JOURNAL OF PROTEINS AND PROTEOMICS
Vol. 1, No. 2, July-December 2010, pp. 77–90

Alternative Conformational Model of a Seed Protein DeK1 for Better Understanding of Structure-Function Relationship

Surya Bhushan Kumar[*, 1], Kavya Venkateswaran[**, 1], and Suman Kundu[*, **, #]

Abstract: *Understanding of the cell development and differentiation processes in plant seeds in general is poor. One gene, among others, that predominantly regulates the aleurone cell formation, differentiation and specification in seeds is called defective kernel (DeK1) and several cell biology and genetic experiments have unequivocally established this fact. However, the mechanism behind such processes is still unclear and understanding the protein functionality of DeK1 is vital to elucidating its role in endosperm cell development. Only preliminary investigations have been performed for just one domain of the protein in vitro and its functional implications have been highlighted lately. An initial attempt at in silico modeling of the protein has shown promise and necessitated thorough investigation of the protein to help understand structure-function relationship in details thus corroborating experimental findings and laying foundation for further studies. DeK1 sequences in public databases were used as raw material for elaborate computational analysis of the protein. DeK1 is a multi-pass membrane protein with interesting structural features and the present analysis provides an alternative model for DeK1 structure that can help pan both in vitro and in vivo studies. The transmembrane helices were shown to have a number of conserved charged and polar residues that can form salt bridges and help in ligand binding or transmitting an external signal in addition to maintaining structural integrity. The protein possesses a big loop of about 280-300 amino acid residues on the cytoplasmic side of the membrane. It has a number of putative phosphorylation sites, multiple cysteine residues and a high density of charged residues, all of which could be important for protein-protein interaction and signaling pathways. The loop has a nuclear localization propensity as well. The long C-terminal tail of DeK1 with homology to calpain domain may be activated by the big loop, or conversely the big loop could be a substrate for the calpain protease in addition to its demonstrated autocatalytic property. Any or all of these features could be important in signaling events and several hypotheses have been forwarded for the structure-function relationship of the novel protein. The results provide a platform for deciphering the biochemical characteristics of DeK1.*

Keywords: *DeK1, defective kernel, seed protein, aleurone layer, big loop, transmembrane topology.*

I INTRODUCTION

A typical plant seed contains an outer layer called the pericarp. The inner endosperm represents the storage tissue of the seed, providing nourishment to the growing seedling at germination. The endosperm is a simple tissue composed of three major cell types - the starchy endosperm which makes up most of the bulk of the seed, the basal transfer layer and the aleurone layer sandwiched between the pericarp and the starchy endosperm (Becraft et al, 2001; Olsen at al, 2008). The transfer layer functions in nutrient uptake from the mother plant during seed development. The starchy endosperm is the major storage site for starch and proteins. While performing some storage function, the primary role of the aleurone is as a digestive tissue. At germination, it secretes amylase into the starchy endosperm causing the breakdown of the stored starch, providing the growing seedlings with sugars for energy and growth. The aleurone layer is also a natural storehouse for oil and fat bodies (Ritchie and Gilroy, 1998). The apparent simplicity of the cellular arrangement, however, is deceptive and endosperm development is a highly specialized process with many unique features. How the endosperm cells specificate into aleurone cells or starchy endosperm cells or transfer cells is yet largely unknown (Olsen, 2001; Berger, 1999; Johnson et al, 2008) and needs to be investigated.

One protein that seems to play crucial role in aleurone and endosperm differentiation is DeK1 (Becraft, 2001; Becraft et al, 2002; Lid et al, 2002; Olsen et al, 2008; Johnson et al, 2008). DeK1 is an acronym for *defective kernel* (Sheridan and Neuffer, 1980; Lid et al, 2002). Absence or deletion of this gene in seeds has resulted in abnormal or defective seed kernels and hence the name (Lid et al, 2005). Though most seeds have a single layer of aleurone cells the presence of multiple layers have been reported in Coroico seeds (Galiant, 1990; Nelson

[*1] School of Biotechnology, Banaras Hindu University, Varanasi 221005, India.
[**1] Department of Biochemistry, University of Delhi South Campus, Benito Juarez Road, New Delhi-110021, India,
[#] Corresponding author: E-mail: suman.kundu@south.du.ac.in.

and Chang, 1974). This indicates scope for engineering multiple layers of aleurone cells in seeds provided we understand the genes that control aleurone cell fate specification. Investigating DeK1 in detail could provide an initiation in this direction.

DeK1 is distributed all over the endosperm cells. However, the greatest effect of this gene in cell specification is observed in the aleurone cells (Lid et al, 2002; Becraft et al, 2002). Mutants that have disrupted aleurone development have been identified. Such mutants lack an aleurone layer (Becraft and Asuncion-Crabb, 2000; Lid et al, 2005). This is further exemplified in a gain-of-function experiment where overexpression of the gene results in an extra layer of aleurone (Becraft and Asuncion-Crabb, 2000; Lid et al, 2005). It is thus evident that aleurone cell formation is a highly specialized event and that endosperm cells, which are only on the surface, get differentiated into aleurone cells (Gifford et al, 2003). DeK1 plays a critical role in this developmental process. Further investigation into such plant proteins and genes could help improve our poor understanding of cell development and differentiation in general.

While genetic analyses and cell biology studies have highlighted the importance of DeK1 gene (Becraft and Asuncian-Crabb, 2000; Becraft et al, 2002; Lid et al, 2002; Lid et al, 2005; Ahn et al, 2004; Johnson et al, 2005), there is minimal information at the protein level. A portion of the protein was found to have homology to animal calpains and the protease activity of this domain in isolation was demonstrated experimentally (Wang et al, 2003). It was recently shown that this protease domain can complement DeK1 function in mutants through autocatalysis (Johnson et al, 2008). Successful understanding of the regulation of aleurone cell formation by DeK1 needs complimentary knowledge about the total DeK1 protein as well. It is vital to know how this protein is synthesized, degraded and transported, which proteins it interacts with, how it triggers the signaling events, how it senses the presence of endosperm cells on the surface directing aleurone formation, what ligands and substrate it would bind to, what is the nature and biochemical characteristics of the protein, what kind of secondary structure and three-dimensional structure it forms, how are structural elements related to its biological function and finally how the protein can be engineered to manipulate its function. The fact that DeK1 is linked to distinct phenotype (defective kernel, absence of aleurone layer) makes it a viable candidate for such investigation. The only attempt at visualizing the protein structure (Lid et al., 2002) predicted one model for DeK1 protein. However, it is unable to answer several questions and the topology of membrane helices and the orientation and role of a big loop is in doubt. Alternative topologies and possibilities need to be explored.

As a preliminary initiative to our objective, the present investigation deals with "*in silico*" modelling of DeK1 protein. We have analysed Dek1 sequences from three plant species namely Rice *(Oryza sativa*), Maize *(Zea mays*), and Arabidopsis *(Arabidopsis thaliana*), the former two being monocotyledonous and the latter being dicotyledonous. The results help us to visualize a tentative secondary structure of the protein and help us to formulate hypothesis regarding the structure-function relationship of the protein. The studies build a basis for detailed research into the DeK1 system.

II MATERIALS AND METHODS

All results were obtained computationally. PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) and popular Internet search engines were used for literature survey. The DNA and amino acid sequences of the target protein DeK1 from different plant sources, as well as all other proteins used as controls or for reference, were obtained either from NCBI (http://www.ncbi.nih.gov) or EMBL-EBI (http://www.ebi.ac.uk/). These sequences were downloaded and analyzed in detail. The individual amino acid sequence file, in each case, was converted to BLAST or FASTA format.

The theoretical physical properties of the DeK1 proteins were calculated using Proteomics tools from the Expasy home page (www.expasy.org). CLUSTALW (http://www.ebi.ac.uk/clustalw/) from Expasy home page was used for sequence alignment (Thompson et al, 1994; Jeanmougin et al, 1998). PSORT (http://www.psort.org/) provided links to the PSORT family of programs for subcellular localization prediction as well as other datasets and resources relevant to localization prediction. The following PSORT programs for localization prediction were used: WoLF PSORT is a recently updated version of PSORT II for the prediction of eukaryotic sequences; PSORT II (Nakai and Horton, 1999; Horton and Nakai, 1997) for eukaryotic sequences; PSORT (Nakai and Kanehisa, 1991) for plant sequences. SIGNALP 3.0 server (www.cbs.dtu.dk/services/SignalP) was used to predict the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms (Bendtsen et al, 2004).

