VNU Journal of Science, Natural Sciences and Technology 28 (2012) 28-36

Identification and Sequencing analysis of a P68 DEAD-box RNA helicase from *Pisum sativum*

Pham Xuan Hoi[1,*], Narendra Tuteja[2]

[1]*The Institute of Agricultural Genetics, 2 Pham Van Dong, Hanoi, Vietnam*
[2]*The International Centre for Genetic Engineering and Biotechnology, New Delhi, India*

Received 27 December 2011

Abstract. Helicases catalyse the unwinding of energetically stable duplex DNA (DNA helicase) or inter - and intra molecular base -paired duplex RNA (RNA helicase) structures by disrupting the hydrogen bonds between the two strands and thereby plays an important role in all DNA/RNA metabolisms. Many DNA and RNA helicases share a core region of highly conserved sequence motifs, and belong to the rapidly growing DEAD-box protein family that contains the same eight conserved helicase motifs. Using 1.93 kb cDNA fragment of P68 DEAH box protein from *Arabidopsis thailiana* as probe for screening pea cDNA library, we identified a full length cDNA of p68 DEAH box protein that has 2058 bps with poly[A^+] tail of 28 nucleotides at the 3' end. It contains coding region of 1869 bps, 5'-end untranslated region of 53 bps and 3'-end untranslated region of 136 bps. The deduced amino acids sequence revealed a protein consisting of 623 amino acid residues with a predicted molecular mass of about 68 kDa (p68). All 8 helicase conserved domains have been observed in amino acid sequence of the protein. The nucleotide sequence alignment of Pea P68 DEAH box and homolog p68 DEAH box from different species shows that Pea P68 DEAH box has striking homology with soybean, castorbean and tomato. The deduced amino acid sequence of P68 was used for searching similar sequences with other two pea DEAH box proteins (pdh45 and p72) by using FASTA computer program reveals a common core-region around 300 amino acids that contains all the known conserved helicase domains and localizes in the middle of the genes. At molecular level, the DEAD-box RNA helicases function in process such as transcriptional regulation, regulation of RNA stability, ribosome biogenesis and post-translational regulation.

Keywords: P68, DEAD-box protein family, RNA helicase, Pisum sativum.

1. Introduction

The majority of RNA helicases belong to the superfamily 2 (SF2) subclass of helicases characterized by sequence homology within a helicase domain consisting of eight conserved amino acid motifs. SF2 consists of three subfamilies, termed DEAD, DEAH and DExH/D, based on variations within a common DEAD (Asp-Glu-Ala-Asp) motif. Amino acid sequences outside this 'core' helicase domain are not conserved and are believed to provide helicase specificity for target RNAs or protein-protein interactions [1]. DEAD box protein has

* Corresponding author. Tel: 84-4-37557764.
E-mail: xuanhoi.pham@gmail.com

been defined by Linder et al. [2] and originated from a wide range of organisms ranging from prokaryotes, including viruses, to lower and higher eukaryotes. They are involved in variety of RNA metabolic processes, such as RNA maturation functioning in ribosome biogenesis, RNA splicing, transport, and turnover, transcription, translation initiation, RNAi, RNA editing, and development [3]. Recently, it is becoming increasingly evident that RNA helicases are associated with a diverse range of biotic cellular functions and there have been relatively a lot reports of RNA helicase involvement in cellular response to abiotic stress [4].

P68 is one of the prototypic members of the DEAD-box family of ARN helicases, which includes a large number of proteins that participate in virtually all processes involving RNA metabolism [5]. Several reports have demonstrated that P68 expression is growth and development regulated [6], while P68 knockout in mice results in embryonic lethality at approximately embryonic day 11.5 (~ E11.5), underscoring the importance of this protein. Additionally, P68 is aberrantly expressed and/or post-translationally modified in a range of cancers, suggesting that changes in P68 levels and/or function may be important in cancer development [7]. As a transcriptional co-regulator, P68 has been found to coactivate several transcription factors that are themselves highly regulated [8], the tumor suppressor p53 [9].

