

## Protein characterization and sequence analysis of ALLCE antimicrobial peptide from *Allium cepa*

Raid Al Akeel<sup>1</sup>, Ayesha Mateen<sup>2\*</sup>, V.C. Gupta<sup>2</sup> and K. Janardhan<sup>3</sup>

<sup>1</sup>Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

<sup>2</sup>Central Research Institute for Unani Medicine, Opp. Employees' State Insurance, Hyderabad, India

<sup>3</sup>P.G. Department of Biotechnology, A.V. College, Hyderabad, India

\*Corresponding Author E-mail: [ayeshamateen@gmail.com](mailto:ayeshamateen@gmail.com)

### Publication Info

Paper received:

11 June 2015

Revised received:

14 September 2015

Re-revised received:

07 October 2015

Accepted:

28 October 2015

### Abstract

Antimicrobial proteins/peptides produced by plant seeds participate in protection of seeds against pathogenic organisms. A study was carried out to investigate the *in silico* analysis of protein sequence localization, structure, homology modeling and 3D structure prediction of ALLCE-AMP in *Allium cepa*. Primary structure prediction and physico-chemical characterization were performed by computing theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY). In the present study, homology modeling, a high quality of peptide 3D structure, was predicted by submitting the peptide sequence (target) to EYPred3D web server. The template (1T12 chain A) was found to share 18.2% identity with the Query (B2CZN8). The model was validated using protein structure checking tools PROCHECK and ERRAT VALUE (62.353). The present study would be useful in studying protein-protein interactions and drug designing.

### Key words

*Allium cepa*, ALLCE-AMP, PROCHECK, Structure prediction, Subcellular localization

### Introduction

Plants have established numerous defense systems to guard themselves from invasion of pathogens during their evolution (Liu *et al.*, 2000). Plants produce secondary metabolites in the form of antimicrobial peptides or proteins showing vast variety in antimicrobial spectrum, structure, function and mechanism of action. Most interestingly, there is growing proof that AMPs also fulfil essential biological functions (Nikoletta and Florentine, 2013). Previous studies have shown that EtOH extract from fly maggots can potently inhibit MRSA and VRE strains, with a ready supply of inhibitory compound being recovered in separated butanol fraction (Sang *et al.*, 2010). AMPs are gene-encoded and are either constitutively expressed or quickly copied upon induction. In higher eukaryotes invading microbes and their products, e.g. lipopolysaccharides (Mendez-Samperio *et al.*, 2007; Amlie-Lefond *et al.*, 2005), or host cellular

compounds, such as butyrate (Murakami *et al.*, 2002), vitamins (Schauber *et al.*, 2006) cytokines (Wolk *et al.*, 2004; Wilson *et al.*, 2007; Lai and Gallo, 2009) and encourage AMP manufacture. Studies of DNA and protein sequence homology are essential for variety of purposes and have consequently developed routine in computational molecular biology. The amino acid sequences and construction of Ace-amp1 protein have been determined which differs in few amino acid sequences to ALLCE antimicrobial peptide (ALLCE-AMP) from *Allium cepa*. AMPs function is strictly dependent on structural and physico-chemical properties to increase antimicrobial activity usually by changing molecular size and charge, residues arrangement, amphipathicity, hydrophobicity and helix folding probability (Tossi *et al.*, 2000; Tain *et al.*, 2009).

Allocating subcellular localization to protein is an important step towards interpreting its interaction partners,

function and its probable role in cellular machinery (Rost *et al.*, 2003). Despite current experimental determination, technological advancements subcellular localization remains laborious and time-consuming. Therefore, computational approaches for assigning localization a proteome-wide scale offer an attractive complement. Methods for predicting subcellular localization can be categorized according to the fundamental theory, e.g. classification stands on N-terminal targeting sequences, overall amino acid composition, and sequence homology. TargetP, (Hoglund *et al.*, 2006) utilize neural networks for perceptive four localizations: mitochondrial, chloroplast, secretory pathway and additional proteins, based on their N-terminal sequence information. An alternative and comparable method, PSORT, (Emanuelsson *et al.*, 2000) uses biologically interpretable rules of N-terminal sequences for assigning same localizations as TargetP.

