In Silico characterization of STK11- a protein involved in PJ Syndrome

M Reddy, B Bhanu, M Prasad, N Ramya, K Madhulika

Citation

M Reddy, B Bhanu, M Prasad, N Ramya, K Madhulika. *In Silico characterization of STK11- a protein involved in PJ Syndrome*. The Internet Journal of Genomics and Proteomics. 2008 Volume 4 Number 1.

Abstract

Studies of hereditary cancer syndromes have contributed greatly to our understanding of molecular events involved in tumor genesis. One such syndrome of our interest is Peutz-Jeghers Syndrome, a hereditary disease, in which there is predisposition to benign and malignant tumors of many organ systems. Here, we investigate the genes responsible for the Peutz-Jeghers syndrome (PJS) and the pathway of disease causing mechanism. The STK11, Serine/Threonine Kinase 11, commonly known as LKB1 is the gene responsible for this syndrome. This gene is coding a protein of length 433 amino acids, 48.6 kDa. The protein sequence is retrieved from the protein database of NCBI. The analysis of this protein is performed using various in silico methods.

INTRODUCTION

Peutz-Jeghers syndrome is characterized by the development of growths called hamartomatous polyps in the gastrointestinal tract (particularly the stomach and intestines) [1]. Patients with this syndrome have a 15-fold increased risk of developing intestinal cancer compared to that of the general population [2,3]. Almost 50% of patients with Peutz-Jeghers syndrome develop and die from cancer around the age of 57 years. The cumulative risk for developing any cancers associated with Peutz-Jeghers syndrome in patients aged 15–64 years is 93% [4,5]. Peutz-Jeghers syndrome has been described in all races. The occurrence of cases in males and females is about equal. The average age at diagnosis is 23 years in men and 26 years in women. The gastrointestinal polyps found in this syndrome are typical hamartomas. Gastrointestinal polyps can result in chronic bleeding and anemia and cause recurrent obstruction and intussusceptions requiring repeated laparotomies and bowel resections. Mucocutaneous hyper pigmentation appears in childhood as dark blue to dark brown macules around the mouth, eyes, and nostrils, in the perianal area, and on the buccal mucosa [6,7]. The signaling pathway of the STK11 gene product currently is not identified, hence the mechanism of hamartomatous polyp formation and mucocutaneous pigmentation is not known. A locus for this condition (PJS) was assigned to chromosome 19p. The mutations of the STK11 gene mapped to chromosome 19p13.3 are responsible for PJS [8].

Chromosome: 19; Location: 19p13.3

Figure 1

The reference sequence has been curated by NCBI staff. The reference sequence was derived from AC011544.6 and AC004221.2. These sequences are submitted by DOE Joint Genome Institute, Stanford Human Genome Center & DOE Joint Genome Institute, Lawrence Livermore National Laboratory, Livermore, CA respectively.

METHODOLOGY

IDENTIFICATION OF MUTATIONS FROM THE UNIPROT DATABASE

The STK11 gene is a tumor suppressor gene, which means that it normally prevents cells from growing and dividing too rapidly or in an uncontrolled way. The mutations are found to occur in the domain region, disrupting its ability to restrain cell division. Thus, the mutations in the STK11 gene are responsible for the Peutz Jeghers Syndrome. These mutations are identified from the protein information given in Uniprot Database [9] listed in Table 1.

SEQUENCE ANALYSIS OF THE STK11 PROTEIN

The sequence analysis is performed using blastp [10] from

NCBI. Sequences with high similarity are structurally conserved and are functionally related. Hence, the similar protein is identified to analyze if it has any role in the syndrome. Similarity search is performed against the non redundant database taking the expect threshold as 1 and blosum matrix 80. The query protein is compared with various model organisms. The protein having good score, high identities & similarities and fewer gaps is selected from the results of the similarity search in all the model organisms. This resulted in the identification of STK11 proteins in all the organisms selected (Table 2) and also the AMPK (Activated AMP Protein Kinase) as the closest protein to STK11 (Table 3). The AMPK has high sequence similarity and a common domain (kinase) with STK11.

PHYLOGENETIC STUDIES USING BIOLOGY WORKBENCH

Evolutionary studies are performed on the sequences taken from the similarity search results. A multiple sequence alignment is built and the dendrogram is constructed (Figure 1) by Clustal W [11] to find the evolutionary relationship between STK11 and AMPK. The conservations in these proteins are analyzed from the Multiple Sequence Alignment built using both the local (Figures 2&3) and the global alignments (Figure 4) by Texshade [12] and Boxshade respectively. The distance matrix (Figure 5) is constructed using ClustalDist [13] to confirm the results of the dendrogram.

STRUCTURAL COMPARISON OF STK11 & AMPK

The structures of STK11 and the AMPK are analyzed to find the structural similarity. The secondary structure elements, alpha helices, beta sheets and coils are predicted using the tool Hierarchical Neural Network (HNN) [14] from the Expasy proteomics server. The composition of the secondary structure elements are listed in table 5. The 3D structures are retrieved using comparative modeling (CPH Models) [15] and threading methods (3DPSSM [16] server & HHPred [17]). The same structural entry, 2H6D is obtained from both the principles for the two proteins. The cartoon view of the structure is given in Figure 6. Also, the other properties of the protein viz, the molecular weight, atomic composition, iso electric point, half-life, extinction co efficient, Gravy, amino acid composition, instability index and aliphatic index are retrieved using the Protparam [18] tool, the results of which are listed in the table 4.