In higher plant, a number of RNA helicase genes whose expression and polyadenylation patterns are tissue specific. However, detailed characterization of plant RNA helicases are less common. Two DEAD-box – related helicases, termed pea DNA helicase 47 (PDH47) and 45 (PDH45) are induced by a variety of abiotic stresses, suggesting that they are components of a general stress response mechanism.PDH47 expression is differentially induced in a tissue specific manner with induction by cold and salinity stress in shoots and roots and head and ABA treatment in roots [10]; while PDH45 transcript is induced in pea seedlings in response to a range of abiotic stresses including salt, dehydration, wounding and low temperature, leading to the suggestion that pdh45 transcript accumulates in response to general water stress caused by desiccation. The physiological importance and conservation of PDH45 function in the salt-stress response was demonstrated by the observation that constitutive expression of PDH45 conveys salt tolerance in tobacco [11].

2. Materials and Methods

Plant Material and Growth Conditions

Seeds of *Pisum sativum* were grown at 25^0C± 2^0C for 7-10 days in light /dark conditions in trays containing moist vermiculite in the greenhouse. The harvested material was frozen in liquid nitrogen until use.

Construction of Pea cDNA Library

Total RNA was isolated from top four leaves of 7-day old pea (*Pisum sativum*) seedling using guanidinium isothiocyanate method (Tuteja and Farber, 1988). Poly(A) RNA was purified from total RNA by fractionation on an oligo (dT)-cellulose type 7 column (Pharmacia, Uppsala, Sweden). A cDNA library was constructed from 5 µg of poly (A) RNA in Uni-Zap XR vector using Zap-cDNA synthesis kit (Stratagene, La Jolla, CA) following the manufacturer's protocol.

The cDNA was cloned between the *Eco*RI and *Xho*I sites of the vector followed by packaging with Gigapack II gold extracts and amplification according to the manufacturer's instructions (Stratagene). The resulting phage library contained 1 x 10^9 plaque forming units per ml.

Preparation of Probe for Screening of Pea cDNA Library

Gene encoding P68 RNA helicase (1.9 kb) from *Arabidopsis thaliana* was obtained by *EcoR*I digestion. DNA fragment was cut out from agarose gel and purified by phenol/chloroform method. Nick translation reaction was carried out in a 50 µl reaction mixture containing 50-100 ng DNA template, 5 µl dNTPs without dCTP, 5 µl $\propto^{32}$PdCTP, and 5 µl of DNA Polymerase I/DNase I mixture for 1 hour at 16 ^{0}C. Reaction was stopped by adding 5 µl of stop reaction buffer. Reaction mixture was purified through Sephadex G-25 column and purified probe was used for screening of pea cDNA library.

Screening of Pea cDNA Library

About 2 µl of pea cDNA library suspension containing 20,000 plaque-forming bacteriophage was infected into 200 µl of XLI-Blue MRF cells at an OD of 0.5 in falcon 2059 polypropylene tubes and incubated for 15 min at 37 ^{0}C with gentle shaking. Infected cells were mixed well with 3 ml of melted NZY Top agar at 48 ^{0}C and spread onto NZY agar plates (100 x 15 mm) at 42 ^{0}C and incubated at 42 ^{0}C for 6 - 8 hours. Hybond filters (Amersham) were placed on the surface of the chilled plates without trapping air bubbles and a needle was used to prick through the NZY and nylon membrane for orientation. The filters were removed carefully with the help of Millipore forceps and placed inverted plaques side up onto 3 MM Whatman sheets moistened with denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 3 min. The filters were transferred to a Whatman sheet soaked in neutralizing solution (1.5 M NaCl, 0.5 M Tris-HCl, pH 8) for 5 min. and then rinsed on a Whatman sheet soaked in neutralizing solution (0.2 M Tris – HCl, pH 7.5, 2X SSC) for 5-10 min. The DNA on the filter was dried and fixed by crosslinking (12,000 µjoules of UV energy) for 30 sec. The filters were put in a hybridization bag and rinsed thoroughly with 5X SSC buffer and then all the SSC was removed. The prehybridization solution containing 6X SSC, 20 mM NaH_2PO_4, 0.4 % w/v SDS, 5X Denhardt's reagent, and 100 µg/ml denatured salmon sperm DNA was added and prehybridization was carried out for 4-6 hours at 45 ^{0}C. After that, the purified probe was added and hybridized at 45 ^{0}C for 15-20 hours. The filters were washed with 2X SSC, 0.1 % SDS for 15 min at room temperature and then the temperature was increased stepwise to 37, 40, and 42 ^{0}C, simultaneously decreasing concentration of SSC to 1, 0.5 and 0.25 X respectively while checking radioactivity counts every 15 minutes until background became negligible. The filters were then exposed for autoradiography. The positive clones obtained from first screening were oriented on the agar plates using the numbers and "dots" where the needle poked through, which had been marked on the membranes. One ml pipette tips were cut out to obtain a square centimetre "window" for transferring the positive plaque into 1ml of SM buffer and 20 µl chloroform. Stock solution was titered before each screening to get a ratio of around 200 plaques/plate. Screening is continued until positive clones getting pure. PCR amplification and Southern hybridization

were used during screening to confirm and determine the purity and the size of each positive clone.