The following algorithms were used to predict the membrane topology of DeK1 proteins:

TMHMM (www.cbs.dtu.dk/service/TMHMM) (Krogh et al, 2001)

SOSUI (http://sosui.proteome.bio.tuat.ac.jp/sosui_submit.html) (Hirokawa et al, 1998; Mitaku and Hirokawa, 1999; Mitaku et al, 1998)

HMMTOP (http://www.enzim.hu/hmmtop/) (Tusnady and Simon, 1998)

TMPRED (http://www.ch.embnet.org/software/TMPRED_form.html) (Hofmann and Stoffel, 1993; Arai et al, 2004)

TOPPRED (www.pasteure.fr/seqanal/interfaces/toppred) (Stoffel et al, 1993)

TMAP (http://bioinfo.limbo.ifm.liu.se/tmap/) (David et al, 2002; Persson and Argos, 1996; Persson and Argos, 1994)

PHOBIUS (http://phobius.cgb.ki.se/and http://phobius.binf.ku.dk/) (Kall et al, 2004).

Of these TMHMM had earlier been reported to be the best (Melen et al, 2003; Moller et al, 2001). However, PHOBIUS is the newest algorithm available (2004) and seems to have certain advantages over the other as seen in the Results section.

The NetPhos 2.0 server (http://www.cbs.dtu.dk/services/NetPhos/) was used to predict serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins (Blom et al, 1999). Nuclear localization signal (NLS) was predicted using a software from the website mentioned (http://cubic.bioc.columbia.edu/PredictNLS) (Cokol et al, 2000).

III RESULTS AND DISCUSSION

DeK1 Sequences

A search of public databases for DeK1 amino acid sequences revealed that such sequences from four plant species are available. For the present investigation, DeK1 sequences of Rice *(Oryza sativa*), Maize *(Zea mays*) and *Arabidopsis (Arabidopsis thaliana*) plants were used, with NCBI accession numbers Q8RYA5, Q8RUQ1 (Q8RVL1) and Q8RVL2, respectively. While rice and maize are monocotyledonous plants, *Arabidopsis* is a dicotyledonous plant. Hence an analysis of these three sequences would be interesting for comparative purposes. Evolutionarily conserved and functionally important residues can be extracted from such analysis. The consistency of *in silico* analysis across species and across monocots and dicots would make the predictions and models more reliable. Arab DeK1 has 2151 amino acid side chains, while maize DeK1 has 2159 and rice DeK1 has 2162 amino acid side chains. As a prelude to experimental studies we have resorted to computational methods here to investigate DeK1 structure. The three sequences were thus subjected to various prediction algorithms.

Predicted Physical Parameters of DeK1 Proteins

There is minimal information about the biochemical characteristics of DeK1 protein, since the full length protein has not been studied yet. Since biochemical and physical properties define a protein, it is interesting to know some of these properties and whether they differ between DeK1 proteins from different species of plants. Thus, the molecular mass (M_w) of these proteins are predicted to be 240, 751 Da for Rice DeK1, 238, 865 Da for *Arabidopsis* DeK1 and 239, 945 Da for Maize DeK1. Their isoelectric points (pI) are 5.85, 6.15 and 5.69, respectively. These results indicate that the DeK1 proteins would have similar physical properties.

Sequence Alignment of DeK1's

Sequences of the three DeK1's were aligned by CLUSTALW to check the overall identity and similarity of the sequences (data not shown). Rice and Maize DeK1 have 90% identity, while each of these sequences are only 70% similar to *Arabidopsis* DeK1. This is expected since while rice and maize are monocotyledonous plants, *Arabidopsis* is dicotyledonous. The high similarity in sequences indicates that DeK1 performs an important function and has been conserved evolutionarily. It also indicates that DeK1 proteins probably have a similar protein fold, three-dimensional structure and consensus functional conformation.

Subcellular Localization

The probability of finding a protein in a particular location of the cell, or a particular organelle of the cell can be predicted using its sequence information. For all the three DeK1 sequences the prediction algorithms (PSORT) produced the same results:

56.0%: extracellular, including cell wall; 12.0%: cytoplasmic; 12.0%: endoplasmic reticulum; 8.0%: vesicles of secretory system; 8.0%: nuclear; 4.0%: mitochondrial.

The prediction indicates that DeK1 is majorly an extracellular protein and this has already been observed experimentally (Johnson et al, 2008), raising confidence in *in silico* analysis of DeK1 sequences. So DeK1 might either be a membrane protein and docked in the cell wall or secreted out of the cytoplasm as a soluble extracytoplasmic protein. In either case, DeK1 would have a signal peptide or a stretch of sequence that leads the protein to its specific location of the cell.

Hemoglobin from *Pseudoterranova decipiens* (codworm), which has been investigated experimentally (Dixon et al, 1991; Gibson et al, 1993), was used as a control protein to test the accuracy of PSORT algorithm. Keeping with experimental observations that the nematode hemoglobin is an extracellular protein, the following results were obtained: 48.0%: extracellular, including cell wall; 16.0%: cytoplasmic; 12.0%: nuclear; 8.0%: vesicles of secretory system; 8.0%: endoplasmic reticulum; 4.0%: vacuolar; 4.0%: mitochondrial.

Signal Peptide Presence and Location Prediction

Signal peptide is a stretch of amino acid sequence at the N-terminus of a protein that targets the protein or helps to transport (or translocate) the protein after synthesis to its site of action. Once the protein reaches its proper location, the signal peptide is cleaved off and the mature protein performs its biological function. If the signal peptide is not cleaved then it is called a signal anchor. Previously, signal anchor was called "uncleaved signal peptide". The presence, if any, of signal peptides in the three DeK1 proteins were predicted by SignalP algorithm. A typical result for Rice DeK1 protein is shown in Table 1.

The software for prediction picks up the first seventy residues at the N-terminus. The output is given in two parts. In the first part of Table 1 (left column), scores are provided for residues with the highest probability of being cleaved. In the second part, the final prediction is given and the overall probabilities are mentioned. From the above results, it is clear that the probability of the presence of a signal peptide in rice DeK1 is very poor (0.2% only). However, this protein has a signal anchor sequence at the N-terminus (probability = 99.7%). This sequence can help the DeK1 protein to hold on to its target location, probably the cell membrane. Similar results were obtained for the other two DeK1 proteins. This means that DeK1 has a signal sequence at the N-terminus that helps the protein to anchor but is not cleaved. So DeK1 is probably not a soluble extracytoplasmic protein but an intrinsic membrane protein as suggested earlier (Lid et al, 2002; Becraft et al, 2002). As such, while building structural model for DeK1 proteins, we have included the amino acid sequence at the N-terminus as well in topology prediction.

Table 1: Prediction of the Presence and Location of Signal Peptide in Rice DeK1 Amino Acid Sequence.

SignalP-NN result:

>RiceDeK1 *length = 70*

Measure	*Position*	*Value*	*Cutoff*	*Signal peptide?*
max. C	34	0.284	0.32	NO
max. Y	34	0.325	0.33	NO
max. S	19	0.981	0.87	YES
mean S	1-33	0.593	0.48	YES
D	1-33	0.459	0.43	YES

Most likely cleavage site between pos. 33 and 34: LWA-VN.

SignalP-HMM result:

>RiceDeK1

Prediction:	*Signal anchor*
Signal peptide probability:	0.002
Signal anchor probability:	0.997
Max cleavage site probability:	0.001 between pos. 33 and 34

The strength and accuracy of this prediction algorithm was tested with a protein sequence for which the presence of signal peptide has been established experimentally. Hemoglobin from *Pseudoterranova decipiens* has been shown to contain 333 amino acids of which the first 18 amino acids form the signal sequence which finally gets cleaved resulting in a mature peptide from residue 19-333 (Dixon et al, 1991; Gibson et al, 1993). SignalP predicts this result with accuracy as shown below.

SignalP-NN result for hemoglobin from *P. decipiens*:

>Sequence length = 70

Measure	Position	Value	Cutoff	Signal peptide?
max. C	19	0.753	0.32	YES
max. Y	19	0.781	0.33	YES
max. S	13	0.995	0.87	YES
mean S	1-18	0.916	0.48	YES
D	1-18	0.849	0.43	YES

Most likely cleavage site between pos. 18 and 19: ASA-SK

SignalP-HMM result for hemoglobin from *P. decipiens*: >Sequence

Prediction: Signal peptide

Signal peptide probability: 0.999; Signal anchor probability: 0.000

Max cleavage site probability: 0.868 between pos. 18 and 19

Unlike the DeK1 sequences, the hemoglobin sequence has a signal peptide probability of 99.9% and conforms to the experimental results. Similarly, another protein lactose permease has experimentally been shown not to have either a signal peptide or a signal anchor sequence (Fortina et al, 2003). The SignalP predictions yielded results of 9.6% probability for signal peptide and 22.6% probability for signal anchor. These numbers are not significant indicating the absence of signal sequence in the protein, as per the experimental results.