As the studies reveal good similarity of AMPK with STK11, the KEGG [19] pathway (Figure 7) is obtained to gain more insights into the disease causing mechanism. It is observed that AMPK is involved in the mTOR signaling pathway and that it is activated by its phosphorylation through STK11. Therefore, the phosporylation sites of AMPK are predicted using the tool NetPhos [20]. Also, the Prosite [21] pattern (Table 6), domain (ProDom [22] (Table 7) and family (Pfam [23]) of the proteins (Tables 8 & 9) is retrieved. The protein localization and the solubility are identified from Psort [24] (Table 10) & Sosui [25] (Table 11) respectively.

RESULTS

Figure 2

Table 1: List of Mutations from the Uniprot Database

Natural variations

	Region	Residue Length	
Natural variant	49	1	Y → D in melanoma; sporadic malignant; somatic mutation.
Natural variant	67	1	L → P in PJS.
Natural variant	135	1	G → R in melanoma; sporadic malignant; somatic mutation.
Natural variant	162 –164	3	DGL → NDM in PJS.
Natural variant	163	1	G → D in testicular tumors
Natural variant	194	1	D → N in PJS.
Natural variant	194	1	D → Y in melanoma; sporadic malignant; somatic mutation.
Natural variant	239	1	W → C in PJS; late onset suggests reduced Penetrance.
Natural variant	247	1	Missing in PJS.
Natural variant	297	1	R → K in PJS.
Natural variant	303 – 306	4	IRQH → N in PJS.
Natural variant	315	1	P → S in PJS; pathogenicity uncertain.

The sequence of STK11 is retrieved from the protein database of NCBI. The accession number of the protein is NP_000446.1. The protein is 433aa long. The FASTA format of the protein is given below and the similarity search results are listed in Table 2 & 3.

FASTA FORMAT of the query sequence (STK11):

>gi|4507271|ref|NP_000446.1| serine/threonine protein kinase 11 [Homo sapiens]

MEVVDPQQLGMFTEGELMSVGMDTFIHRIDSTEVIYQ
PRRKRAKLIGKYLMGDLLGEGSYGKVKEVLDSETLC
RRAVKILKKKKLRRIPNGEANVKKEIQLLRRLRHKNV
IQLVDVLYNEEKQKMYMVMEYCVCGMQEMLDSVPE
KRFPVCQAHGYFCQLIDGLEYLHSQGIVHKDIKPGNL
LLTTGGTLKISDLGVAEALHPFAADDTCRTSQGSPAF

QPPEIANGLDTFSGFKVDIWSAGVTLYNITTGLYPFEG
DNIYKLFENIGKGSYAIPGDCGPPLSDLLKGMLEYEPA
KRFSIRQIRQHSWFRKKHPPAEAPVPIPPSPDTKDRWR
SMTVVPYLEDLHGADEDEDLFDIEDDIIYTQDFTVPGQ
VPEEEASHNGQRRGLPKAVCMNGTEAAQLSTKSRAE
GRAPNPARKACSASSKIRRLSACKQQ

Figure 3

Table 2: STK11 Sequences – Results of Similarity Searches

Organism	Accession Number	Bit Score	Identities & Positives	E-Value & Gaps	Length of the subject protein in amino acids
Homo sapiens	ABR45718.1	1033	99% & 99%	0.0 & 0%	433
Mus musculus	NP_035622.1	889	89% & 92%	0.0 & 1%	436
Rattus norvegicus	NP_001101539.1	866	89% & 92%	0.0 & 1%	436
Bos taurus	XP_593229.3	825	82% & 84%	0.0 & 6%	456
Gallus gallus	NP_001039298.1	865	88% & 92%	0.0 & 1%	440
Aedes aegypti	XP_001651729.1	509	67% & 78%	1e-144 & 2%	460
Drosophila melanogaster	NP_650302.1	503	60% & 78%	2e-142 & 1%	567

Figure 4

Table 3: AMPK Sequences – Results of Similarity Searches

Organism	Accession Number	Bit Score	Identities & Positives	E-Value & Gaps	Length of the subject protein in amino acids
Homo sapiens	BAA36547.1	182	37% & 58%	3e-45 & 3%	550
Mus musculus	Q5EG47	181	36% & 58%	6e-45 & 3%	548
Rattus norvegicus	NP_062015.1	181	36% & 58%	3e-45 & 3%	548
Macaca mulatta	XP_001086410.1	181	36% & 58%	2e-45 & 3%	550
Bos taurus	NP_001103272.1	133	37% & 58%	3e-31 & 3%	458
Gallus gallus	NP_001034692.1	184	37% & 58%	1e-46 & 3%	560
Aedes aegypti	XP_001652572.1	175	35% & 56%	4e-44 & 3%	545

The similarity search result has given significant similarity between the query and the AMPK in all the model organisms listed above in the tables. The studies of the proteins in the results indicate that the STK11 protein and the AMPK are having a kinase domain in common.

Figure 5

Figure 1: Dendrogram (Rooted Tree) constructed using Clustal W

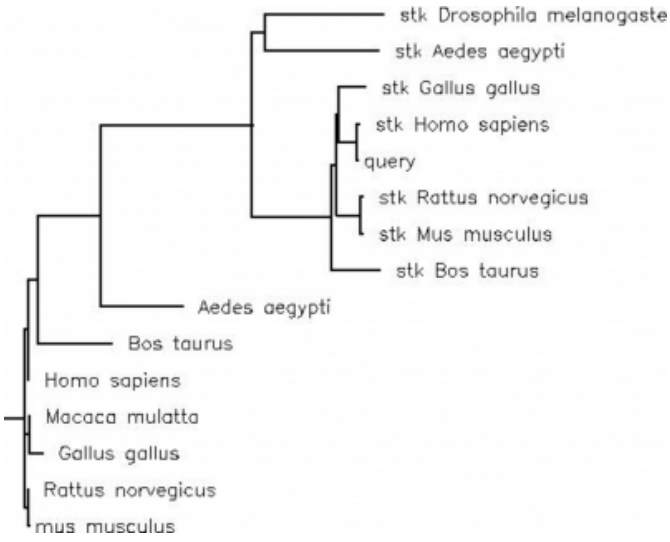


Figure 6

Figure 2: Multiple Sequence Alignment built on local alignment to identify the conserved regions (Pg 1).