3. Results

Cloning and Sequencing of Pea p68 DEAH-box Protein

A cDNA library constructed in λ ZAP II (Stratagene) prepared from leaves of eight–day old light/dark grown pea seedlings was used for isolation of the gen encoding *Pea p68 DEAH-box Protein*. The probe used for screening was 1.93 kb cDNA fragment of P68 DEAH box protein from *Arabidopsis thailiana* (kindly provided by Tetsuo Meshi of Kyoto University of Japan). Screening of 1.5 x 10^5 plaques from pea cDNA library for the purpose of isolating the RNA helicase revealed two positive clones (fig.1A). These positive recombinant clones were purified independently to homogeneity after three rounds of screening (Fig. 1A, B, C).

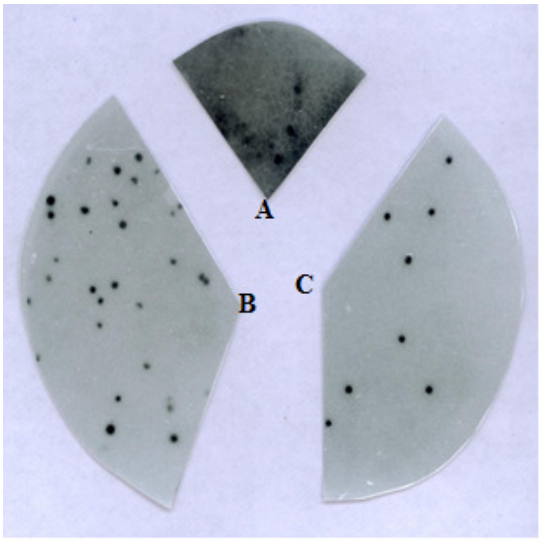

Figure 1: Screening of cDNA library from Pisum sativum. The cDNA library constructed in Lambda Zap II vector was screened using . A total of 1.5x105 flaques was screened and positive plaques were isolated and purified till homogeneity. (A) primary, (B) secondary, (C) final screening.

The inserts were excised out as pBluescript (SK-) phagemid and amplified by PCR using T3 and T7 primers and PCR products were subjected to Southern hybridization. Results show that both clones 1 and 2 yielded fragment size of 2.3 kb (fig 2A, B). The cDNA clones were digested with selected enzymes (*Xho* I; *Bam* H1; *Not* I; *EcoR* I; *Bgl* II; *Xba* I; *Sma* I and *Hind* III) and digested plasmids were fractionated on 0.9% agarose gene and stained with ethidium bromide. Restriction enzymes digestion of these clones revealed that both clones contain the same restriction sizes, suggesting that both positive clones are the same (data not shown).