Membrane Topology Prediction

The subcellular localization prediction of DeK1 and previous literature indicates that DeK1 is a membrane protein (Lid et al, 2002; Wang et al, 2003). To understand the structure-function relationship in membrane proteins it is important to know how the protein spans the membrane. The number of transmembrane domains, their topology or orientation, the number and orientation of loops, the orientation of the N-terminal and C-terminal domains and other related structural information regarding membrane proteins are important. Lid et al. had modeled DeK1 transmembrane domain based on only one topology prediction algorithm, TMHMM (Lid et al, 2002). *In silico* predictions should be based on multiple algorithms since each algorithm has its own limitations. Using multiple algorithms enhance the confidence in prediction, especially if the sequence information is robust and produces consensus across different algorithms. Moreover, such an extensive exercise allows prediction of alternate possibilities for structural models and corroboration of experimental findings become easier. Lid et al's model also fails to explain facts like the presence of a big loop on the extracytoplasmic side while it has a large number of charged residues and probability for phosphorylation and therefore more likely to be cytoplasmic. Further, the model was not analyzed thoroughly for structure-function relationship in relation to its role in aleurone specification. Hence, we have revisited DeK1 model and investigated alternative conformation for DeK1 to provide a boost to in solution studies. A number of membrane topology prediction algorithms, as mentioned in the Materials and Methods section, were used to predict the transmembrane domains (TM). These topology prediction algorithms predict the number and position of transmembrane helices (TMH).

A typical prediction result for rice DeK1 using Phobius as the prediction algorithm is shown in Table 2. The predictor identifies the regions which can form transmembrane helices, e.g. amino acid residues 6-34, 68-90, 96-119, etc (denoted as TRANSMEM in the above results). This then delineates the non-membrane portions of the protein. The non-membrane portions called loops are predicted to be on the cytoplasmic or non-cytoplasmic side. The N-terminus, which for rice DeK1, is from amino acid 1 to amino acid 5 (above) is on the "outside" (non-cytoplasmic) of the membrane, while the long C-terminus tail from amino acid side chain number 1108 to 2162 is on the "inside" of the membrane. Thus rice DeK1 has 23 transmembrane helices, 11 loops on the outside and 11 loops on the inside. One of the loops, we call "Big Loop" is exceptionally long in length (~280 residues) and is on the cytoplasmic side or inside of the membrane, unlike proposed by Lid et al (2002).

Table 2: Topology Prediction of RiceDeK1 using PHOBIUS.

ID	*RiceDEK1*			
FT	DOMAIN	1	5	NON CYTOPLASMIC.
FT	TRANSMEM	6	34	
FT	DOMAIN	35	67	CYTOPLASMIC.
FT	TRANSMEM	68	90	
FT	DOMAIN	91	95	NON CYTOPLASMIC.
FT	TRANSMEM	96	119	
FT	DOMAIN	120	130	CYTOPLASMIC.
FT	TRANSMEM	131	152	
FT	DOMAIN	153	163	NON CYTOPLASMIC.
FT	TRANSMEM	164	186	
FT	DOMAIN	187	233	CYTOPLASMIC.
FT	TRANSMEM	234	258	
FT	DOMAIN	259	263	NON CYTOPLASMIC.
FT	TRANSMEM	264	282	
FT	DOMAIN	283	293	CYTOPLASMIC.
FT	TRANSMEM	294	315	
FT	DOMAIN	316	320	NON CYTOPLASMIC.
FT	TRANSMEM	321	341	
FT	DOMAIN	342	626	CYTOPLASMIC.
FT	TRANSMEM	627	646	
FT	DOMAIN	647	665	NON CYTOPLASMIC.
FT	TRANSMEM	666	687	
FT	DOMAIN	688	698	CYTOPLASMIC.
FT	TRANSMEM	699	717	
FT	DOMAIN	718	722	NON CYTOPLASMIC.
FT	TRANSMEM	723	744	
FT	DOMAIN	745	773	CYTOPLASMIC.
FT	TRANSMEM	774	794	
FT	DOMAIN	795	822	NON CYTOPLASMIC.
FT	TRANSMEM	823	849	
FT	DOMAIN	850	855	CYTOPLASMIC.
FT	TRANSMEM	856	877	
FT	DOMAIN	878	888	NON CYTOPLASMIC.
FT	TRANSMEM	889	912	
FT	DOMAIN	913	923	CYTOPLASMIC.
FT	TRANSMEM	924	944	
FT	DOMAIN	945	949	NON CYTOPLASMIC.
FT	TRANSMEM	950	973	
FT	DOMAIN	974	984	CYTOPLASMIC.
FT	TRANSMEM	985	1005	
FT	DOMAIN	1006	1016	NON CYTOPLASMIC.
FT	TRANSMEM	1017	1039	
FT	DOMAIN	1040	1059	CYTOPLASMIC.
FT	TRANSMEM	1060	1083	
FT	DOMAIN	1084	1088	NON CYTOPLASMIC.
FT	TRANSMEM	1089	1107	
FT	DOMAIN	1108	2162	CYTOPLASMIC.

The results for *Arabidopsis* and maize DeK1 TMH prediction were also obtained (data not shown) and are similar to those for rice DeK1. Other predictors were used as well for comparative purposes. The results are however, not shown individually but summarized in Table 3 and analyzed in the next section. At this point it would be worthwhile to mention that depending on just one algorithm to model a protein can often be misleading and leave us unaware of all the possibilities. Just to drive the point we show here how TMHMM can lead to wrong prediction for proteins. We have seen earlier that hemoglobin from *P. decipiens* has a signal peptide from amino acid side chain 1-18. However, TMHMM cannot predict the signal peptide and thus considers the whole sequence (1-333) as a mature protein unlike the experimental observations. For a multi-pass protein like DeK1 with a signal peptide such a shortcoming would throw off the entire prediction since the correct prediction of the N-terminus is very essential. SOSUI, on the other hand, does recognize the N-terminal signal peptide region but considers it as a TMH thus predicting the hemoglobin to be a membrane protein, which is not actually the case. PHOBIUS actually predicts amino acids 1-18 to be the signal peptide and thus renders the correct prediction. So using three predictors improves our prediction capabilities, and justifies our use of multiple predictors. As seen in the model for DeK1 proposed by Lid et al. based only on TMHMM, the C-terminal tail was considered to be cytoplasmic to fit in with experimental observations even though the algorithm predicted it to be non-cytoplasmic. The big loop, on the other hand, was considered to be non-cytoplasmic against prediction since it seemed that big loop and the C-terminal tail would have opposite orientations. Here, we observe that certain algorithms like SOSUI, PHOBIUS, etc. do indeed predict C-terminal to be cytoplasmic in accordance with experimental observations, and that it is possible that the big loop and C-terminal share the same orientation. Thus, some inherent problem in prediction of signal peptide by TMHMM is obvious and using multiple algorithms improves the prediction.

Table 3: Summary of TMH Prediction for Rice, *Arabidopsis* and Maize DeK1 using a Number of Prediction Tools.

Arabidopsis DeK1	*No. of TMHs*	*No. of big loops*	*Big loop Orientation*	*N-Terminal Orientation*	*C-Terminal Orientation*
TMHMM	23	1	Inside	Inside	Outside
HMMTOP	24	2	Inside	Outside	Outside
SOSUI	23	1	Inside	Outside	Inside
TMPRED	23	2	Inside	Inside	Outside
TMAP	23	2	Outside	Outside	Inside
TOPPRED	23	2	Outside	Outside	Inside
PHOBIUS	23	1	Inside	Outside	Inside
Maize DeK1	*No. of TMHs*	*No. of big loops*	*Big loop Orientation*	*N-Terminal Orientation*	*C-Terminal Orientation*
TMHMM	22	1	Inside	Outside	Outside
HMMTOP	24	1	Inside	Outside	Outside
SOSUI	23	1	Inside	Outside	Inside
TMPRED	23	1	Outside	Inside	Outside
TMAP	24	1	Outside	Outside	Outside
TOPPRED	24	2	Outside	Inside	Inside
PHOBIUS	23	1	Inside	Outside	Inside
Rice DeK1	*No. of TMHs*	*No. of big loops*	*Big loop Orientation*	*N-Terminal Orientation*	*C-Terminal Orientation*
TMHMM	22	1	Inside	Outside	Outside
HMMTOP	26	1	Inside	Outside	Inside
SOSUI	23	1	Inside	Outside	Inside
TMPRED	24	2	Inside	Outside	Outside
TMAP	23	2	Outside	Outside	Inside
TOPPRED	25	2	Outside	Outside	Inside
PHOBIUS	23	1	Inside	Outside	Inside

Another observation that gains prominence is the constant success of the tool PHOBIUS in predicting different parameters, atleast for DeK1. To verify this observation further, we used this tool for other well-studied multi-pass membrane proteins like lactose permease and non-membrane proteins like hemoglobin from *P. decipiens*. The predictions of TMH in the first case (data not shown) tally closely with experimental findings (Fortina et al, 2003). Similar results from other predictors do not agree as well as Phobius does with experimental findings (Kundu et al, Personal communications). Phobius is thus a reliable predictor of transmembrane topology. This tool also worked well for the non-membrane proteins in predicting accurately their signal peptides (data not shown).

The Model

Inspecting the membrane topology predictions taking into account all prediction algorithms in Table 3 above, the following general features for DeK1 has been observed irrespective of the species it belongs to. In these features the most consensus prediction has been considered.