In view of the above, the present study was carried out to predict three dimensional structure of ALLCE-AMP and to describe *in silico* analysis of primary, secondary, 3D structure prediction and homology modeling which would provide insight into its structure and knowledge of subcellular localization of a protein.

### Materials and Methods

**Primary and secondary structure prediction :** The primary peptide sequence of ALLCE-AMP were retrieved from Uniprotkb database in FASTA format and used for further analysis. Percentages of hydrophilic and hydrophobic residues were calculated from primary structure examination. The physico-chemical parameters were computed with the help of Expasy's ProtParam prediction server. The in Expasy, SOPMA tool was used for the secondary structure class identification, secondary structure prediction and for computation of percentages of  $\alpha$ -helical,  $\beta$ -strand and coiled regions.

**Comparative modelling and validation :** Swiss model is automated modelling software which develops the 3D structure model of indefinite structure protein built on the sequence homology with known structured protein. It is significant to note that for structure prediction, the sequence homology must be higher than 30%. ESYPred3D server predicts the putative 3D modelled structure (Ashwani and Anshul, 2010). The theoretical structure of ALLCE-AMP was built using ESYPred3D server by submitting the target sequence (Lambert *et al.*, 2002). Validation of tertiary structure of protein was done by PROCHECK.

**Protein location and localization sites prediction :** ALLCE-AMP localisation sites were predicted using PSORT tool. It analyzes the input sequence by applying

stored rules for several sequence features of known protein sorting signals using Target P, a neural network-based tool; it is used to predict the subcellular location in newly identified proteins.

### Results and Discussion

ALLCE antimicrobial peptide sequence were retrieved from Uniprotkb, in FASTA format and used for further analysis. In the present study, the primary structure of peptide were predicted by using the Expasy's ProtParam server (Gasteiger *et al.*, 2005). The results showed that ALLCE-AMP had 120 amino acid residues and estimated molecular weight was 13737.3. The calculated isoelectric point would be beneficial because at pI, solubility was least and mobility in an electro focusing system was zero. Isoelectric point is the pH at which the surface of protein is covered with charge but net charge of protein is 0. The computed pI value of ALLCE-AMP was 11.74. Computed pI value of protein greater than 7 (pI>7) indicates that ALLCE-AMP was as basic. The computed isoelectric point will be beneficial for developing buffer system for purification by isoelectric directing system. The total number of negatively charged residues (Asp + Glu) was 2 and Total number of positively charged residues (Arg + Lys) was 20. Aliphatic index (AI) is defined as relative volume of a protein employed by aliphatic side chains was observed as a positive factor for increase in thermal stability of globular proteins. Aliphatic index for ALLCE-AMP sequence was 99.08. Whereas, Expasy's ProtParam computed extinction coefficient at 280 nm wavelengths and was preferred because proteins absorb light strongly, while other substances usually do not. ALLCE-AMP 's Extinction coefficient at 280 nm ranged from 14480 to 13980  $\text{m}^{-1} \text{cm}^{-1}$  with respect to the concentration of Cys residues. This computed extermination coefficients supported the quantitative study of protein–ligand interactions in solution.

**Table 1 :** Amino acid percentage analysis of ALLCE antimicrobial peptide

AA	%tage	AA	%tage	AA	%tage
Ala	7.5%	Arg	16.7%	Asn	10.0%
Asp	1.7%	Cys	6.7%	Gln	2.5%
Gly	4.2%	Ile	8.3%	Leu	8.3%
Met	1.7%	Phe	4.2%	Pro	8.3%
Ser	4.2%	Thr	3.3%	Trp	1.7%
Tyr	1.7%	Val	9.2%	GLU	0.0%
HIS	0.0%			LYS	0.0%