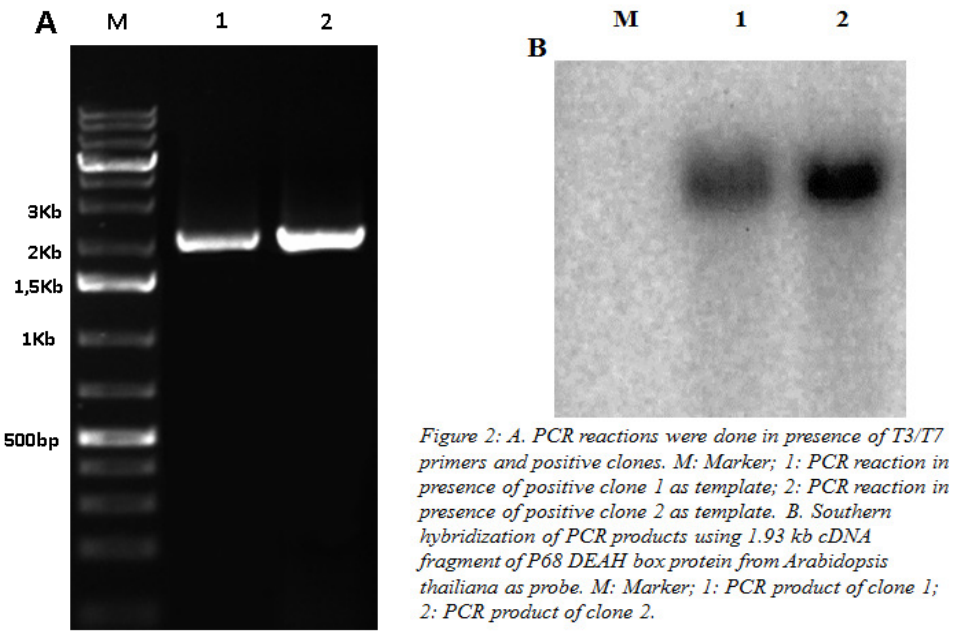

Figure 2: A. PCR reactions were done in presence of T3/T7 primers and positive clones. M: Marker; 1: PCR reaction in presence of positive clone 1 as template; 2: PCR reaction in presence of positive clone 2 as template. B. Southern hybridization of PCR products using 1.93 kb cDNA fragment of P68 DEAH box protein from Arabidopsis thailiana as probe. M: Marker; 1: PCR product of clone 1; 2: PCR product of clone 2.

Based on the restriction sites, clone 1 was subjected for sequencing using T3, T7 primers. Sequence analysis shows that the clone 1 encoded a full length cDNA of p68 DEAH box protein. Pea P68 DEAH box protein has 2058 bps with poly[A^+] tail of 28 nucleotides at the 3' end. It contains coding region of 1869 bps, 5'-end untranslated region of 53 bps and 3'-end untranslated region of 136 bps. These sequence data have been submitted to the DDBJ/EMBL/GenBank databases under accession number AF271892. The deduced amino acids sequence revealed a protein consisting of 623 amino acid residues with a predicted molecular mass of about 68 kDa (p68). All 8 helicase conserved domains have been observed in amino acid sequence of the protein (fig. 3).

The nucleotide sequence alignment of Pea P68 DEAH box protein with other p68 DEAH box proteins from difference species using Blast search show that Pea P68 DEAH box protein has maximum of 92% identity with **soybean** (XR136957.1), 79% with **castorbean** (XM 002523901.1) and 72% with tomato (BT 013308.1). The deduced amino acid sequence of Pea P68 DEAH box protein was used for searching similar sequences with other two new pea DEAH box proteins (pdh45 and p78 DEAH box protein) by using FASTA computer program (Fig.4). Although different in size, the alignment among three pea DEAH box proteins reveals a common core-region around 300 amino acids that contains all the known conserved helicase domains and localizes in the middle of the genes. Homology search revealed that pea p68 and p72 DEAD-box proteins have DEAD-box, while pdh45 has a DESD-box (domain VI); three DEAD-box proteins (p68, p72 and pdh45) have different SAT-box respectively as SAT, TAT and SRT-box (domain VI). A region of rich Glycine and Arginine (G and R) sequences with many RGG like domains was observed at the C terminal of the sequence was only observed in p72 DEAD box. In the core-region of the pea DEAD box

proteins reveal a 40 % identity of amino acid whereas in the amino - and carboxy-terminal parts (divergent region), high variability of amino acid sequence as well as length of polypeptide is observed. Sequence Alignment and Homology analysis among Pea DEAH-box Proteins suggest that they have specific function in DNA/RNA metabolism.

