. 17- Dendrogram of interacting proteins

The phylogenetic analysis of the proteins shows that TSC1 and SHANK family of proteins are distantly related. Some of the proteins in the interacting net works are closely related. TSC1 and Shank1 proteins are closely related to mTOR and SHANK2 proteins respectively. TSC1 and SHANK family of proteins are distantly related.

Functional regions using ScanProsite

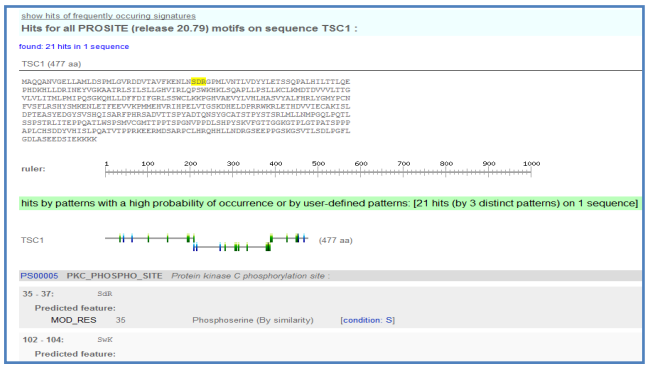

Fig. 18- Domain regions in the TSC1 protein

The domains of TSC1 have no significant region. The few residues of TSC1 are act as the functional effective region.

Fig. 19- Domain regions in the TSC2 protein

In TSC2 the Rap GTPase activating proteins profile is the domain. Rap proteins acts as molecular switches, with an active GTP-bound form and an inactive GDP-bound form

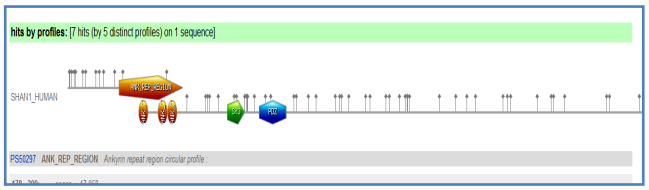

Fig. 20- Domain regions in the SHANK1 protein

In SHANK1 the sterile α motif (SAM) domain, Ankyrin repeats (ANK), PDZ domains are the functional parts.

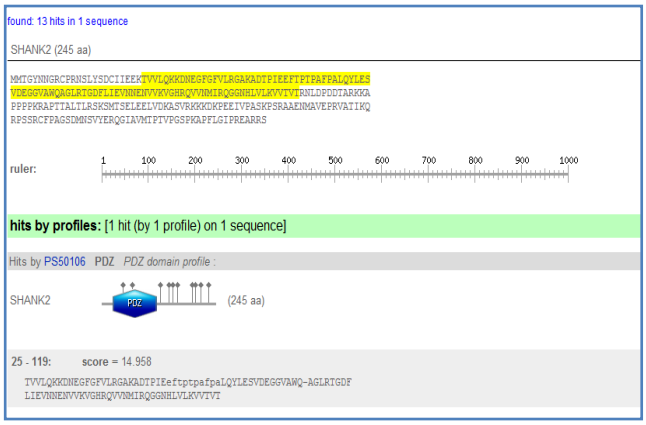

Fig. 21- Domain regions in the SHANK2 protein

In SANK2 the functional part is the PDZ domain.

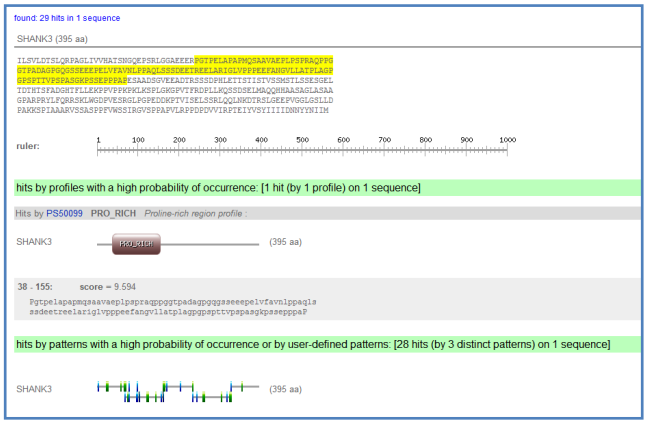

Fig. 22- Domain regions in the SHANK3 protein

In SHANK3 prolin rich profile domain is the functional part, and it has no biologically significant function.