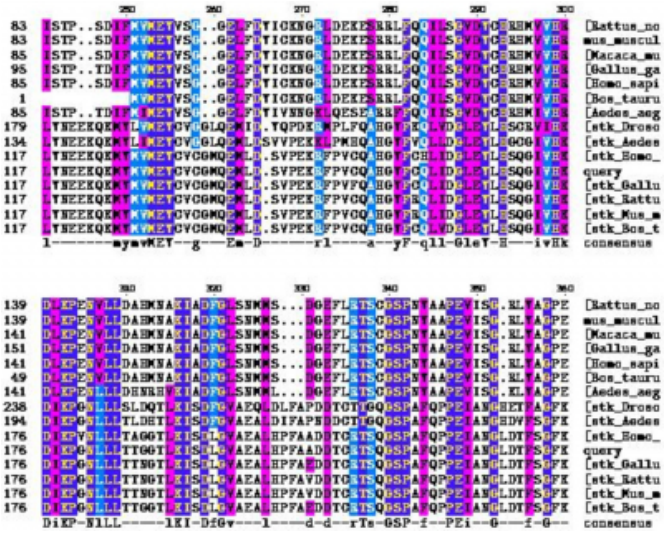


Figure 7

Figure 3: Multiple Sequence Alignment built on local alignment to identify the conserved regions (Pg2).

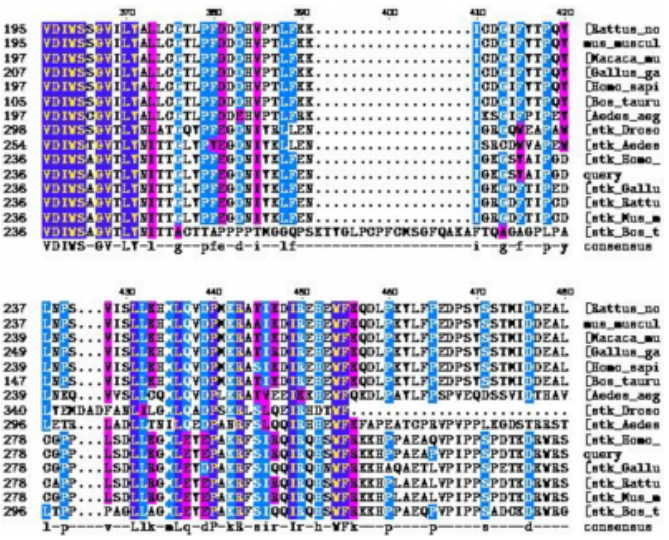


Figure 8

Figure 4: Multiple Sequence Alignment built on global alignment to identify the conserved regions

Violet and green colors are observed dominantly in the results of TEXSHADE and BOXSHADE respectively. These colors indicate the residues conserved in all the species taken.

Figure 9

Figure 5: Clustal Distance Matrix

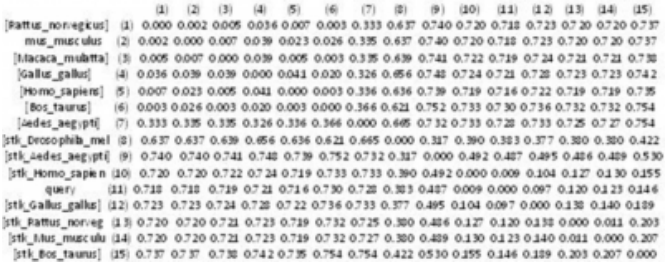


Figure 10

Table 4: Physico-chemical properties of STK11 & AMPK

VALUE FOR STK11	PROPERTY	VALUE FOR AMPK
433	Number of amino acids	550
48635.8	Molecular weight	62793.7
7.12	Theoretical pI	7.32
	Amino acid composition	
6.0%	Ala (A)	4.9%
6.0%	Arg (R)	6.4%
2.8%	Asn (N)	3.3%
6.2%	Asp (D)	7.1%
2.3%	Cys (C)	2.0%
4.6%	Gln (Q)	3.6%
7.2%	Glu (E)	6.0%
7.9%	Gly (G)	4.9%
2.3%	His (H)	4.0%
5.8%	Ile (I)	6.0%
9.0%	**Leu (L)**	**9.5%**
7.4%	Lys (K)	6.7%
3.0%	Met (M)	2.5%
3.2%	Phe (F)	3.5%
6.5%	Pro (P)	5.3%
5.5%	Ser (S)	8.5%
4.4%	Thr (T)	4.4%
0.7%	Trp (W)	0.7%
3.5%	Tyr (Y)	4.0%
5.8%	Val (V)	6.7%
58	Total number of negatively charged residues (Asp + Glu)	72
58	Total number of positively charged residues (Arg + Lys)	72
	Atomic composition	
2151	Carbon C	2785
3420	Hydrogen H	4394
598	Nitrogen N	778
640	Oxygen O	826
23	Sulfur S	25
$C_{2151}H_{3420}N_{598}O_{640}S_{23}$	Formula	$C_{2785}H_{4394}N_{778}O_{826}S_{25}$
6832	Total number of atoms	8808
*39475	#Extinction coefficients	*55405
**38850	#Extinction coefficients	**54780
30 hours	~Estimated half-life	**30 hours**
52.60 **Protein is unstable**	'Instability index	48.23 **Protein is unstable**
80.39	Aliphatic index	**84.69**
-0.463 **Protein is Hydrophilic**	Grand average of hydropathicity (GRAVY)	-0.443 **Protein is Hydrophilic**

Extinction coefficients are in units of M-1 cm-1, at 280 nm measured in water.