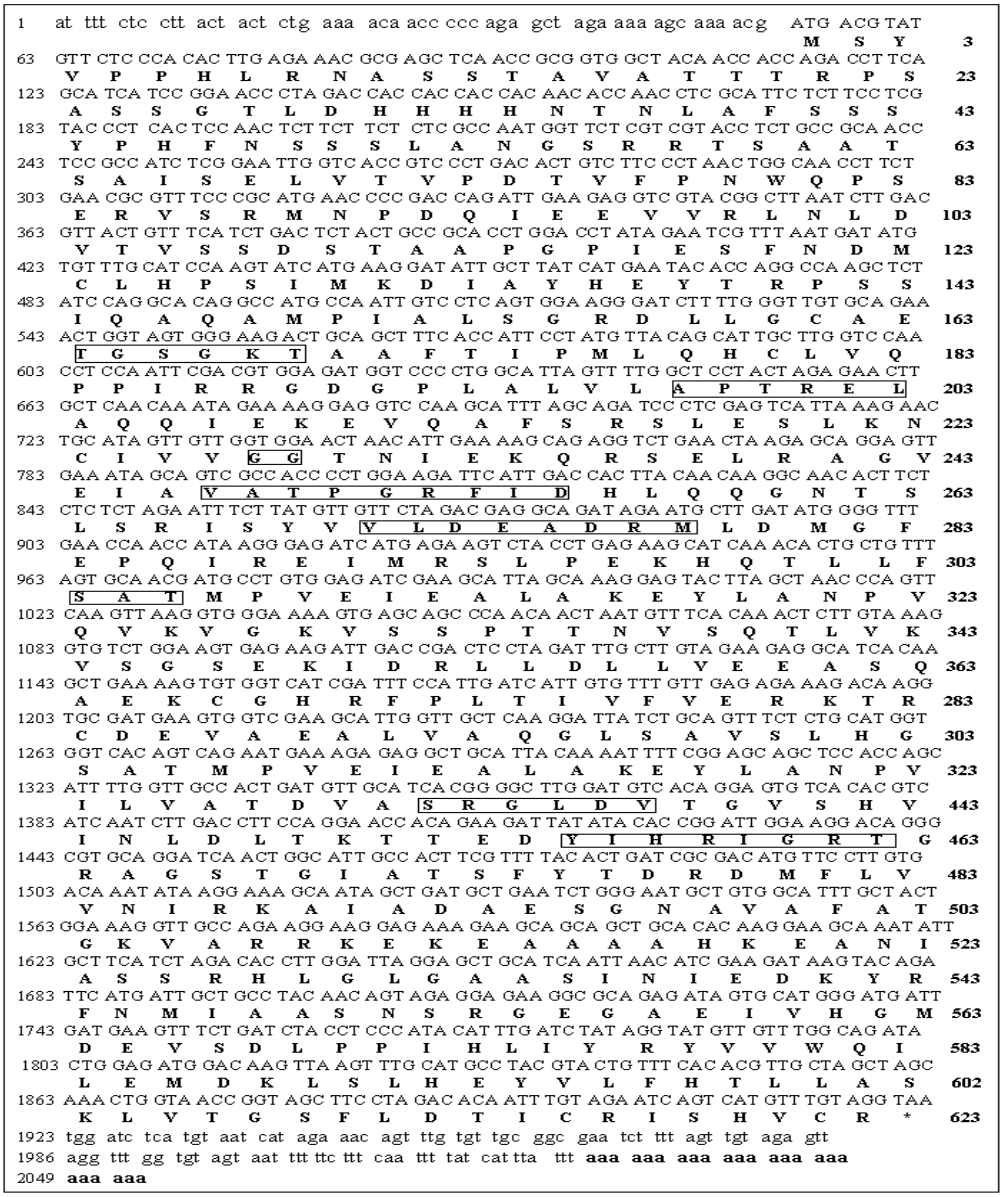

Figure 3: Nucleotide and deduced amino acid sequence of the p68 DEAD box protein (Accession number: AF 271892). ORF is shown in capital letters and 5' and 3'- untranslated regions are shown in small letters. Eight conserved helicase domains are shaded. Position of the nucleotide or the amino acid residue in a sequence are indicated on both sites of the sequence.