Molecular weight of the protein: Higher than 230 kDa

Nature of the protein: Slightly Acidic

Subcellular Location: Extracellular, cell wall

Type of protein: Multi-spanning membrane protein

N – terminal orientation: Outside

Number of TM helices: 22-24

Number of Big Loops: 1-2

The big loop predicted by all softwares = ~amino acid position 330-630

Big loop orientation: Inside

C-terminal length: ~1000 amino acids

C-terminal orientation: Inside

Signal peptide at the N-terminus: Signal anchor

Based on these findings, a schematic model for DeK1 can be proposed as shown in Figure 1. The PHOBIUS prediction stands out as similar to that of the general consensus.

It is immediately evident that the model proposed here after extensive analysis is different from that proposed earlier (Lid et al, 2002) based on just one topology predictor. Thus we provide an alternative model and enhance the scope for formulating better hypothesis for experimental verification of these findings.

The Loops

A special feature of the DeK1 model is the presence of at least one big loop of 280-300 amino acid residues in addition to other smaller loops. Loops, in proteins, are mostly involved in protein-protein interactions and ligand binding. The big loop of DeK1 might be important for ligand or substrate binding, especially in light of the function of DeK1 in regulating the formation of aleurone cells as a signaling molecule. A key step in signaling is the phosphorylation of proteins. The typical phosphorylation sites in the big loop of Rice DeK1 were predicted (Figure 2). Only those positions which have probabilities higher than 50% are considered; these are the ones above the threshold line indicated in the graph. Of these only serine and threonine are most commonly phosphorylated in case of plant proteins, while animal proteins are phosphorylated at tyrosine. For rice DeK1, it is seen that 23 serine residues have a high potential for phosphorylation against only 3 for threonine. The lone tyrosine residue predicted could well be a false positive. Similar results were obtained for *Arabidopsis* and maize DeK1 (data not shown). So, the big loop is susceptible to phosphorylation and could be important for the function of DeK1 and is expected to be cytoplasmic as shown in the model (Figure 1).

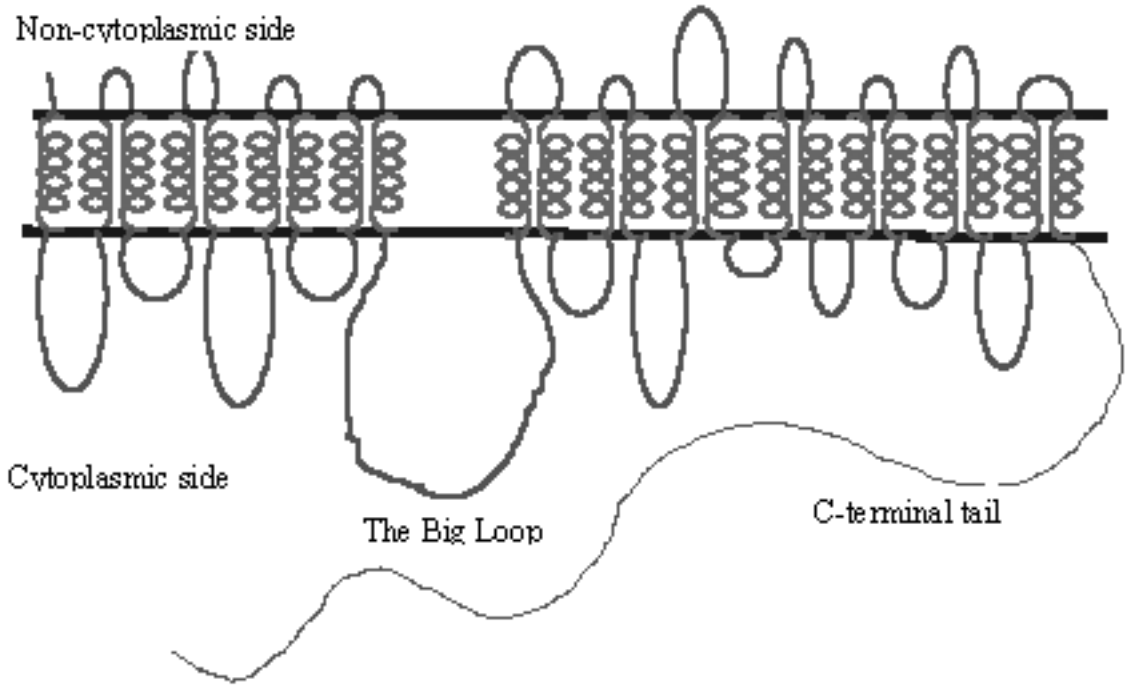

Figure 1: Consensus Model for DeK1 Protein. The Long *C*-terminal Tail and the Big Loop are both Cytoplasmic While the N-terminal is Non-cytoplasmic. The Protein has 23 Transmembrane Helices and Several Smaller Loops of Varying Sizes. The Spirals Represent TMHs, the Ovals Represent Small Loops and the Thick Lines the Membrane Boundaries

A closer inspection of the big loop sequences of the three DeK1 proteins also reveals other interesting features. The sequence of the big loop of rice DeK1 is shown in Figure 3. The sequence shows a high density of charged (K, R, D, E) and strongly polar (Q, N, H) residues. The total percentage is over 40%. None other regions of the DeK1 sequence show such crowding of charged and polar residues. Charged residues can form salt bridges

and bind ligands. Another interesting feature is the presence of six cysteine residues in the big loop sequence. It would be interesting to know whether they are all positioned to form disulfide bridges, or can they be free for thiol-based reactions and how they affect ligand binding and stability. The presence of dense charged residues in the big loop probably justify its prediction on the cytoplasmic side of the cell membrane.

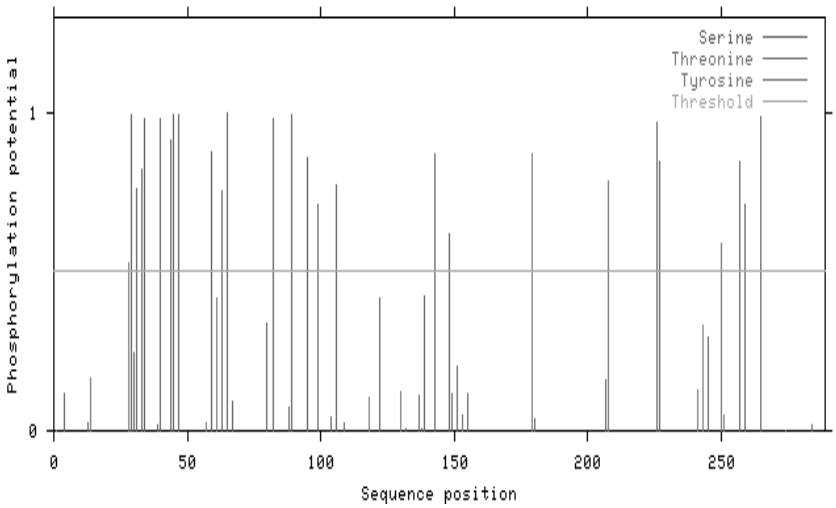

Figure 2: Predicted Phosphorylation Sites in the Big Loop of Rice DeK1.

SNPSVARIDALRSTVIKLREGFRRKGQTSSSNSSDG*C*GSSVKRSSGSVEAGPHGNA
TDSMYRSNSQSD*C*VNWNNVPFDRSNS*C*QEGQSSDKNIDSGRASLAHRSNS*C*LSA
VAVQDPETAVVSADRHGDPTASLVV*C*SSSGLESQG*C*ESSGSATASGNQQLLDLN
LAAIFQDRLNDPRITSMLKRNGGLGDVELANLLQDKGLDPNFSYMMKDKVMDP
RILALLQRSSLDADREHQDDVDVTGTDSDRLDTTIANQISLSEELRRSGLENWLN
LSRLMFHQVAGS

Figure 3. Sequence of the Big Loop of Rice DeK1 Protein. Charged Residues are Colour Coded (Positive Amino Acids are Blue, Negative ones are in Red and the Polar ones in Green). The Cysteine Residues are in Italics and Underlined

The loops can also have other contributions to the structure and/or function of DeK1. For example, they can come together to form channels in the membrane for passage of components necessary for the aleurone development. Also, it is not necessary for the DeK1 protein to function as one component. The membrane portion of the protein might just be needed to translocate the whole protein to the membrane. Then the big loop for example may be cleaved for its function and released into the cytoplasm. If that happens then it is interesting to see at what subcellular location it would end up. A nuclear localization signal (NLS) prediction algorithm was used and the big loop sequence of DeK1 was found not to contain any such signal. However, subcellular localization prediction software predicts that the highest probability of its location would be the nucleus (approximate probability is 50%). The transport of a putative big loop to the nucleus is possible by another intermediary protein carrier since it lacks a NLS. The nuclear localization of the big loop can be very interesting and important for the function of DeK1. Since the cytoplasmic tail is a protease (Wang et al, 2003), it is also possible for the big loop to be an internal substrate for the C-terminal tail calpain. Even if the big loop is not cleaved, it can still phosphorylate substrates or bind ligands and transmit signal into the cytoplasm or even activate the C-terminal tail. The possibilities need to be verified experimentally.