* Abs 0.1% (=1 g/l) 0.812, assuming ALL Cys residues appear as half cystines

** Abs 0.1% (=1 g/l) 0.799, assuming NO Cys residues appear as half cystines

~ The N-terminal of the sequence considered is M (Met). The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo), >10 hours (Escherichia coli, in vivo).

` The instability index provides an estimate of the stability of a protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed that there are certain dipeptides, the occurrence of which is significantly different in the unstable proteins compared with those in the stable ones. The authors of this method have assigned a weight value of instability to each of the 400 different dipeptides (DIWV). Using these weight values it is possible to compute an instability index (II) which is defined as:

Figure 11

$$II = (10/L) * \sum_{i=1}^{i=L-1} DIWV(x[i]x[i+1])$$

Where: L is the length of sequence

DIWV(x*x [i+1]) is the instability weight value for the dipeptide starting in position i.*

A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable.

Both these proteins exhibit similar physico-chemical properties, the properties that are similar are highlighted in the table above.

Figure 12

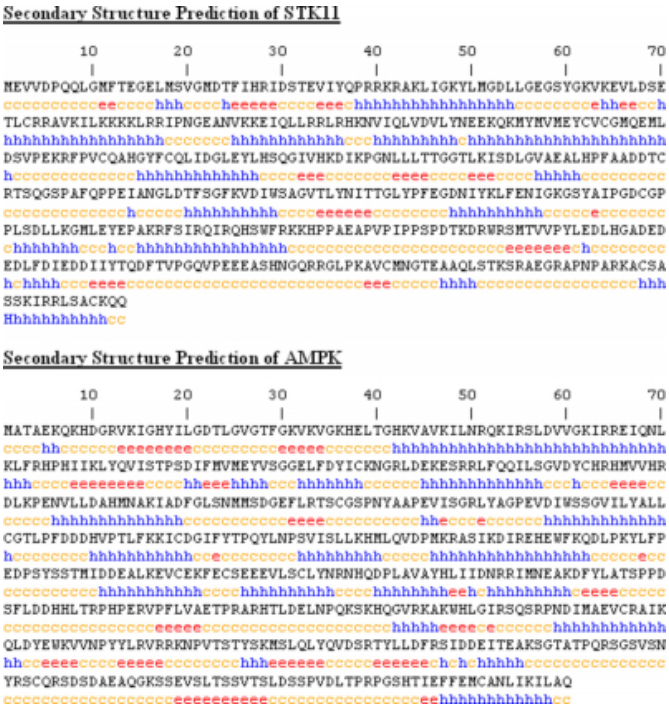


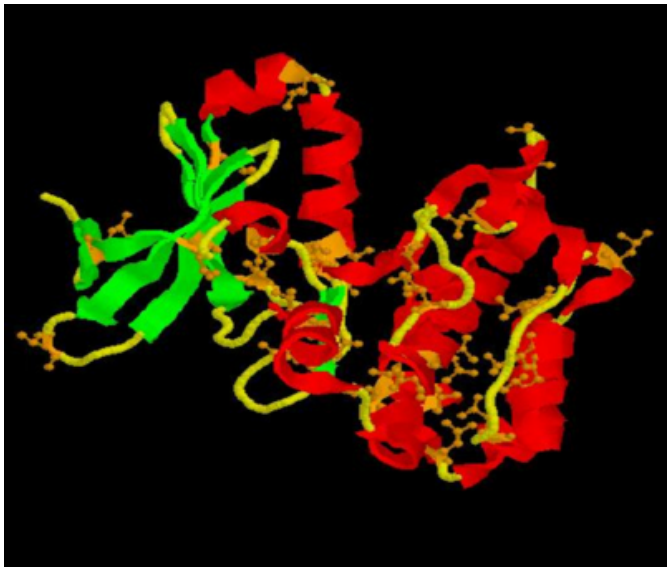
Figure 13

Table 5: %Composition of the Secondary Structure Elements in STK11 & AMPK

SECONDARY STRUCTURE	STK11		AMPK	
	No. of Amino acids	%Composition	No. of Amino acids	%Composition
Alpha helix (Hh)	**171**	**39.49%**	**208**	**37.82%**
3_{10} helix (Gg)	0	0.00%	0	0.00%
Pi helix (Ii)	0	0.00%	0	0.00%
Beta bridge (Bb)	0	0.00%	0	0.00%
Extended strand (Ee)	**44**	**10.16%**	**85**	**15.45%**
Beta turn (Tt)	0	0.00%	0	0.00%
Bend region (Ss)	0	0.00%	0	0.00%
Random coil (Cc)	**218**	**50.35%**	**257**	**46.73%**
Ambiguous states (?)	0	0.00%	0	0.00%
Other states	0	0.00%	0	0.00%

Figure 14

Figure 6: The 3d structure of the PDB entry 2H6D – Rasmol View



The cartoon view of the secondary structural elements, alpha helices, beta sheets & coils are colored red, green & yellow respectively. The amino acid Leucine contributing the highest percent in composition is highlighted using the ball& stick rasmol view (orange color).