The protein networking of TSC1 and SHANK family of proteins showed 21 interacting proteins. Only first few residues in the proteins of interacting networks showed the conserved regions. The phylogenetic analysis of the proteins showed that TSC1 and SHANK family of proteins are distantly related and the functional part of the domains also found.

Discussion

Proteins are commonly the target of therapeutic drugs, ranging from antimicrobial to anticancer drugs. With the rise of drug re-

sistant and multi-drug resistant forms of many diseases, it has become increasingly important to develop new strategies to identify alternative drug targets. Strong and Eisenberg (2007) reported that protein networks help define individual proteins within the context of all other cellular proteins and suggested methods for the identification and analysis of genome-wide protein networks and how protein networks can be used to aid the identification of novel drug targets [1]. Wu *et al.,* (2010) built a highly reliable functional interaction network upon expert-curated pathways and applied network to the analysis of two genome-wide GBM and several other cancer data sets. The network patterns revealed the results suggested common mechanisms in the cancer biology and system should provide a foundation for a network or pathway-based analysis platform for cancer and other diseases [2].

Using protein interaction networks, the functional similarity between two proteins which involved in different symptoms of autism is identified and this may leads to produce more effective drugs for autism. The proteins which involved in different autism related syndrome are TSC1 and Shank protein. Lin *et al*., (2009) investigated network patterns for breast and colorectal cancers using a similar but smaller data set and predicted that over half of the mutated proteins (59 out of 83) in breast cancers participate in an interaction cluster, but only a very small percentage of mutated proteins in colorectal cancers form an interaction cluster, which contains 12 proteins [3]. In this present study, the protein networking of TSC1 and SHANK family of proteins showed 21 interacting proteins.

Among the 21 interacting proteins, GSK3B, AKT1, CDC42 and ARHGEF were the main interacting proteins.GSK3B is the glycine synthase kinase 3 beta proteins which are involved in the neuronal cell development and body pattern formation. AKT1 is the alpha theronine or serine kinase protein. CDC42 is the cell division control protein 42 which regulate the signaling pathway. ARHFEF 7 is the rho guanine exchange factor 7 which plays an important role in positive regulation of apoptosis. The network view summarizes the network of predicted associations for a particular group of proteins. The other interacting proteins of this network are TSC2, RHEB, AXIN1, MTOR, IKBKB, CTNNB1, IQGAP1, WASL, PARD6A, NFKBIA, CBL, RAC1, WAS, PAK2, PAK1, GIT1 and ARHGH7.

According to Sitbon and Pietrokovski (2007), conserved protein sequence regions are extremely useful for identifying and studying functionally and structurally important regions. By means of an integrated analysis of large-scale protein structure and sequence data, structural features of conserved protein sequence regions were identified and they reported helices and turns were found to be underrepresented in conserved regions, while strands were found to be overrepresented. Similar numbers of loops were found in conserved and random regions [4]. According to Keskin *et al.,* (2004) structurally conserved residues at protein–protein interfaces correlate with the experimental alanine-scanning hot spots, investigated the organization of these conserved, computational hot spots and their contribution to the stability of protein associations and found computational hot spots are not homogeneously distributed along the protein interfaces; rather they are clustered within locally tightly packed regions [5].

The T-COFFEE showed the conserved regions in the interacting protein. Taly *et al.,* (2011) T-Coffee is a versatile multiple sequence alignment (MSA) method suitable for aligning most types of biological sequences. The main strength of T-Coffee is its ability to combine third party aligners and to integrate structural (or homology) information when building MSAs [6]. The multiple alignment of the proteins within the interacting network of TSC1 and Shank family of proteins, only first few residues were showed conserved region and these regions may responsible for the functional interaction between these proteins.