Total number of negatively charged residues (Asp + Glu): 2; Total number of positively charged residues (Arg + Lys): 20; Aliphatic index: 99.08; Grand average of hydropathicity (GRAVY): 0.037; Number of amino acids: 120; Molecular weight: 13737.3; Theoretical pI: 11.74; The instability index (II) is computed to be 53.13

A high aliphatic index of ALLCE-AMP sequence indicated that proteins might be stable for a wide temperature range. Instability index (II) provides an estimate of stability of protein. A protein whose instability index is lesser than 40 is expected as stable, a value exceeding 40 expected that the protein may be unstable (Dehury *et al.*, 2013). Instability

index value for ALLCE-AMP was ranged from 53.13 which indicated ALLCE-AMP as unstable protein. The Grand Average hydropathicity (GRAVY) value for a protein or peptide was calculated as the amount of hydropathy values of all amino acids, distributed by the quantity of residues in the sequence. GRAVY indices of ALLCE-AMP was 0.037. This

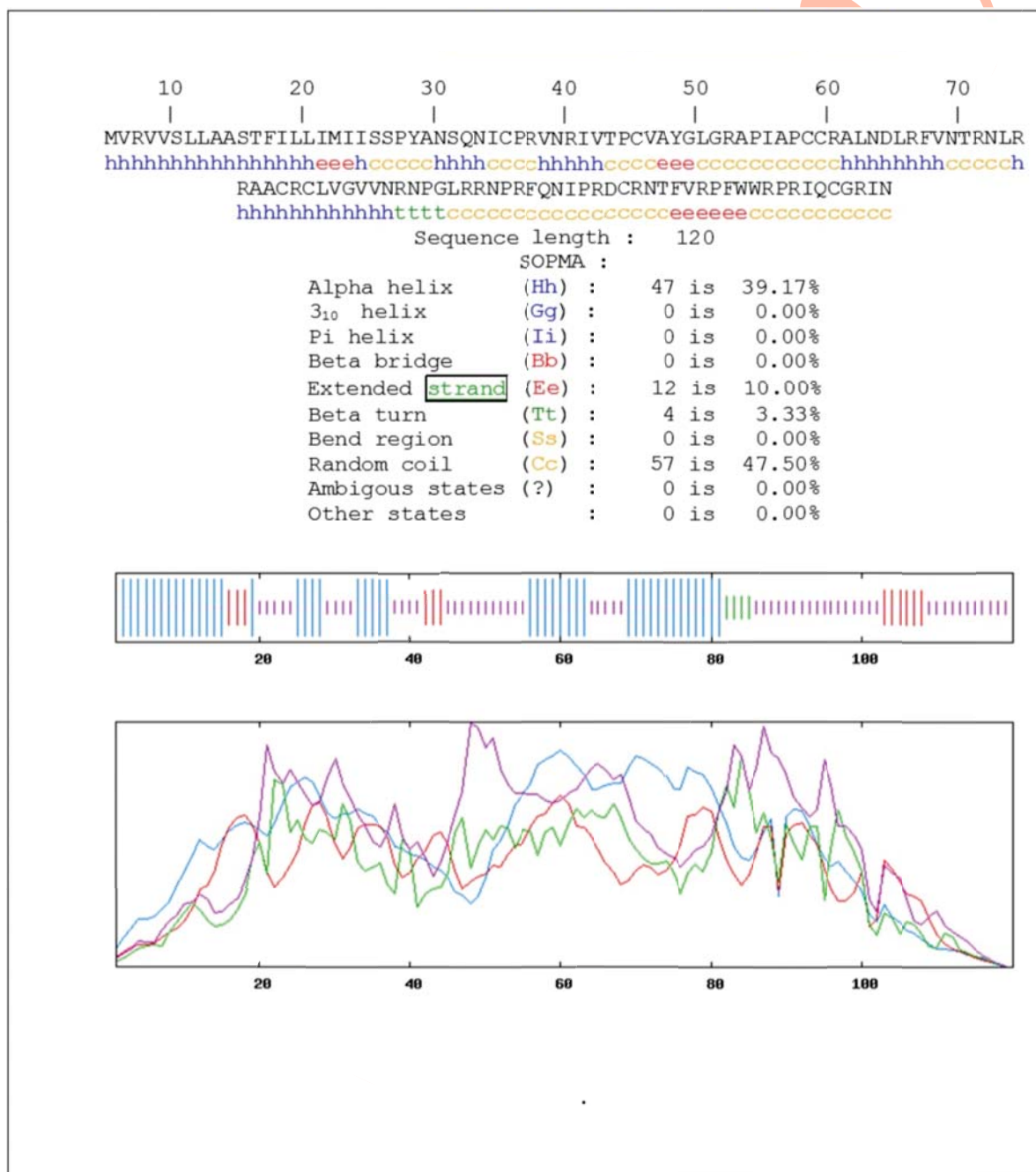


Fig. 1 : Represents secondary structure elements of ALLCE antimicrobial peptide sequence were predicted using SOPMA tool in Expsy

positive range of value indicated the possibility of worst interaction with water (Table 1). Construction of alpha helix and beta sheet were entitled as secondary structure of protein. It was made up of two networks: A structure-to-structure network and a sequence-to-structure network. Secondary structure of protein sequence were predicted using SOPMA

tool in ExPasy. The results revealed that random coil (47.50 %) dominated among secondary structure elements and alpha helices (39.17 %) and extended strand (10 %) were presented at equal percentage (Fig.1).