Figure 15

Table 6: Prosite Pattern of STK11 & AMPK

PROTEIN_KINASE_ST, PS00108, **Serine/Threonine protein kinases active-site signature** (PATTERN)	
Consensus pattern:	**[LIVMFYC] - x - [HY] - x - D - [LIVMFY] - K - x(2) - N - [LIVMFYCT](3)** D is an active site residue
Sequences known to belong to this class detected by the pattern:	Most serine/ threonine specific protein kinases with 10 exceptions (half of them viral kinases) and also Epstein-Barr virus BGLF4 and Drosophila ninaC which have respectively Ser and Arg instead of the conserved Lys and which are therefore detected by the tyrosine kinase specific pattern described below

The PROSITE pattern observed in the results of these proteins is the signature pattern of the Serine – Threonine protein kinase family. The kinase domain is shared in common for both the proteins. This confirms the hypothesis made in the similarity search.

Figure 16

Table 7: Domain identification using ProDom

RESULT	STK11	AMPK
Entry	PD000001	PD000001
Closest domain	STKB_ HUMAN 55-309	AAK1_ HUMAN 20-270
Number of domains in family	8496	8496
Commentary	KINASE ATP-BINDING TRANSFERASE SERINE/THREONINE-PROTEIN RECEPTOR PHOSPHORYLATION SERINE/THREONINE TYROSINE-PROTEIN	KINASE ATP-BINDING TRANSFERASE SERINE/THREONINE-PROTEIN RECEPTOR PHOSPHORYLATION SERINE/THREONINE TYROSINE-PROTEIN
Length	255	251
Score	1269 (493.4 bits)	1294 (503.1 bits)
Expect	3e-139	3e-142
Identities	91%	95%
Positives	91%	95%

Figure 17

Table 8: Pfam Results of AMPK

Pfam-A	Description	Entry type	Sequence		HMM		Bits score	E-value
			Start	End	From	To		
	Pkinase	Protein kinase domain	Domain	18	270	1	287	345.2
	PPE	PPE family	Family	181	195	1	15	5.4
	UPF0224	Uncharacterised protein family (UPF0224)	Family	307	329	1	27	4.0
	CPSase_L_D3	Carbamoyl-phosphate synthetase large chain, oligomerisation domain	Domain	354	446	1	128	-84.6
	XPA_C	XPA protein C-terminus	Domain	434	456	29	53	6.1

Figure 18

Table 9: Pfam Result for STK11

Pfam-A	Description	Entry type	Sequence		HMM		Bits score	E-value
			Start	End	From	To		
	Pkinase	Protein kinase domain	Domain	49	309	1	287	232.7
	Peptidase_C24	2C endopeptidase (C24) cysteine protease family	Family	275	284	98	107	2.9

Both these proteins belong to a family of PKinase. However, the protein AMPK is a member of many other families as seen in the result. This is because this protein is involved in many metabolic pathways.

Figure 19

Table 10: Identification of Protein Localization using Psort

STK11	AMPK
cyto: 15.0, nucl: 11.0, pero: 3.0	cyto: 24.0, mito: 5.0

These results confirm that these are mostly cytoplasmic proteins.

Figure 20

Table 11: Confirmation of the protein's solubility from Sosui

STK11	AMPK
Total length: **433** A. A. Average of hydrophobicity : **-0.463048** **This amino acid sequence is of a SOLUBLE PROTEIN.**	Total length: **550** A. A. Average of hydrophobicity : **-0.442909** **This amino acid sequence is of a SOLUBLE PROTEIN.**

Figure 21

Figure 7: KEGG Metabolic Pathway

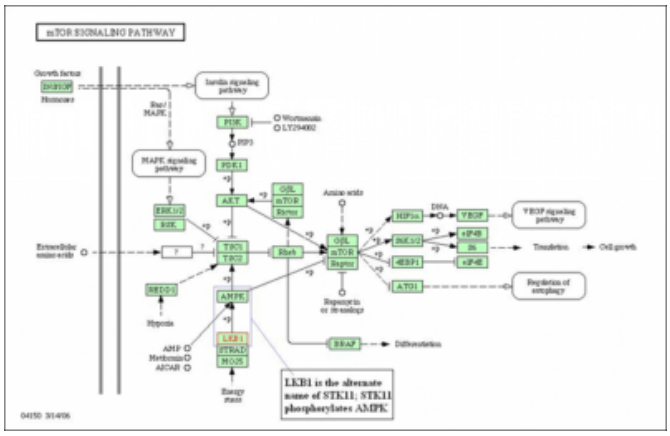
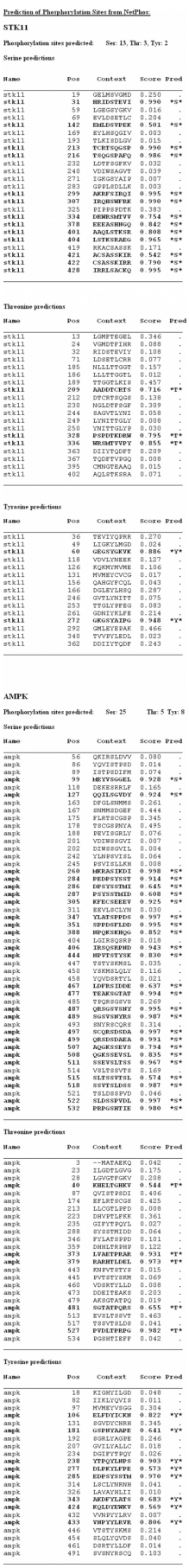


Figure 22



Prediction of Phosphorylation Sites from NetPhos:

STK11

Phosphorylation sites predicted: Ser: 13, Thr: 3, Tyr: 2

Serine predictions

Name	Pos	Context	Score	Pred
stk11	19	GELMSVGMD	0.250	.
stk11	**31**	**HRIDSTEVI**	**0.990**	***S***
stk11	59	LGEGSYGKV	0.016	.
stk11	69	EVLDSETLC	0.204	.
stk11	**142**	**EMLDSVPEK**	**0.501**	***S***
stk11	169	EYLHSQGIV	0.003	.
stk11	193	TLKISDLGV	0.015	.
stk11	**213**	**TCRTSQGSP**	**0.990**	***S***
stk11	**216**	**TSQGSPAFQ**	**0.986**	***S***
stk11	232	LDTPSGFKV	0.032	.
stk11	240	VDIWSAGVT	0.039	.
stk11	271	IGKGRYAIS	0.007	.
stk11	283	GPPLSDLLK	0.003	.
stk11	**299**	**AKEFSIRQI**	**0.995**	***S***
stk11	**307**	**IRQHSWTEK**	**0.990**	***S***
stk11	325	PIPPSPDTK	0.383	.
stk11	**334**	**DNWHSMTVV**	**0.754**	***S***
stk11	**378**	**EEEASHNQQ**	**0.842**	***S***
stk11	**401**	**AAQLSTKSE**	**0.808**	***S***
stk11	**404**	**LSTKSEAEG**	**0.965**	***S***
stk11	419	RKACSASSK	0.171	.
stk11	**421**	**ACSASSKIR**	**0.542**	***S***
stk11	**422**	**CSASSKIRE**	**0.790**	***S***
stk11	**428**	**IRRLSACKQ**	**0.995**	***S***

Threonine predictions

Name	Pos	Context	Score	Pred
stk11	13	LGMPTEGEL	0.346	.
stk11	24	VGMDTFINR	0.088	.
stk11	32	RIDSTEVIY	0.108	.
stk11	71	LDSETLCRR	0.077	.
stk11	185	NLLLTTGGT	0.157	.
stk11	186	LLLTTGGTL	0.012	.
stk11	189	TTGGTLKIS	0.457	.
stk11	**209**	**AADDTCRTS**	**0.716**	***T***
stk11	212	DTCRTSQGS	0.138	.
stk11	230	NGLDTFSGF	0.309	.
stk11	244	SAGVTLYNI	0.058	.
stk11	249	LYNITTGLY	0.008	.
stk11	250	YNITTGLYP	0.030	.
stk11	**328**	**PSPDTKDRW**	**0.795**	***T***
stk11	**336**	**WRSMTVVPY**	**0.855**	***T***
stk11	363	DIYYQDFT	0.209	.
stk11	367	TQDFTVPGQ	0.008	.
stk11	395	CMNGTEAAQ	0.015	.
stk11	402	AQLSTKSRA	0.071	.

Tyrosine predictions

Name	Pos	Context	Score	Pred
stk11	36	TEVIYQPRR	0.270	.
stk11	49	LIGRYLMGD	0.024	.
stk11	**60**	**GEGSYGKVK**	**0.886**	***Y***
stk11	118	VDVLVNEEK	0.127	.
stk11	126	KQHVMVME	0.106	.
stk11	131	MVMEYCVCG	0.017	.
stk11	156	QAHGVFCQL	0.043	.
stk11	166	DGLEYLHSQ	0.287	.
stk11	246	GVTLYNITT	0.075	.
stk11	253	TTGLYPEG	0.083	.
stk11	261	GDNIYKLFE	0.214	.
stk11	**272**	**GKGSYAIPG**	**0.948**	***Y***
stk11	292	GWLSYEAK	0.466	.
stk11	340	TVVPYLEDL	0.023	.
stk11	362	DDIIYTQDF	0.243	.

AMPK

Phosphorylation sites predicted: Ser: 25 Thr: 5 Tyr: 8

Serine predictions

Name	Pos	Context	Score	Pred
ampk	56	QKIRSLDVV	0.080	.
ampk	86	YQVISTPSD	0.014	.
ampk	89	ISTPSDIFM	0.074	.
ampk	**99**	**MEYYSGGEL**	**0.928**	***S***
ampk	118	DSKESRRLF	0.165	.
ampk	**127**	**QQILSGVDY**	**0.924**	***S***
ampk	163	DFGLSNMMS	0.261	.
ampk	167	SNMMSDGEF	0.444	.
ampk	175	FLRTSCGSP	0.345	.
ampk	178	TSCGSPNYA	0.495	.
ampk	188	PEVISGRLY	0.076	.
ampk	201	VDIWSSGVI	0.007	.
ampk	202	DIWSSGVIL	0.004	.
ampk	242	YINPEVISL	0.064	.
ampk	245	PEVISLLKH	0.008	.
ampk	**260**	**MKEASIKDI**	**0.998**	***S***
ampk	**284**	**PEDPSYSST**	**0.914**	***S***
ampk	**286**	**DPSYSSTMI**	**0.645**	***S***
ampk	**287**	**PSYSSTMID**	**0.608**	***S***
ampk	**305**	**KFECSEEEV**	**0.925**	***S***
ampk	311	EEVLSCLYN	0.030	.
ampk	**347**	**YLATSPPDS**	**0.997**	***S***
ampk	**351**	**SPPDSFLDD**	**0.995**	***S***
ampk	**388**	**MPQKSKRQG**	**0.852**	***S***
ampk	404	LGIRSQSRF	0.018	.
ampk	**406**	**IRSQSRPHD**	**0.943**	***S***
ampk	**444**	**HPTSTYSK**	**0.830**	***S***
ampk	447	TSTYSKMSL	0.035	.
ampk	450	YSKMSLQLY	0.116	.
ampk	458	YQVDSRTYL	0.021	.
ampk	**467**	**LDFRSIDDE**	**0.637**	***S***
ampk	**477**	**TEAKSGTAT**	**0.994**	***S***
ampk	485	TPQRSGTVS	0.269	.
ampk	**487**	**QRSGSVSHY**	**0.995**	***S***
ampk	**489**	**SGSVSHYRS**	**0.987**	***S***
ampk	493	SHYRSCQRS	0.314	.
ampk	**497**	**SCQRSDSDA**	**0.997**	***S***
ampk	**499**	**QRSDSDAEA**	**0.993**	***S***
ampk	**507**	**AQGKSSEVS**	**0.794**	***S***
ampk	**508**	**QGKSSEVSL**	**0.835**	***S***
ampk	**511**	**SSEVSLTSS**	**0.967**	***S***
ampk	514	VSLTSSVTS	0.169	.
ampk	**515**	**SLTSSVTSL**	**0.574**	***S***
ampk	**518**	**SSVTSLDSS**	**0.987**	***S***
ampk	521	TSLDSSPVD	0.046	.
ampk	**522**	**SLDSSPVDL**	**0.997**	***S***
ampk	**532**	**PEPGSHTIE**	**0.980**	***S***