Transmembrane Helices

The DeK1 proteins were predicted to have 22-24 putative TMHs. It would be interesting to know why so many and what they do. One obvious reason is of course to embed the protein in the membrane and maintain the structural integrity of the biomolecule. However, that can be done by a single TMH, or for a large membrane protein, by a few TMHs. These helices thus must also be important for the function of the protein. To analyze these helices better, the sequence for each of them from

all the three DeK1 proteins were identified. A closer inspection of these helix sequences reveals that many of them have charged or strongly polar residues (Table 4). This observation is important since the membrane proteins have to pay a very high-energy price to include charged/strongly polar residues in its sequence (MacKinnon, 2005; Hessa et al, 2005a, b). There must be a functional relevance for such energy sacrifices.

Table 4: Sequence Alignment of Respective Transmembrane Helices of *Arabidopsis*, Maize and Rice DeK1 Proteins. Conserved Charged and Polar Residues are Colour Coded, where Blue Stands for Positive Charge, Red for Negative Charge and Green for Neutral Polar Residues.

ArabTMH1	----------LLACVISGTLFTVFGSGSFWILWAVNWR		
MaizeTMH1	EKMEGEGHHGVVLACSICGFLFAVLSPF----------		
RiceTMH1	EKMEEEEHRGVVLVCSICGFLFAVLGPLS----------		
ArabTMH2	WPTYVQGPQLSTLCSLLTLCAWLVV---------		
MaizeTMH2	--------QLSTLCSLLTLCAWLVVISPIAVLL-		
RiceTMH2	----------TLCSFFTLFAWLVVVSPITVLLV		
ArabTMH3	VLIALLERNIIGLAVIMAGVALL---	ArabTMH4	QWQSSKAVAYLLLLAVGLLCA---
MaizeTMH3	--IALLERNIIGLAVIMAGVALLLS-RiceTMH3	MaizeTMH4	-WQSSKAVAYLLLLAVGLLCAYD-RiceTMH4
	--IALLERNIIGLAVIMVGVALLLSF		--QSSKAVAYLLLLAVGLLCAYEF
ArabTMH5	SASELNSPSGFFFGVSVISLAI---	ArabTMH6	----SSRVKHLGLLYISSLLVLV--
MaizeTMH5	-ASELNSPSGFFFGVSVISLAINM-RiceTMH5	MaizeTMH6	----SSRVKHLGLLYISSLLVLVG-
	--SELNSPSGFFFGVSAISLAINML	RiceTMH6	YMTKSSRVLHLGLLYLCSLMVLVVY
ArabTMH7	GLTSKEARWLGALTSVAVVILDW	ArabTMH8	RFELLKSRMIVLFVAGTSRAF--
MaizeTMH7	GLTSKEARWLGALTSVAVV----RiceTMH7	MaizeTMH8	RFELLKSRMIVLFVAGTSRAFL-RiceTMH8
	-LTSKEARWLGALTSVAVVI---		-FELLKSRMIALFVAGTSRVFLI
ArabTMH9	-YWYLGHCISYAFVASVLLAA-	ArabTMH10	ISRLIFHHLAGSPIRAFI--------
MaizeTMH9	HYWYLGHCISYAFVASVLLSA-	MaizeTMH10	-----FHHLAGSPIRAFIVFTVMFII
RiceTMH9	-YWYLGHCISYAFVASVLLAAA	RiceTMH10	------HQVAGSPIRAFVVFTLIFII
ArabTMH11	IKVINATHEQFEFGFSILLL---------	ArabTMH12	SLRAEEMLMTSKPQKYGF---------
MaizeTMH11	-------HEQFEFGFSILLLSPVVCSIM-RiceTMH11	MaizeTMH12	-------LMTSKPQKYGFIAWLLSTC-
	-------HEQFEFGFSILLLSPVVCSIMA	RiceTMH12	--------MTSKPRKYGFIAWLLSTCV
ArabTMH13	LSTCVGLFLSFLSKSSVILGLS----------	ArabTMH14	FANRENVSQAPGEKERAL------------
MaiZeTMH13	---------SFLSKSSVILGLSLTVPLMVACL	MaizeTMH14	--------QAPGEKERALFVITIAVFTAS-
RiceTMH13	----------FLSKSSVILGLSLTVPLMVACL	RiceTMH14	---------APGKKERALFAISITVFTASV
ArabTMH15	--PYTSSVYLGWAMSSGIALVVTAILPIV	ArabTMH16	--RFSHSSAVCLMIFSVVLVAF-
MaizeTMH15	YSPYATSMYLGWALSSTIAVITTGLIP--	MaizeTMH16	TYRFSPSSAICVGLFATVLVSF-
RiceTMH15	-SPYATSMYLGWALSSTIAVLATGVIPI-	RiceTMH16	-YRFSPSSAICVGLFATVLVSFC
ArabTMH17	----LPTKGDFLAALLPLACIPALLS	ArabTMH18	--CWILSRGVYVFFSIGLLLLFGA
MaizeTMH17	REDGVPLKADFLAALLPLLCIPAF--	MaizeTMH18	DDDWKISRGVYLFVGMGMLLL---
RiceTMH17	--DGVPLKADFLAALLPLLCIPAVF-	RiceTMH18	-DDWKISRGVYLFVGMGVLLLL--
ArabTMH19	---AVKPWTIGVSFLLVLFLMVVTIG	ArabTMH20	---LTRKQTSFVCFLALLLGLAA
MaizeTMH19	VIVTIRPWTVGVACLVAILFLVF---	MaizeTMH20	NFYLTRTQMLLVCSIAFLLAL--
RiceTMH19	VIVTIRPWTVGVACLLVILFLVFA--	RiceTMH20	-FYLTRTQMLLVCSLAFLLALA-
ArabTMH21	--AGASVGYFTFLSLLAGRALAVLL	ArabTMH22	-ADCGKNVSAAFLVLYGIALA-------
MaizeTMH21	PFVGASIGYFSFIFLLTGRALTV--	MaizeTMH22	HADSAKNVSYAFLILYGIALATE-----
RiceTMH21	-FVGASIGYFSFLFLLTGRALTVL-	RiceTMH22	HADSAKNVSYAFLILYGIALATEVWGVI
ArabTMH23	VVASLIIYPPFAGAAVSAITLVVAF		
MaizeTMH23	--ASLIMNPPFVGAGVSATTL----		
RiceTMH23	---SLILNPPFIGAAISAITLV---		

Referring to literature on various membrane proteins, it was found that this is indeed true. Charged and polar residues do play an important role in the structure and function of membrane proteins (Sal-man and Shai, 2005; Zhou et al, 2001; Li et al, 2004). Engelman (Zhou et al, 2001) has aptly remarked that "…despite their rare presence in TM helices, strongly polar residues are highly 'conserved', especially in multispanning TM proteins, suggesting molecular interactions that either functionally or structurally favor these residues…". Thus, lactose

permease from *E.coli* has six amino acid residues that are absolutely necessary for its function (Zhou et al, 2001; Zhao et al, 1999). All these residues are polar and present in TM helices forming three pair of salt bridges. Voltage-gated K^+, Na^+ and Ca^{2+} channels have "conserved" polar residues in their TM helices that help in voltage sensing and/or association of helices (Bezanilla, 2000). Among the members of G-protein coupled receptor superfamily, family of rhodopsin-like receptors have "conserved" polar residues extremely important for ligand binding and signal transduction (Flower, 1999; Gether and Kobilka, 1998). In light-harvesting complex, ion pairs in TM helices help stabilize the membrane protein structure (Kuhlbrandt et al, 1994). This suggests that the charged/strongly polar residues in the TMHs of DeK1 must serve a functional/ structural role.

The sequences of the corresponding TM helices from three DeK1 proteins were aligned as shown in Table 4 in light of the above knowledge from literature. That is TM helix number 1 from rice, *Arabidopsis* and maize DeK1 were aligned. Then TM helix number 2 from rice, *Arabidopsis* and maize DeK1 were aligned, so on and so forth. From the alignment, the "conserved charged and strongly polar residues" were identified. From these results it is evident that many of the TM helices have either conserved charged residues or strongly polar residues or both. So, one reason for DeK1 to have so many TMHs may be to load the protein with charged or polar residues while still being embedded in the hydrophobic environment of the membrane. This would argue that the ligand that prompts signaling in DeK1 is charged in nature. The charged or polar residues in DeK1 TMHs would help to bind the ligand and transmit it through the membrane.