Glaser et al., (2002) reported projecting the conservation grades onto the molecular surface of proteins reveald patches of highly conserved (or occasionally highly variable) residues that are often of important biological function [7]. Tamura *et al.,* (2008) suggested that phylogenetic and molecular evolutionary analysis was conducted using MEGA version 4.0 and he multiple alignments of sequences of CS and MSP-1 proteins were used to create phylogenetic trees. The evolutionary history was inferred using the Neighbour-Joining method [8]. The phylogenetic analysis of the proteins using MEGA showd that TSC1 and SHANK family of proteins are distantly related. Some of the proteins in the interacting net works are closely related. TSC1 and Shank1 are closely related to mTOR and SHANK2 respectively. The SHANK3 and TSC1 proteins were distantly related.

Domains are the functional part of a protein. Royler et al., (2008) reveled, in the case of protein interaction networks, their topology has been explored through the clustering of proteins into groups that share the same biological function, are similarly localized in the cell, or are part of a complex [9]. The functional regions of TSC1 and Shank family of proteins were identified using Scanprosite. Castero *et al.,* (2006) reported PROSITE, through ScanProsite, provides broad intra-domain feature prediction via a flexible context-dependent annotation transfer system. Associating domain detection with an automated annotation system can significantly increase functional predictive power of profiles [10]. The domains of TSC1 have no significant region. The few residues of TSC1 are act as the functional effective region. In TSC2 the Rap GTPase activating proteins profile is the domain. Rap proteins act as molecular switches with an active GTP-bound form and an inactive GDP-bound form. In SHANK1 the sterile α motif (SAM) domain, Ankyrin repeats (ANK), PDZ domains are the functional parts. In SANK3 the functional part is the PDZ domain. In SHANK3 prolin rich profile domain is the functional part and it has no biologically significant function. These are the functional regions of the TSC1 and Shank family of proteins. Protein interaction networks between these proteins helps to find a solution for Autism disorder, opening new avenues for future research with novel prospects and contributing new boon to tackle this unanswered genetic disorder.

References

[1] Strong M. and Eisenberg D. (2007) *Prog Drug Res*., 64, 191, 193-215.

[2] Wu G., Feng X. and Stein L. (2010) *Genome Biology,* 11, R53

[3] Lin J., Gan C.M., Zhang X., Jones S., Sjöblom T., Wood L.D., Parsons D.W., Papadopoulos N., Kinzler K.W., Vogelstein B., Parmigiani G. and Velculescu V.E. (2007) *Genome Res*., 17, 1304-1318.

[4] Sitbon E. and Pietrokovski P. (2007) *BMC Structural Biology*, 7,

3.

[5] Keskina O.B., Mab B. and Nussinovb R.C. (2005) *Journal of Molecular Biology,* 345(5), 1281-1294.

[6] Taly J.F., Magis C., Bussotti G., Chang J.M., Tommaso P.D., Erb I., Espinosa-Carrasco J., Kemena C. and Notredame C. (2011) *Nature Protocols,* 6, 1669-1682.

[7] Glaser F., Pupko T., Paz I., Bell R.E., Bechor-Shental D., Martz E. and Ben-Tal1 N. (2002) *Oxford Journals,* 19(1), 163-164.

[8] Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. (2011) *Oxford Journals,* 28(10), 2731-2739.

[9] Royer L., Reimann M., Andreopoulos B. and Schroeder M. (2008) *PLoS Comput. Biol.,* 4(7), e1000108.

[10] Castro E.I., Christian J.A., Sigrist, Gattiker A., Bulliard V., Petra S. Langendijk-Genevaux, Gasteiger E., Bairoch A. and Nicolas (2006) *Nucleic Acids Res*., 34, W362-W365.