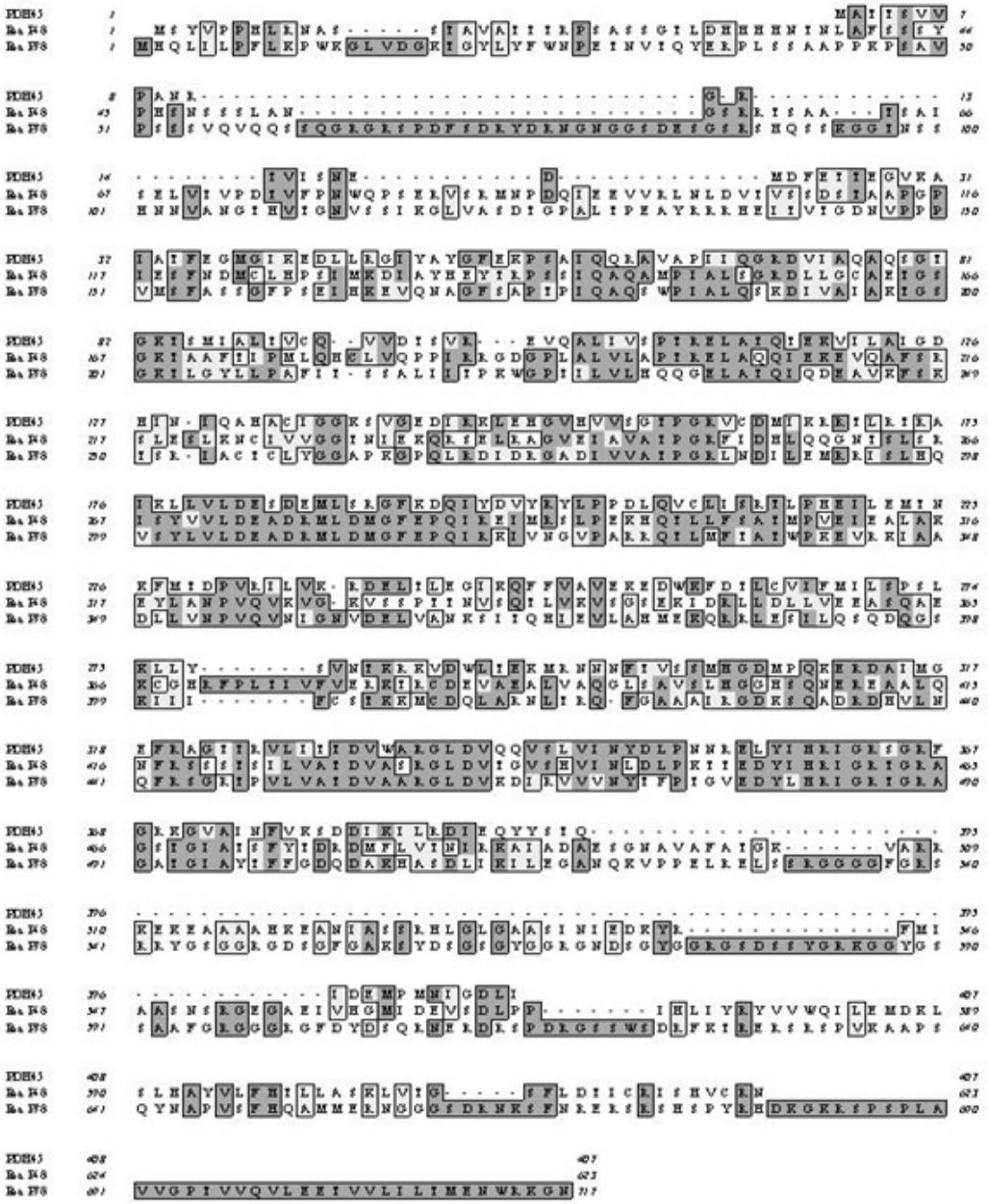

Figure 4: Multi-sequence alignment of deduced amino acid sequence of three pea DEAD box protein (pdh45, p68 and p78). Multiple alignment was done using Macvector Clustalw program. The most identical amino acids at each protein are black-boxed and similar ones are shaded. The position of the amino acid residues in original sequence are indicated on both side.

Acknowledgements

This research was financially supported by TWAS (Academy of Sciences for the Developing world) through project No. 09-235 RG/Bio/A S_G – UNESCO FR: 3240230329.

References

[1] N. K. Tanner and P. Linder, DExD/H boa RNA helicases: from genetic motors to specific dissociation function. *Mol. Cell.* 8 (2001) 251.

[2] P. Linder, P. F. Lasko, M. Ashburner, P. Leroy, P. J. Nielsen, K. Nishi, J. Schneir and P. P.Slonimski. Birth of the DEAD-box. *Nature* 337 (1989) 121.

[3] S. Rocak and P. Linder, DEAD-box proteins: the driving forces behind RNA metabolism. *Nature Rev. Mol. Cell Biol.* 5 (2004) 232

[4] G. W. Owttrim. Survey and summary RNA helicases and abiotic stress. *Nucleic Acids Res.,* 34 (2006), 3220.

[5] P. Linder, DEAD-box proteins: a family affair-active and passive players in RNP-remodeling. Nucleic Acids *Res.* 34 (2006) 4168.

[6] K. Kahlina, I. Goren, J. Pfeilschifter, S. Frank. P68 DEAD box RNA helicase expression in keratinocytes. Regulation, nucleolar localization, and functional connection to proliferation and vascular endothelial growth factor gene expression. *J. Biol. Chem.* 279 (2004) 44872.