It is striking to note that some of the adjacent TM helices have conserved charged residues that are oppositely charged. This results in a possibility of forming "salt bridges" or ionic bonds or ion pairs. Thus TM helices 3 and 4, 5 and 6 and 7 and 8 have oppositely charged conserved residues. Then the big loop interferes. This is followed by the presence of opposite charges again in helices 11 and 12, 13 and 14, 16 and 17 and 21 and 22. Similarly, helices 3, 4, 5, 6, 9, 10, 11, 20 and 22 have conserved strongly polar residues (asparagine, glutamine, and histidine). We can place these charges and polar residues on the model for DeK1 represented in figure 1 and the resulting model is shown in Figure 4. The helices have been numbered.

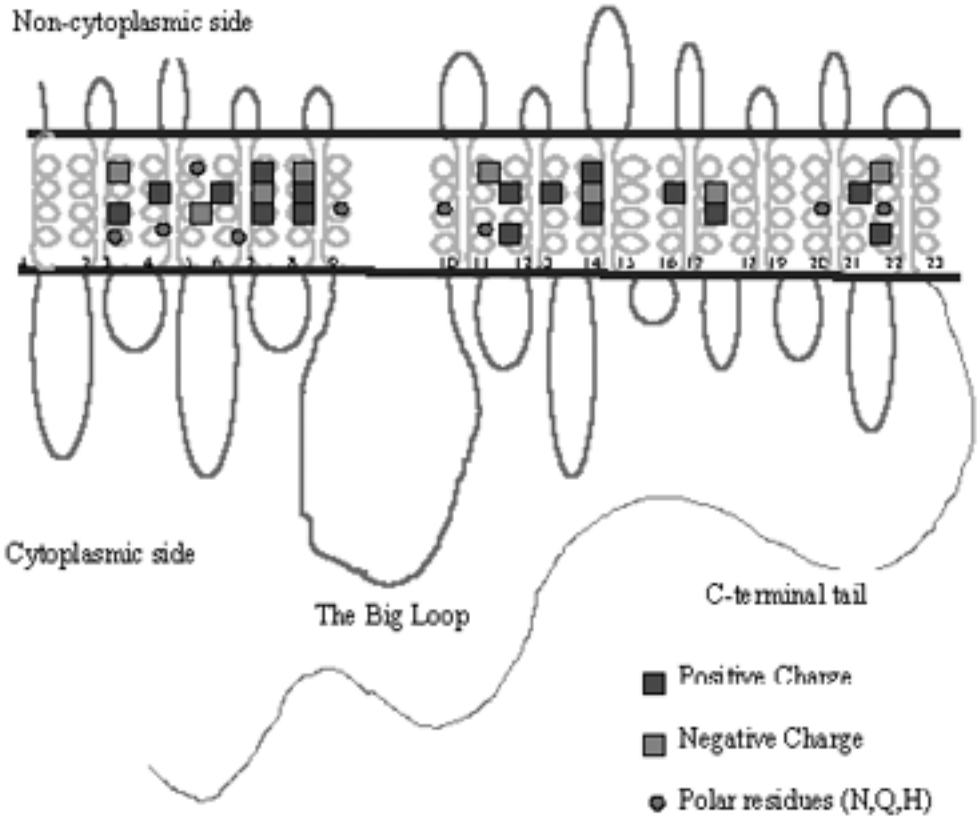

Figure 4: Model of DeK1 Showing Various Features. It has 23 Transmembrane Helices. The Conserved Charged and Polar Residues in Transmembrane Helices are Indicated in the Model and Mostly are Concentrated Around the Big Loop. The Big Loop is on the Cytoplasmic Side and Contains a Large Number of Charged Residues and Putative Phosphorylation Sites. The C-terminal Tail is also Cytoplasmic in Orientation.

It is interesting to see how in the model the conserved charged and polar residues are crowded mostly near the big loop. Oppositely charged amino acid residues can interact to form salt bridges either in the same loop or between adjacent loops. Such salt bridges can be dynamic in the sense that they can form and open up reversibly in presence of a stimuli or a ligand. The conserved strongly polar residues (spheres in Figure 4) can neutralize charges. Polar residues when present in TM helices can act both as acids and bases and so can form salt bridges themselves. As in lactose permease and some other membrane proteins, these salt bridges along with conserved charged and polar residues could be important structurally and/or functionally for DeK1 protein.

C-terminal Tail

The long C-terminal tail (Figure 4) is the only domain of DeK1 for which some biochemical information is available (Wang et al, 2003) and functional implications have been highlighted (Johnson et al, 2008). Constituting the major portion of DeK1 (about 1000 amino acids out of 2159 amino acid residues), the C-terminal tail codes for a calpain like domain. The maize DeK1 calpain like domain was shown to have high similarity to animal calpain, including a conserved catalytic site (Lid et al, 2002; Wang et al, 2003). *Arabidopsis thaliana* and maize DeK1 share 70% overall identity; the identity of calpain like domain being even higher at 85% (Lid et al, 2002). Calpains are cysteine proteinases that are activated by rise in intracellular calcium concentration and are involved in mediating signal transduction leading to cell differentation, proliferation and cell death in animal system (Sato and Kawashima, 2001; Ahn et al, 2004). *In vitro* studies have demonstrated that the maize DeK1 calpain has proteinase activity similar to animal m-calpain and that the activity depends on cysteine residue of conserved active site (Wang et al, 2003). In contrast to m-calpain, the DeK1 calpain was active in the absence of calcium, the activity being stimulated by addition of calcium. However, the specific activity of the putative proteinase was extremely low (Wang et al, 2003) and it might be necessary to find the proper substrate. Lately, it was demonstrated that the calpain like proteinase domain undergoes autocatalysis and this domain seems to play a major role in imparting DeK1 its physiological function (Johnson et al, 2008).

The conventional calpains typified by human m-calpains, are cytosolic enzymes activated by a rise in intracellular Ca, as well as by signaling that includes the protein kinase C-tyrosine kinase or adhesion molecule-derived cascade (Sato and Kawashima, 2001). Upon activation, the enzyme associates with proteins or phospholipids in the plasma membrane and undergoes autolysis (Sato and Kawashima, 2001). The substrates of calpains are typically part of signal transduction cascades, such as protein kinase C's (Kishimoto et al, 1989; Temmblay et al, 2000). Through these mechanisms conventional calpain fulfill essential function in multiple developmental pathways in animals (Arthur et al, 2000). By analogy, the C-terminus of DeK1, could be involved in developmental processes in the endosperm cells that are at the surface thus converting them to aleurone cells. Like animal calpains, the activated C-terminal calpain domain of DeK1 might associate with the plasma membrane with the help of charged residues in the TM helices to undergo autolysis. The big loop, with its capacity to be heavily phosphorylated, might activate the calpain domain of DeK1 either directly or through other intermediary proteins. It is also possible that the big loop is itself a putative substrate for the C-terminal proteinase. All these processes could trigger the downstream signaling that regulates cell differentiation.

IV SUMMARY

The *in silico* investigations performed here delineate the DeK1 sequence and help to model the protein. The model provides us with domains that individually or in combination can be important for the structure and function of the protein. It also indicates various structural elements that can be studied experimentally to understand the protein better. The model provides a platform for testing various hypotheses regarding the protein function.

Previously, during genetic analysis, a preliminary attempt was made to model DeK1 (Lid et al, 2002). However, only one predictor TMHMM was used. We have shown here how one predictor and especially TMHMM alone can lead to wrong predictions. The model that was obtained (Lid et al, 2002) did not fit into their observations and expectations, and it was turned upside down to fit into their hypothesis regarding the function of the C-terminal tail of DeK1. However, our attempt at modeling DeK1 has yielded a better model that can fit into their hypothesis without having to turn it upside down or changing features of the theoretical model.

The bioinformatic analysis has supported many of the results from genetic analysis. It can be said unequivocally that DeK1 is a membrane protein and that it is localized in the cell wall. However, some of the previous results were contradicted as well. It was suggested that the C-terminal tail and the big loop are oppositely oriented to the membrane. However, we have shown that they are on the same side (cytoplasmic side) of the membrane. Our results help hypothesize the relationship of the structure of the protein to its function better. At the least, we have shown an alternative model

and the benefits of using multiple topology prediction algorithms in decoding a membrane protein sequence. It is also established from our studies that PHOBIUS is a better TM topology predictor, atleast for DeK1.

Our observations and analysis regarding the features of the big loop, its relation to the smaller loops, the presence of conserved charged and polar residues in the TM helices of DeK1, and their functional implications are novel.

We can now visualize a putative DeK1 structure atleast at the secondary structure level, albeit theoretical. We know what to expect when we initiate experimental investigations, what to look at, what to observe closely, what the challenges are and what the various options are. The theoretical studies here have given us a solid footing for experimental approaches.