Modelling of three dimensional structure of ALLCE-

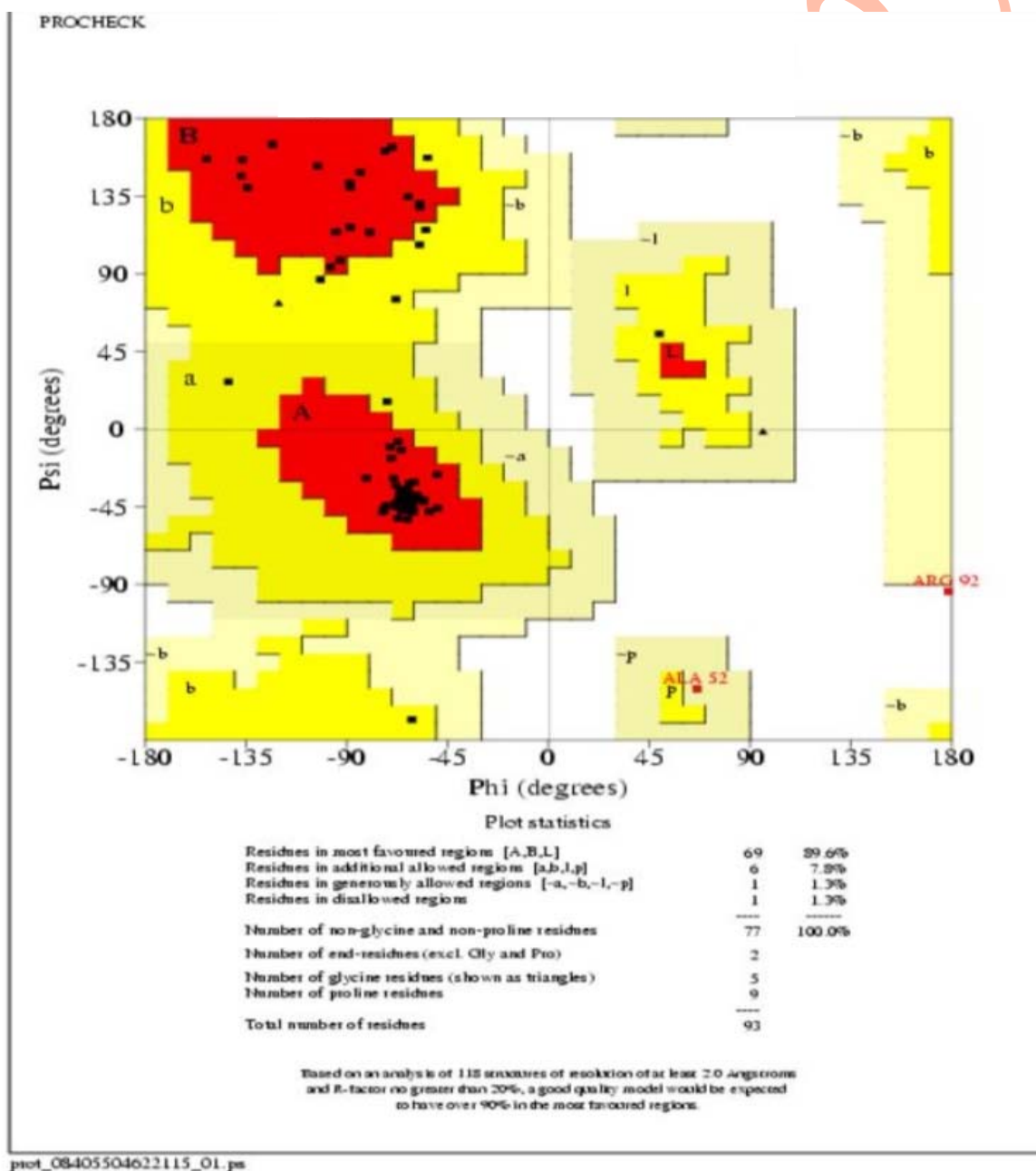


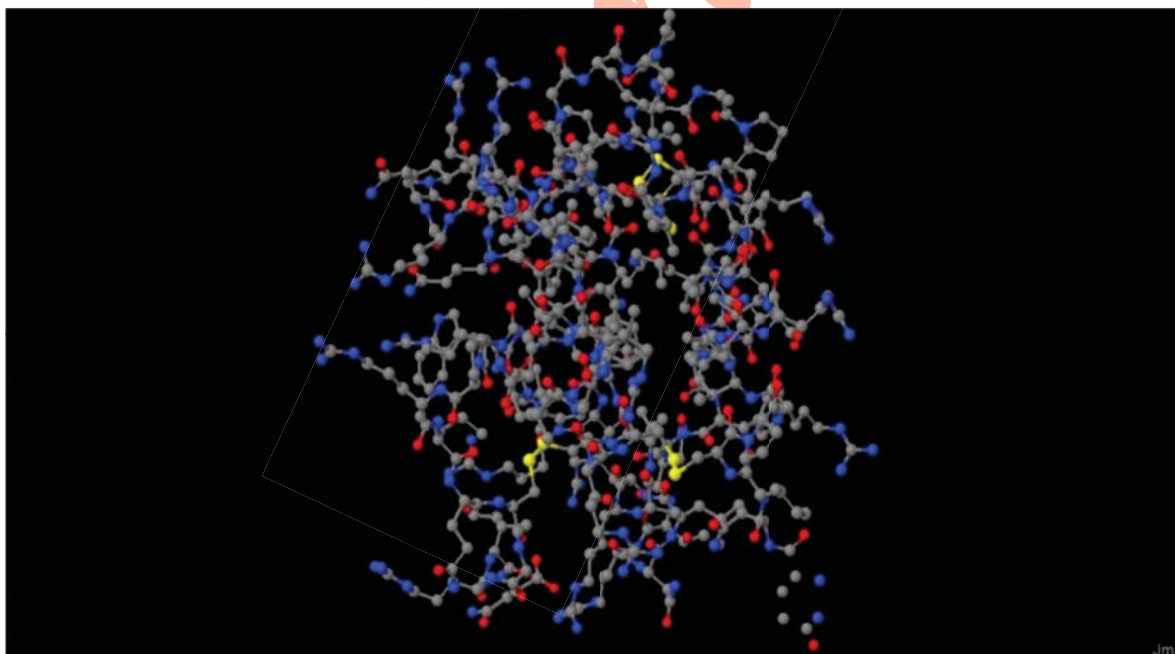
Fig. 2 : Represents the results of Ramachandran plot investigation of the ALLCE antimicrobial peptide sequence percentage residues were categorized according to their regions in the quadrangle