Threonine predictions

Name	Pos	Context	Score	Pred
ampk	3	--MATAEKQ	0.042	.
ampk	23	ILGDTLGVG	0.175	.
ampk	28	LGVGTFGKV	0.208	.
ampk	**40**	**KNELTGHKV**	**0.544**	***T***
ampk	87	QVISTPSDI	0.406	.
ampk	174	EFLRTSCGS	0.425	.
ampk	213	LLCGTLPFD	0.008	.
ampk	223	DHVPTLFKK	0.361	.
ampk	235	GIFYTPQYL	0.027	.
ampk	288	SYSSTMIDD	0.064	.
ampk	346	FYLATSPPD	0.101	.
ampk	359	DNHLTRPHP	0.122	.
ampk	**373**	**LYATTPRAR**	**0.931**	***T***
ampk	**379**	**RARHTLDEL**	**0.973**	***T***
ampk	443	KNPYSTYS	0.015	.
ampk	445	PYTSTYSKM	0.069	.
ampk	460	VDSRTYLLD	0.008	.
ampk	473	DDEITEAKS	0.203	.
ampk	479	AKSGTATPQ	0.019	.
ampk	**481**	**SGTATPQRS**	**0.655**	***T***
ampk	513	EVSLTSSVT	0.463	.
ampk	517	TSSVTSLDS	0.041	.
ampk	**527**	**PVDLTPEPG**	**0.982**	***T***
ampk	534	PGSHTIEPF	0.042	.

Tyrosine predictions

Name	Pos	Context	Score	Pred
ampk	18	KIGHYILGD	0.048	.
ampk	82	IIKLVQVIS	0.011	.
ampk	97	MVMEYVSGG	0.304	.
ampk	**106**	**ELFDYICKH**	**0.822**	***Y***
ampk	131	SGVDYCHRH	0.345	.
ampk	**181**	**GSPHYAAPE**	**0.641**	***Y***
ampk	192	SGRLYAGPE	0.246	.
ampk	207	GVILYALLC	0.018	.
ampk	234	DGIFYTPQY	0.026	.
ampk	**238**	**YTPQYLHPS**	**0.903**	***Y***
ampk	**277**	**DLPKYLFPE**	**0.573**	***Y***
ampk	**285**	**EDPSYSSTM**	**0.970**	***Y***
ampk	314	LSCLYNHNH	0.041	.
ampk	326	LAVAYHLII	0.010	.
ampk	**343**	**AKDFYLATS**	**0.683**	***Y***
ampk	**424**	**KQLDYEWKV**	**0.569**	***Y***
ampk	432	VVNPYYLRV	0.007	.
ampk	**433**	**VNPYYLRVA**	**0.806**	***Y***
ampk	446	VTSTYSKMS	0.214	.
ampk	454	SLQLYQVDS	0.040	.
ampk	461	DSRTYLLDF	0.016	.
ampk	491	SVSNYRSCQ	0.103	.

DISCUSSION

The Insilico analysis of the STK11 reveals the role of AMPK in the Peutz-Jeghers syndrome. AMP-activated protein kinase (AMPK) is a primary regulator of the cellular response to lowered ATP levels in eukaryotic cells [26,27]. AMPK is a serine/threonine protein kinase [26].

The pathway also suggests involvement of both these proteins in the same pathway of cell growth. However, AMPK is involved in many other metabolic processes. AMPK acts as a metabolic master switch regulating several intracellular systems including the cellular uptake of glucose, the -oxidation of fatty acids and the biogenesis of glucose transporter 4 (GLUT4). STK11 has been shown to phosphorylate AMPK on Thr-172 and serves as an essential component of physiological AMPK activation. The identified human mutations of STK11 occur in the catalytic domain and cause a loss of its kinase activity. This failure of STK11 phosphorylation impairs its downstream signaling of AMPK [28], thus disturbing the normal cycle of cell growth to abnormal growth leading to the formation of hamartomatous, cancerous polyps, i.e. the Peutz-Jeghers Syndrome. The role of STK11/AMPK in the survival of a subset of genetically defined tumor cells may provide opportunities for cancer therapeutics.

ACKNOWLEDGEMENTS

We would like to thank Mr. G. Ashok Kumar, Director, Ventura Institute of Biosciences, Hyderabad, for providing necessary facilities during the period of the project work.