Genetic analysis and cell biology studies have shown that DeK1 plays an important role in aleurone cell development (Lid et al, 2005; Lid et al, 2002; Becraft, 2001; Becraft and Asuncion-Crabb, 2000; Becraft et al, 2002). Endosperm cells that are on the "surface" seem to get differentiated into aleurone cells forming a single layer below the pericarp (Gifford et al, 2003). DeK1 regulates this process. It must be doing so by a signaling mechanism yet unknown. The model proposed here supports this role of DeK1. If it has to sense the "surface" it can best be done from the cell membrane than the cytoplasm. The model shows that DeK1 has a number of loops that protrude out in the non-cytoplasmic side that might sense any stimuli. If an external ligand stimulates the signaling process then the loops can come together to form a channel to allow the passage of the ligand. The conserved charged and strongly polar residues in the TM helices can help ligand binding. These residues can form salt bridges as well and in the presence of a putative ligand can open up to help ligand binding. If this is true then the helices near the big loop would be important since the charged and polar residues are crowded in this region. Such opening and closing of salt bridges can also lead to conformational changes in the protein structure and the signal can be sent inside as a result.

The properties of the big loop hint at an important role as well. Signaling events ought to involve kinases and phosphorylation to initiate the signaling cascade. The big loop contains a number of probable phosphorylation sites for this to happen. The high charge density on the big loop may help this protein to interact with other substrate proteins as well. That is to say, the big loop could be a docking site for substrates. The importance of calpains in animal cells in regulating cell development has been mentioned in the results and discussion section. The C-terminal tail of DeK1 has homology to calpain domains and a cysteine residue at its active site typical of a proteinase. The big loop could be a substrate for the calpain-like protease. The protease could cleave the big loop off the protein, which can then reach the nucleus triggering a set of reactions that signal the cells to specify into aleurone cells. We have shown theoretically that the big loop by itself has a high probability of nuclear sublocalization.

As discussed in the previous section, the human m-calpains are cytosolic enzymes that are activated by signaling that includes the protein kinase C-tyrosine kinase or adhesion molecule-derived cascade. The big loop can probably serve this purpose as well. In summary, the model of DeK1 shows that it has many features that can make the protein capable of signaling events leading to regulation of cell development and differentiation.

V ACKNOWLEDGEMENTS

The financial assistance provided by University of Delhi through R&D grants, Department of Biotechnology (DBT) through research grant to SK (BT/PR8901/BRB/10/535/2007), and University Grants Commission (UGC) through Special Assistance (SAP) programme to the Department of Biochemistry is greatly acknowledged.

REFERENCES

[1] Ahn, J.W., Kim, M., Lim, J.H., Kim, G.T., and Pai, H.S. (2004). "Phytocalpain Controls the Proliferation and Differentiation ..... *Plant J*. **38**: 969-981.

[2] Arai, M., Mitsuke, H., Ikeda, M., Xia, J.X., Kikuchi, T., Shatake, M., and Shimizu, T. (2004). "CornPredII: a Consensus Prediction Method for Obtaining Transmembrane Topology Models with High Reliability." *Nucleic Acids Res*. **32**: W390-393.

[3] Arthur, J.S.C., Elec, J.S., Hegadorn, C., Williams, K., and Greer. P.A. (2000). "Disruption of the Murine Calpain Small Subunit Gene, Capn4: Calpain is Essential for Embryonic Development but not for Cell Growth and Division." *Mol. Cell.Biol*. **20**: 4474-4481.

[4] Becraft, P.W. (2001). "Cell Fate Specification in the Cereal Endosperm". *Semin. Cell. Dev. Biol* **12**: 387-394.

[5] Becraft, P.W., and Asuncion-Crabb,Y. (2000). "Positional Cues Specify and Maintain Aleurone Cell Fate in Maize Endosperm Development". *Development*. **127**: 4039-4048.

[6] Becraft, P.W., Brown, R.C., Lemmon, B.E.O., Ferstad, H.G., and Olsen, O.A. (2001). "Endosperm Development. *In* "*Current Trends In The Embryology Of Angiosperms*." (S. S. Bhojwani, Ed.) pp. 353-374.

[7] Becraft, P.W., Li, K., Dey, N., and Asuncion-Crabb, Y. (2002). "The Maize Dek1 Gene Functions in Embryonic Pattern Formation and in Cell Fate Specification." *Development*. **129**: 5217-5225.

[8] Bendtsen, J.D., Nielsen, H., von Heijne G., and Brunak, S. (2004). "Improved Prediction of Signal Peptides: SignalP 3.0." *J. MoI. Biol.* **340**: 783-795.

[9] Berger, F. (1999). "Endosperm Development." *Curr. Opin. Plant Biol*. **2**: 28-32.

[10] Bezanilla, F. (2000). "The Voltage Sensor in Voltage-dependent Ion Channel." *Physiol. Rev.* **80**: 555-592.

[11] Blom, N., Gammeltoft, S., and Brunak, S. (1999). "Sequence- and Structure-based Prediction of Eukaryotic Protein Phosphorylation Sites." *J. Mol. Biol*. **294**: 1351-1362.

[12] Cokol, M., Nair, R., and Rost, B. (2000). "Finding Nuclear Localization Signals." *EMBO Rep*. **2001**: 411-415.

[13] David, D., Das, S.J.N., Thomas, U., Chen-Ni, C., and Jan, W.G. (2002). "Rapid Topological Mapping of *E. Coli* Inner- membrane Protein." *Proc. Natl. Acad. Sci.* USA **5**: 2690-2695.

[14] Dixon, B., Walker, B. Kimmins, W., and Pohajdak, B. (1991). Isolation and Sequencing of a cDNA for an Unusual Hemoglobin from the Parasitic Nematode *Pseudoterranova Decipiens*." *Proc. Natl Acad. Sci*. USA. **88**: 5655-5659.

[15] Flower, D.R (1999). "Modeling G-protein-coupled Receptors for Drug design." *Biochim. Biophys. Acta*. **1422**: 207-234.

[16] Fortina, M.G., Ricci, G., Mora, D., Guglielmetti, S., and Manachini, P.L. (2003). "Unusual Organization for Lactose and Galactose Gene Clusters in *Lactobacillus Helveticus*." *Appl. Environ. Microbiol*. **69**: 3238-3243.

[17] Galinat, W.C. (1990). "Multi-layered Expression of Aleurone Specific Genes." *Maize News Lett*. **64**: 121.

[18] Gether, U., and Kobilka, B.K. (1998). "G. Protein-coupled Receptors. II. Mechanism of Agonist Activation." *J. Biol. Chem.* **273**: 17979-17982.

[19] Gibson, Q.H., Regan, R., Olson, J.S., Carver, T.E., Dixon, B., Pohajdak, B., Sharma, P.K., and Vinogradov, S.N. (1993). "Kinetics of Ligand Binding to *Pseudoterranova Decipiens* and *Ascaris Suum* Hemoglobins and to Leu-29 → Tyr sperm whale Myoglobin Mutant." *J. Biol. Chem*. **268**: 16993-16998.

[20] Gifford, M.L., Dean, S., and Ingram, G.C. (2003). "The Arabidopsis ACR4 Gene Plays a Role in Cell Layer Organisation During Ovule Integument and Sepal Margin Development." *Development*. **130**: 4249–4258

[21] Hessa, T., Kim, H., Bihimaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., Stephan, H., and von Heijne G. (2005a). "Recognition of Transemembrane Helix by Endoplasmic Reticulum Translocon." *Nature*. **433**: 377-381.

[22] Hessa, T., White, S.H., and von Heijne, G. (2005b). "Membrane Insertion of a Potassium-channel Voltage Sensor." *Science* **307**: 1427.

[23] Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998). "SOSUI: Classification and Secondary Structure Prediction System for Membrane Proteins." *Bioinformatics*. **14**: 378-379.

[24] Hofmann, K., and Stoffel, W. (1993). "A Database of Membrane Spanning Protein Segments." *Biol.Chem*. **374**: 166.

[25] Horton, P., and Nakai, K. (1997). "Better Prediction of Protein Cellular Localization Sites with the k Nearest Neighbors Classifier." *Proc. Int. Conf. Intell. Syst. Mol. Biol*. **5**: 147-152.

[26] Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., and Gibson, T.J. (1998). "Multiple Sequence Alignment with Clustal X." *Trends Biochem Sci*. **23**: 403-405.

[27] Johnson, K.L., Degnan, K.A., Walker, J.R., and Ingram, G.C. (2005). "*AtDEK1* is Essential for Specification of Embryonic Epidermal Cell Fate." *Plant J*. **44**: 114-127.

[28] Johnson, K.L., Faulkner, C., Jeffree C.D., and Ingram, G.C. (2008). "The Phytocalpain Defective Kernel 1 is a Novel Arabidopsis Growth Regulator whose Activity is Regulated by Proteolytic Processing." *The Plant Cell,* **20,** 2619–2630.

[29] Kall, L., Krogh, A., and Sonnhammer, E.L. (2004). "A Combined Transmembrane Topology and Signal Peptide Prediction Method." *J. Mol. Biol.* **338**: 1027–1036.

[30] Kishimoto, A., Mikawa, K., Hasimoto, K., Yasuda, I., Tanaka, S., Tominaga, M., Kurada, T. and Nishizuka, Y. (1989). "Limited Proteolysis of Protein Kinase C subspecies by Calcium-dependent Neutral Protease (calpain)." *J. Biol.Chem*. **264**: 4088-4092.