[7] E. L. Clark, A. Coulson, C. Dalgliesh, P. Rajan, S. M. Nicol, S. Fleming, R. Heer, et al. The RNA helicase 68 is a novel androgen receptor coactivator involved in splicing and is overexpressed in prostate cancer. *Cancer Res.* 68 (2008) 7938.

[8] H. Endoh, K. Maruyama, Y. Masuhiro, Y. Kobayashi, M. Goto, H. Tai, J. Yanagisawa, D. Metzger, S. Hashimoto, S. Kato. Purification and identification of P68 RNA helicase acting as a transcriptional coactivator specific for the activation function of human estrogen receptor. *Mol. Cell Biol.* 19 (1999) 5363.

[9] G. J. Bates, S. M. Nicol, B. J. Wilson, A. M. Jacobs, J. C. Bourdon, J. Wardrop, D. J. Gregory. D. P. Lane, N. D. Perkins, F. V. Fuller-Pace. *EMBO* 24 (2005) 543.

[10] A. A. Vashisht, A. Pradhan, R. Tuteja and N. Tuteja. Cold and salinity stress-induced bipolar pea DNA helicase 47 is involved in protein synthesis and stimulated by phosphorylation with protein kinase C. *Plant J.* 44 (2005) 76.

[11] N. Sanan-Mishra, X. H. Pham**,** S. K.Sopory and N. Tuteja. Pea DNA helicase 45 overexpression in tobacco confers high salinity tolerance without affecting yield. *Proc Natl Acad Sci.* USA. 102 (2005) 509.

Phân lập và phân tích trình tự gen mở xoắn ARN của một ARN helicase phân nhóm DEAD-box P68 từ đậu Hà Lan

Phạm Xuân Hội[1], Narendra Tuteja[2]

[1]*Viện Di truyền Nông nghiệp, số 2 Phạm Văn Đồng, Hà Nội, Việt Nam*
[2]*Trung tâm Kỹ thuật gen và Công nghệ sinh học Quốc tế, New Delhi, Ấn Độ*

Helicaz xúc tác việc mở xoắn cấu trúc sợi đôi ADN hoặc ARN bằng việc bẻ gẫy các cầu liên kết hydrogen giữa hai sợi vì vậy đóng vai trò quan trọng trong tất cả các quá trình trao đổi chất ADN/ARN . Rất nhiều ADN và ARN helicaz có chung một vùng trung tâm gồm các kiểu trật tự bảo thủ cao thuộc protein họ DEAD-box chứa 8 kiểu Helicaz bảo thủ. Sử dụng một đoạn cDNA 1.93 kb của protein DEAD-box P68 từ cây mô hình *Arabidopsis* như đầu dò để sàng lọc thư viện cDNA đậu

Hà Lan, chúng tôi đã phân lập một cDNA đầy đủ của protein DEAD-box P68 gồm 2058 bps với một đuôi poly[A] 28 nucleotide ở đầu 3'. cDNA của protein DEAD-box P68 gồm một vùng mã hóa 1869 bps, vùng không mã hóa đầu 5' 53 bps và vùng không mã hóa đầu 3' 136 bps. Trình tự amino acid suy đoán bộc lộ một protein gồm 623 amino acid với trọng lượng phân tử dự đoán khoảng 68 kDa(p68). Tất cả 8 vùng bảo thủ Helicaz đều được quan sát trong trình tự amino acid của protein. So sánh sự tương đồng về trật tự nucleotide của DEAD-box P68 đậu Hà Lan với các DEAD-box P68 tương đồng từ các đối tượng cây trồng khác nhau đã phát hiện DEAD-box P68 đậu Hà Lan tương đồng cao với đậu tương, thầu dầu và cà chua. Trình tự amino acid suy đoán của P68 được sử dụng để tìm kiếm các trật tự tương đồng với 2 protein DEAD-box từ đậu Hà Lan (pdh45 và p72) bằng chương trình phần mềm FASTA cho kết quả một vùng trung tâm chung gồm khoảng 300 amino acid chứa tất cả các vùng helicase bảo thủ được biết và định vị ở vùng trung tâm của các gene. Ở mức độ phân tử, các helicase ARN DEAD-box được chứng minh là có chức năng điều hòa phiên mã, điều hòa độ ổn định của ARN, sinh tổng hợp ribosome và điều hòa sau phiên mã.