References

1. Peter Beighton & Gretha Beighton. The Man behind the Syndrome. Springer-Verlag, Berlin-Heidelberg, 1986.
2. Jeghers, H, Mckusick, VA, KATZ, KH.Generalized intestinal polyposis and melanin spots of the oral mucosa, lips and digits; a syndrome of diagnostic significance. N Engl J Med 1949; 241:993.
3. Peutz, JL. Over een zeer merkwaardige, gecombineerde familiaire pollyposis van de sligmliezen van den tractus intestinalis met die van de neuskeelholte en gepaard met eigenaardige pigmentaties van huid-en slijmvliezen. Ned Maandschr v Gen 1921; 10:134.
4. Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, Cruz-Correa M, Offerhaus JA (2000) Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology Med 119:1447-53 [Medline]
5. Giardiello FM, Welsh SB, Hamilton SR, Offerhaus GJ, Gittelsohn AM, Booker SV, Krush AJ, Yardley JH, Luk GD (1987) Increased risk of cancer in the Peutz-Jeghers syndrome. N Engl J Med 316:1511-4 [Medline]
6. Bruwer A, Bargen Ja, Kierland Rr (1954) Surface pigmentation and generalized intestinal polyposis; (Peutz-Jeghers syndrome). Mayo Clin Proc 29:168-71 [Medline]
7. Burke CA, Santisi J, Church J, Levinthal G (2005) the utility of capsule endoscopy small bowel surveillance in patients with polyposis. Am J Gastroenterol 100:1498-502 [Medline]
8. Jenne DE, Reimann H, Nezu J, Friedel W, Loff S, Jeschke R, Müller O, Back W, Zimmer M. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nature Genetics. 1998 Jan; 18(1):38-43.
9. Bairoch A., Apweiler R., Wu C.H., Barker W.C., Boeckmann B., Ferro S.,Gasteiger E., Huang H., Lopez R., Magrane M., Martin M.J.,Natale D.A., O'Donovan C., Redaschi N., YehL.S. The Universal Protein Resource (UniProt), Nucleic Acids Res. 33:D154-D159 (2005).
10. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410. PubMed
11. Thompson J.D., Higgins D.G., Gibson T.J. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res. 22:4673-4680(1994).
12. Beitz, E. (2000), TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e. Bioinformatics 16: 135-139.
13. Thompson J.D., Higgins D.G., Gibson T.J. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res. 22:4673-4680(1994).
14. Combinaison de classifieurs statistiques, Application a la prediction de structure secondaire des proteins, PhD Thesis , Guermeur, Y
15. O. Lund, M. Nielsen, C. Lundegaard, P. Worning, CPHmodels 2.0: X3M a Computer Program to extract 3D Models. Abstract at the CASP5 conferenceA102, 2002.
16. http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html
17. Söding J, Biegert A, and Lupas AN. (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research 33, W244--W248 (Web Server issue).doi:10.1093/nar/gki40.
18. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; Protein Identification and Analysis Tools on the ExPASy Server;(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005) pp. 571-607
19. http://www.genome.jp/kegg/pathway.html
20. Blom, N., Gammeltoft, S., and Brunak, S. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology: 294(5): 1351-1362, 1999.
21. De Castro E., Sigrist C.J.A., Gattiker A., Bulliard V., Langendijk-Genevaux P.S., Gasteiger E., Bairoch A., Hulo N. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 2006 Jul 1; 34(Web Server issue):W362-5.
22. Catherine Bru, Emmanuel Courcelle, Sébastien Carrère, Yoann Beausse, Sandrine Dalmar, and Daniel Kahn (2005) The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res. 33: D212-D215
23. Robert D. Finn, Jaina Mistry, Benjamin Schuster-Böckler, Sam Griffiths-Jones, Volker Hollich, Timo Lassmann, Simon Moxon, Mhairi Marshall, Ajay Khanna, Richard Durbin, Sean R. Eddy, Erik L. L. Sonnhammer and Alex Bateman, Pfam: clans, web tools and services. Nucleic Acids Research (2006) Database Issue 34:D247-D251
24. Paul Horton, Keun-Joon Park, Takeshi Obayashi, Naoya Fujita, Hajime Harada, C.J. Adams-Collier, & Kenta Nakai, "WoLF PSORT: Protein Localization Predictor", Nucleic Acids Research, doi:10.1093/nar/gkm259, 2007.

25. Hirokawa T., Boon-Chieng S., and Mitaku S., SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics, 14 378-9 (1998)
26. Hardie DG, Scott JW, Pan DA, Hudson ER: Management of cellular energy by the AMP-activated protein kinase system. FEBS Lett 2003, 546:113-20.
27. Kemp BE, Stapleton D, Campbell DJ, Chen ZP, Murthy S, Walter M, Gupta A, Adams JJ, Katsis F, van Denderen B, Jennings IG, Iseli T, Michell BJ, Witters LA: AMP-activated protein kinase, super metabolic regulator. Biochem Soc Trans 2003, 31:162-8.
28. Ojuka EO. Role of calcium and AMP kinase in the regulation of mitochondrial biogenesis and GLUT4 levels in muscle. Proc Nutr Soc. 63: 275-278, 2004.

Author Information

M. Vishwanath Reddy, Ph.D.
Department of Bioinformatics, Ventura Institute of Biosciences

B. Divya Bhanu, M.Sc
Department of Bioinformatics, Ventura Institute of Biosciences

MPSNR Prasad, M.tech
Department of Bioinformatics, Ventura Institute of Biosciences

N. Ramya, B.Tech
Department of Bioinformatics, Ventura Institute of Biosciences

K. Madhulika, B.Tech
Department of Bioinformatics, Ventura Institute of Biosciences