[31] Krogh, A., Larsson, B., von Heijne G., and Sonnhammer, E.L. (2001). "Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genome." *J. Mol. Biol*. **305**: 567-580.

[32] Kuhlbrandt, W., Wang, D.N., and Fujiyoshi, Y. (1994). "Atomic Model of Plant Light-harvesting Complex by Electron Crystallography." *Nature*. **367**: 614-621.

[33] Li, X., Ding, J., Liu, Y., Brix, B.J., and Fliegel, L. (2004). "Functional Analysis of Acidic Amino Acid in Cytosolic Tail of the Na+/H+ exchanger." *Biochemistry* **43**: 16477-16486.

[34] Lid, S. E., Gruis, D., Jung, R., Lorentzen, J.A., Ananiev, E., Chamberlin, M., Niu, X., Meeley, R., Nichols, S., and Olsen, O-.A. (2002). "The *Defective Kernel 1 (dek1*) Gene Required for Aleurone Cell Development in the Endosperm of Maize Grains Encodes a Membrane Protein of the Calpain Gene Superfamily." *Proc. Natl. Acad. Sci. U.S.A.* **99**: 5460-5465.

[35] Lid, S.E, Olsen, L., Nestestog, R., Aukerman, M., Brown, R.C., Lemmon, B., Mucha, M., Opsahl-Sorteberg, H-G, and Olsen, O-A. (2005). "Mutation in the *Arabidopsis Thaliana* DeK1 Calpain Gene Perturbs Endosperm and Embryo Development While Over-expression Affects Organ Development Globally." *Planta,* **221**, 339-351.

[36] MacKinnon, R. (2005). "Membrane Protein Insertion and Stability." *Science*. **307**: 1425-1426.

[37] Melen, K., Krogh, A., and von Heijne, G. (2003). "Reliability Measures for Membrane Protein Topology Prediction Algorithms." *J. Mol. Biol*. **327**: 735-744.

[38] Mitaku, S., and Hirokawa, T. (1999). "Physicochemical Factors for Discriminating between Soluble and Membrane Proteins: Hydrophobicity of Helical Segments and Protein Length." *Protein Eng*. **11**: 953-957.

[39] Mitaku, S., Hirokawa, T., and Ono, M. (1998). "Classification of Membrane Proteins by Types of Transmembrane Helices Using SOSUI System." *Genome Informatics*. **9**: 367-368.

[40] Moller, S., Croning, M.D.R., and Apweiler, R. (2001). Evaluation of Methods for the Prediction of Membrane Spanning Proteins." *Bioinformatics*. **17**: 646-653.

[41] Nakai, K. and Kanehisa, M. (1991). "Expert System for Predicting Protein Localization Site in Gram-negative Bacteria." *Proteins,* **11**, 95-110.

[42] Nakai, K., and Horton, P. (1999). "PSORT: a Program for Detecting Sorting Signals in Proteins and Predicting their Subcellular Localization." *Trends Biochem. Sci*. **24**: 34-36.

[43] Nelson, O.E., and Chang, M.T. (1974). "Effect of Multiple Aleurone Layers on the Protein and Amino Acid Content of Maize Endosperm." *Crop Sci*. **14**: 374-376.

[44] Olsen, L.T., Divon, H. H., Al, R., Fosnes, K., Lid S. E., and Opsahl-Sorteberg, H. (2008). "The Defective Seed5 (des5) Mutant: Effects on Barley Seed Development and HvDek1, HvCr4, and HvSal1 Gene Regulation." *J. Exp. Bot*. **59**: 3753-3765.

[45] Olsen, O-.A. (2001). "Endosperm Development: Cellularization and Cell Fate Specification." *Annu. Rev. Plant Physiol. Plant Mol Biol*. **52**: 233-267.

[46] "Penalties and Weight Matrix Choice." *Nucleic Acid Res*. **22**: 4673-4680.

[47] Persson, B., and Argos, P. (1994). "Prediction of Transmembrane Segments in Proteins Utilising Multiple Sequence Alignments." *J. Mol. Biol*. **237**: 182-192.

[48] Persson, B., and Argos, P. (1996). "Topology Prediction of Membrane Proteins." *Prot. Sci*.**5**: 363-371.

[49] Ritchie, S., and Gilroy, S. (1998). "Abscisic Acid Signal Transduction in the Barley Aleurone is Mediated by Phospholipase D. Activity." *Proc. Natl. Acad. Sci. USA*. **95**: 2697-2702.

[50] Sal-man, N. and Shai, Y. (2005). "Arginine Mutations within a Transmembrane Domain of Tar, an *Escherichia Coli* Aspartate Receptor, can Drive Homodimer Dissociation and Heterodimer Association *in vivo*". *Biochem. J*. **385**: 29-36.

[51] Sato, K. and Kawashima, S. (2001). "Calpain Function in the Modulation of Signal Transduction Molecules." *J. Biochem*. **382**: 743-751.

[52] Sheridan, W.F. and Neuffer, M.G. (1980). Defective Kernel Mutants of Maize. II. Morphology and Embryo Culture Studies." *Genetics*. **95**: 945-968.

[53] Stoffel, W., Duker, M., and Hofmann, K. (1993). "Molecular Cloning and Gene Organization of the Mouse Mitochondrial 3, 2-trans-enoyl-CoA Isomerase." *FEBS Lett*. **333**: 119-122.

[54] Temmblay, R., Chakravarthy, B., Hewitt, K., Tauskela, J., Morly, P., Atkinson, T., and Durkin, J.P. (2000). "Transient NMDA Receptor Inactivation Provides Long-term Protection to Cultured Cortical Neurons from a Variety of Death Signals." *J. Neurosci*. **20**: 7183-7192

[55] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994), Clastal 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.

[56] Tusnady, G.E. and Simon, I. (1998). "Principles Governing Amino Acid Composition of Integral Membrane Proteins: Application to Topology Prediction." *J. Mol. Biol*. **283**: 489-506.

[57] Wang, C., Barry, J.K., Min, Z., Tordsen, G., Rao, A.G. and Olsen, O-.A. (2003). The Calpain Domain of Maize DeK1 Protein Contains the Conserved Catalytic Triad and Functions as a Cysteine Proteinase." *J Biol. Chem.* **278**: 34467-34474.

[58] Zhao, M., Zen, K.C., Hubbell, W.L., and Kaback, H.R. (1999). "Proximity between Glu126 and Arg144 in the Lactose Permease of Escherichia coli." *Biochemistry*. **38**: 7407-7412.

[59] Zhou, F.X., Merianos, H.J., Brunger, A.T. and Engelman, D.M. (2001). "Polar Residues Drive Association of Polyleucine Transmembrane Helices." *Proc. Natl. Acad. Sci*. USA. **98**: 2250-2255.

MILLIPORE

Innovative Solutions for Protein Research

SNAP i.d. Protein Detection System :

Rapid immunodetection in just 22 minutes

The SNAP i.d. Protein Detection System revolutionizes immunodetection every time — in record time! Unlike conventional Western blotting, where diffusion is the primary means of reagent transport, the SNAP i.d. system applies a vacuum to actively drive reagents through the membrane

- **Dynamic** – vacuum actively drives reagents through blotting membranes
- **Quality** – equal or better signal-to-noise ratios than standard western blotting
- **Fast** – reduces immunodetection time from 4 hours to 30 minutes
- **Simple** – incorporates blocking, washing and antibody incubation steps
- **Compatible** – works with standard gel sizes and protocols
- **Efficient** – optimizes your protocol with a new antibody in 30 minutes

Luminata Western HRP Chemiluminescence Substrates :

Premixed for convenience. Formulated for optimal results

The Luminata HRP substrates are a family of premixed, ready-to-use substrates for the detection of HRP-based Westerns. You can pour Luminata substrates directly out of one bottle onto your blot without having to mix two solutions together

- Premixed for convenience
- Stable at room temperature or 4°C
- Most sensitive in their class
- Consistent results with less pipetting error
- Available in three different sensitivities to meet all your detection needs

Bløk :

Noise-Cancelling Reagents

Bløk reagents are a family of uniquely formulated protein-free blocking solutions that provide numerous benefits including :
Reduced background for chemiluminescent or fluorescent detection

- Unique formulation for detection of phosphorylated proteins
- More stable diluent for antibodies than milk
- Allows chromogenic staining of membranes after immunodetection

Immobilon :

Western Blotting Transfer Membrane

Immobilon PVDF membranes have low background, broad solvent compatibility, and superior staining capabilities. In addition, they can be reprobed multiple times

- Unique formulation for detection of phosphorylated proteins
- More stable diluent for antibodies than milk
- Allows chromogenic staining of membranes after immunodetection

For more information, please contact :
Pravin Jain, Mobile : +919313290421, Email : pravin_jain@millipore.com

ADVANCING LIFE SCIENCE TOGETHER®
Research. Development. Production.