AMP was built by submitting the protein sequence (target) to ESYPred3D web server. ESYPred3D is a new automated homology modelling program. The method gets benefit of increased alignment performances of new alignment strategy by means of neural networks. Alignments were acquired by weighting, combining and screening the results of several multiple alignment programs. The absolute 3D structure was constructed using modelling package MODELLER. The predicted 3-D model of ALLCE-AMP PDB File and target-template alignment file attachment was received from ESYPred3D server. The template (1T12 chain A) be originate to share 18.2% identity with our Query (B2CZN8). The 3D structures of template and model are given in (Fig.3). Structural analysis was accomplished and figures illustrations were generated with Rasmol protein visualization tool (Aleksey *et al.*, 2010). Peptide/protein structure homology modeling, prediction was also identified as comparative modeling, is a class of procedures for constructing an atomic-resolution model of protein as of its amino acid sequence (Tripathi *et al.*, 2010). In the present study, roughly all the homology modeling techniques depended on identification of one or more known peptide/protein structures likely to resemble the structure of query sequence, and construction of alignment that maps residues in the query sequence to residues in the template sequence (Messaoudi *et al.*, 2011).

Accuracy of ALLCE-AMP model was validated using SAVES server and was judged by validity report

generated by PROCHECK (Kumar *et al.*, 2014). In the Ramachandran plot investigation, the residues were categorized according to their regions in the quadrangle. The Ramachandran map for ALLCE antimicrobial peptide is represented in Fig. 2. Homology protein modelling adopts experimentally determined protein structures to forecast the 3-D structure of one more protein that has comparable amino acid sequence (the target). This method of modelling is conceivable since a minor change in the protein sequence usually results in small modification in its 3D structure. Ramachandran plot investigation exhibited that main-chain conformations for 89.6 % of amino acid residue were inside the most favored or allowed region, 7.8 % in the allowed and 1.3% in the generously allowed region and only 1.3% in the disallowed region. In general, a score close to 100% implies virtuous stereo chemical superiority of the model (Reddy *et al.*, 2006). ERRAT is a protein structure confirmation algorithm that is particularly well-suited for evaluating the progress of crystallographic model construction and enhancement. The program works by investigating the statistics of non-bonded interactions between different atom types (Colovos and Yeates, 2002). The errat value was found to be 62.353, which was considered as a good quality protein.

The Target P datasets is a tool used for training was achieved from the TargetP web site. These datasets comprise a total of 3678 proteins indicating four plant and three non-plant localizations. The results of Target P for signal peptide (SP), mitochondrial pointing peptide, chloroplast transit



**Fig. 3 :** The predicted 3-D model of ALLCE-AMP received from ESYPred3D server

**Table 2 :** Target P prediction results representing number of query sequence and sub cellular location of ALLCE antimicrobial peptide sequence

### targetp v1.1 prediction results ###							
Number of query sequences: 1							
Cleavage site predictions not included.							
Using PLANT networks.							
Name	Len	CTP	mTP	SP	other	Loc	RC
tr_B2CZN8_B2CZN8_ALL	120	0.028	0.006	0.813	0.042	S	2

**Table 3 :** Psort prediction results showing probability of protein localization sites based on the certainty value of the ALLCE antimicrobial peptide sequence

plasma membrane	—	Certainty= 0.730(Affirmative)	< succ>
endoplasmic reticulum (membrane)	—	Certainty= 0.640(Affirmative)	< succ>
endoplasmic reticulum (lumen)	—	Certainty= 0.100 (Affirmative)	< succ>
outside	—	Certainty= 0.100(Affirmative)	< succ>

peptide and any other location (Other) specificity with 0.95 cut off value were reported to be 0.028, 0.006, 0.813 and 0.042, respectively. Reliability class 2 which indicates strongest prediction and prediction of localization was found to be more at signal peptide with 0.813 specificity as shown in Table 2. A neural network-based device, TargetP, used for important subcellular location prediction of newly identified proteins. Using N-terminal sequence information only, it categorize between proteins destined for the mitochondrion, the secretory pathway, the chloroplast, and "other" localizations with a achievement rate of 85% (plant) or 90% (non-plant) on redundancy-reduced test sets. Recently in TargetP investigation of sequenced *Arabidopsis thaliana* chromosomes 2 and 4 and the Ensembl *Homo sapiens* protein were found to estimate 10% of all plant proteins were mitochondrial and 14% chloroplastic, and that the plenty of secretory proteins, together with *Arabidopsis* and *Homo*, was around 10%. TargetP also forecasted cleavage sites with levels of suitably predicted sites ranging from 40% to 50% and above 70%.

Psort is a computational program for the forecast of protein localization in cells. It receives information of an amino acid sequence and its source origin . It analyzes the input sequence by applying the stored rules for various sequence features of recognized protein sorting signals. Finally, it reports the possibility for input protein to be localized at each candidate site with supplementary information. PSORT showed the probability of ALLCE-AMP localization sites in plasma membrane (0.730), endoplasmic reticulum (membrane) (0.640), endoplasmic reticulum (lumen)(0.100) and outside (0.100) based on the

certainty of 0.73 as given in Table 3. Protein localization is important because it supports a proposed role of a particular protein or peptide.

In conclusion, based on the template structure it is evident that the theoretical structure created was structurally similar to the template structure, which was sufficient for the development of specific ligand for ALLCE antimicrobial peptide. Subcellular localization is important for understanding peptides/protein function and plays a critical role in gene annotation. The predicted 3D structure of ALLCE-AMP PDP file were submitted to the Protein Model Database(PMDB) assigned the PMDB ID PM0078074.

### Acknowledgments

We are grateful to the Director General of Central Council for Research in Unani Medicine, Ministry of health and family welfare, New Delhi, and Department of Laboratory Sciences, College of Applied Medical sciences, King Saud University for encouragement and providing facilities for carrying out present study.

### References

- Amlie-Lefond, C., D.A. Paz, M.P. Connelly, G.B. Huffnagle, N.T. Whelan and H.T. Whelan: Innate immunity for biodefense: a strategy whose time has come. *J. Allergy Clin. Immunol.*, **116**, 1334–1342 (2005).
- Aleksey P. and J. Meller: POLYVIEW-MM: web-based platform for animation and analysis of molecular simulations. *Nucl. Acids Res.*, **38**, 662–666 (2010).
- Colovos, C. and T.O. Yeates: Verification of protein structures: patterns of non-bonded atomic interactions. *Protein Sci.*, **2**, 1511–1519 (2002).
- Dehury, B., M. Sahu, K. Sarma, J. Sahu, P. Sen, M.K. Modi, G.D. Sharma, M.D. Choudhury and M. Baroah: Molecular phylogeny, homology modeling, and molecular dynamics simulation of race-specific bacterial blight disease resistance protein [xa5] of rice: A comparative agriproteomics approach. *OMICs*, **17**, 423–438 (2013).
- Emanuelsson, O., H. Nielsen, S. Brunak and G.V. Heijne: Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Bio.*, **300**, 1005–1016 (2000).
- Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel and A. Bairoch: Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook* (Ed.: John Walker M.). Humana Press., pp. 571–607 (2005).
- Hoglund, A., P. Donnes, T. Blum, H. Adolph and O.W. Kohlbacher. MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics.*, **22**, 1158–1165 (2006).
- Kumar, P.S., Y.N. Kumar, U.V. Prasad, S. Yeswanth, V. Swarupa, D. Vasu, K. Venkatesh, L. Srikanth, V. K. Rao and P.V.G.K. Sarma: Comparative structural and functional analysis of *staphylococcus aureus* glucokinase with other bacterial glucokinases. *Ind. J. Pharm. Sci.*, **76**, 430–436 (2014).

- Ai, Y. and R.L. Gallo: AMPed up immunity: How antimicrobial peptides have multiple roles in immune defense. *Trends Immunol.*, **30**, 131–141 (2009).
- Lambert, C., N. Leonard and E. De Bolle Xand Depiereux: ESyPred3D: Prediction of protein 3D structure: *Bioinformatics.*, **18**,1250-1256 (2002).
- Liu, Y., J. Luo, C. Xu, F. Ren, C. Peng, G. Wu and J. Zhao: Purification characterization and molecular cloning of the gene of a seed-specific antimicrobial protein from pokeweed: *Plant Physiol.*, **122**,1015–1024 (2000).
- Messaoudi, A., H. Belguith and J.B. Hamida: Three-Dimensional Structure of Arabidopsis thaliana Lipase Predicted by Homology Modeling Method: *Evol. Bioinform Online.*, **7**, 99–105 (2011).
- Mendez-Samperio, P., L. Alba and A. Trejo: Mycobacterium bovis-mediated induction of human beta-defensin-2 in epithelial cells is controlled by intracellular calcium and p38MAPK. *J. Infect.*, **54**, 469–474 (2007).
- Murakami, M., T. Ohtake, R.A.B. Dorschner, Schitteck, C. Garbe, R.L. Gallo: Cathelicidin anti-microbial peptide expression in sweat, an innate defense system for the skin. *J. Invest. Dermatol.*, **119**, 1090–1095 (2002).
- Nikoletta, H. and F. Marx: Antifungal proteins: More than antimicrobials?. *Fungal Biol. Rev.*, **26**, 132–145 (2013).
- Reddy, C.H., K. Vijayasathy, E. Srinivas, G.M. Sastry and G.N. Sastry: Homology modeling of membrane proteins: A critical assessment: *Comput Biol Chem.*, **30**,120 (2006).
- Rost, B., J. Liu, R. Nair, K.O. Wrzeszczynski and Y. Ofran :Automatic prediction of protein function: *Cell Mol Life Sci.*, **60**, 2637–2650 (2003).
- Sang, O. P., J.H. Shin, W.K. Choi, B.S. Park, J.S. Oh and A. Jang: Antibacterial activity of house fly-maggot extracts against MRSA (Methicillin-resistant *Staphylococcus aureus*) and VRE (Vancomycin-resistant enterococci). *J. Environ. Biol.*, **31**, 865-871 (2010).
- Schauber, J., R.A. Dorschner, K. Yamasaki, B. Brouha and R.L. Gallo: Control of the innate epithelial antimicrobial response is cell-type specific and dependent on relevant microenvironmental stimuli. *Immunology*, **118**, 509–519 (2006).
- Sharma, A. and A. Nigam: Structure modeling of novel DNA glycosylase enzyme from oral pathogen *Streptococcus sanguinis*. *Bioinformation*, **5**, 136–140 (2010).
- Tossi, A., L. Sandri and A. Giangaspero: Amphipathic alpha-helical antimicrobial peptides: *Biopolymers*, **55**, 4–30 (2000).
- Tian Z., T. Dong, D. Teng, and J. Yang Yand Wang: Design and characterization of novel hybrid peptides from LFB15[W4,10], HP[2-20] and cecropin a based on structure parameters by computer-aided method. *Appl. Microbiol. Biotechnol.*, **82**, 1097–1103 (2009).
- Tripathi, V., R. Sinha and D.K. Gupta: Molecular docking study on hemagglutinin protein of H1N1 virus with recommended antiviral drugs: *Der Pharma Chemica.*, **2**, 53-59 (2010).
- Wolk, K., S. Kunz, E. Witte, M. Friedrich, K. Asadullah and R. Sabat: IL-22 increases the innate immunity of tissues. *Immunity.*, **21**, 241–254 (2004).
- Wilson, N.J., K. Boniface, J.R. Chan, B.S. McKenzie, W.M. Blumenschein, J.D. Mattson, B. Basham, K. Smith, T. Chen, F. Morel, J.C. Lecron, R.A. Kastelein, D.J. Cua, T.K. McClanahan, E.P. Bowman and R. de Waal Malefyt: Development, cytokine profile and function of human interleukin 17-producing helper T cells. *Nat. Immunol.*, **8**, 950–957 (2007).