Protein and Peptide Letters, Vol. 11, No. 4, pp. 345-352, 2004
Bentham Science Publishers Ltd.
0929-8665/04 $ 45.00 + .00

AN OPAQUE-2-LIKE TRANSCRIPTION FACTOR FROM PEARL MILLET

Lucilia Helena Marcellino[1], Goran Neshich[2], Antonio Furtado Neto[1] and Eugen Silvano Gander[1]*

[1] Embrapa-Recursos Genéticos e Biotecnologia – Cenargen
Parque Estação Biológica – PqEB Final W5 Norte, 70770-900 Brasília – DF, Brazil
(*): egander@cenargen.embrapa.br; Fax: (++ 55 61) 340 3658
Centro de BioInformática Estrutural, Embrapa-Informática Agropecuária, Cidade Universitária "Zeferino Vaz", 13083-970 Campinas -SP, Brazil

Abstract: Pearl millet (*Pennisetum glaucum* L.), a species of the *Poaceae* family, is an important food crop in Africa, Asia and South America. Its nutritional value is due to storage prolamins accumulated in the seeds. In other species of the same family, the expression of the genes coding for storage prolamins is mediated by the regulatory protein opaque-2. In this paper we show that an opaque-2 –like protein is present in pearl millet too and is expressed during the early stages of seed development. The organization of the gene coding for this protein is similar to that of orthologous genes in other *Poaceae* species, i.e. six exons separated by five introns. A comparison of amino acid homologies with other described opaque-2 proteins is presented.

Keywords: Transcription factors, opaque-2, pearl millet, storage proteins.

INTRODUCTION

The astonishing amount of information on eukaryotic genome organization, brought about by the various genome projects, has not yet been matched with a comparable amount of functional insights, neither with respect to structural genes and much less with respect to genes and proteins involved in the regulation of ordered gene expression. With respect to plant regulatory proteins, the bZIP protein opaque-2, originally identified in maize, is among the best investigated so far [1]. Within the *Poaceae* family, orthologous genes have been identified and characterized physical-chemically and biologically within various species (e.g. [2]). All these proteins so far described are characterized by a basic region which, in general, recognizes DNA sequences containing an ACGT or ACCT core [3] and a region containing 4 to 7 leucine residues spaced by 6 amino acids that mediates homo- and/or heterodimerization and is thought to be responsible for the biological activity of these regulatory molecules [4]. Functional analyses revealed that O2 is involved in the regulation of the expression of endosperm-specific prolamin genes in maize, in

Job's tears (coix) and in sorghum [5, 1, 6] as well as in the expression of the maize *b-32* gene [5] coding for a RIP protein.

Pearl millet (*Pennisetum glaucum* L.), another member of the *Poaceae* family, is another important food crop. It is tolerant to drought and can grow under very harsh conditions such as infertile soils and excessive heat, conditions that are unsuitable for maize or sorghum. These characteristics account for its importance in Africa and Asia. Like in almost all the major cereals, the most abundant seed storage protein fractions is composed of prolamins [7, 8] and since expression of maize-, corn- and sorghum-prolamin is mediated by opaque-2 proteins, we initiated a search for an orthologue gene in pearl millet. In this work we report the physical-chemical and biological characterization of an *opaque-2*-like pearl millet protein.

MATERIALS AND METHODS

Plant Material: Seeds of pearl millet were obtained from EMBRAPA's Maize and Sorghum Research Center (Sete Lagoas-MG, Brazil).

***Genomic DNA and partial cDNA Cloning*:** A genomic library was constructed by cloning *Sau*3AI fragments (9-23 kb) from pearl millet DNA into the λ dash vector (Stratagene®). Pearl millet opaque-2 clones were obtained by screening the library using maize opaque-2 cDNA as a probe.

For the cDNA cloning, total RNA from seeds at an early developmental stage (stage1) was used with the 3' and 5'RACE kit (Gibco®). Genomic and cDNA clones were sequenced.

Computer analyses: The sequences alignment and homology trees were performed using *DNA*MAN software (Lynnon BioSoft).

***Modeling of Protein Structure*:** Molecular modeling of pearl millet opaque-2 was carried out on a "Silicon Graphics Challenge R10000" workstation using the program "Modeller" [9]. The program 'Sting Millennium Suite' [10] was used for analysis and presentation of structure and sequence. The atomic coordinates of the GCN4 b-ZIP and basic region (Brookhaven Protein Data code:1ZIL.pdb [11], and 2DGC.pdb) were used to build the three-dimensional model of the basic and the ZIP region. Adequate superimpositions and substitutions, accompanied by energy minimization resulted in a modeled dimer in the case of the ZIP region. The energy minimization procedure used the steepest descent and conjugate gradient, set to run for up to 18.000 steps, or until the maximum derivative of the energy with respect to the atomic positions was less than 0.0002 kcal/mol/Å.

Western Blot: Proteins were extracted by grinding 0.3 g of leaves and seeds in 1 ml solution of 50 mM Tris-HCl (pH 6.8), 1% β-mercaptoethanol, 1 mM leupeptin and 2 mM PMSF. The homogenate was centrifuged at 4°C for 5 min at 12,000 g. Soluble proteins were separated on 15% SDS-polyacrylamide gels (SDS-PAGE), transferred to nitrocellulose membranes and probed with specific polyclonal rabbit antibodies raised against the bZIP domain of maize O2. Following immunodetection with anti-rabbit Ig horseradish peroxidase conjugate, the blot was developed according to ECL kit specifications (Amersham-Pharmacia®) and exposed to Hyperfilm ECL for 60 seconds.

RT-PCR: Total RNA from seeds at different developmental stages were prepared as described by Jones *et al.* [12]. The RT-PCR was performed using specific primers designed to generate a 250 bp fragment. The generated fragments were analyzed by Southern blots and expression levels were visually evaluated.

RESULTS AND DISCUSSION

Genomic DNA and cDNA cloning: The combined genomic and cDNA cloning allowed us to determine the gene's sequence and hence structure (Figure 1), as well as the complete amino acid sequence of the mature protein (Figure 2).

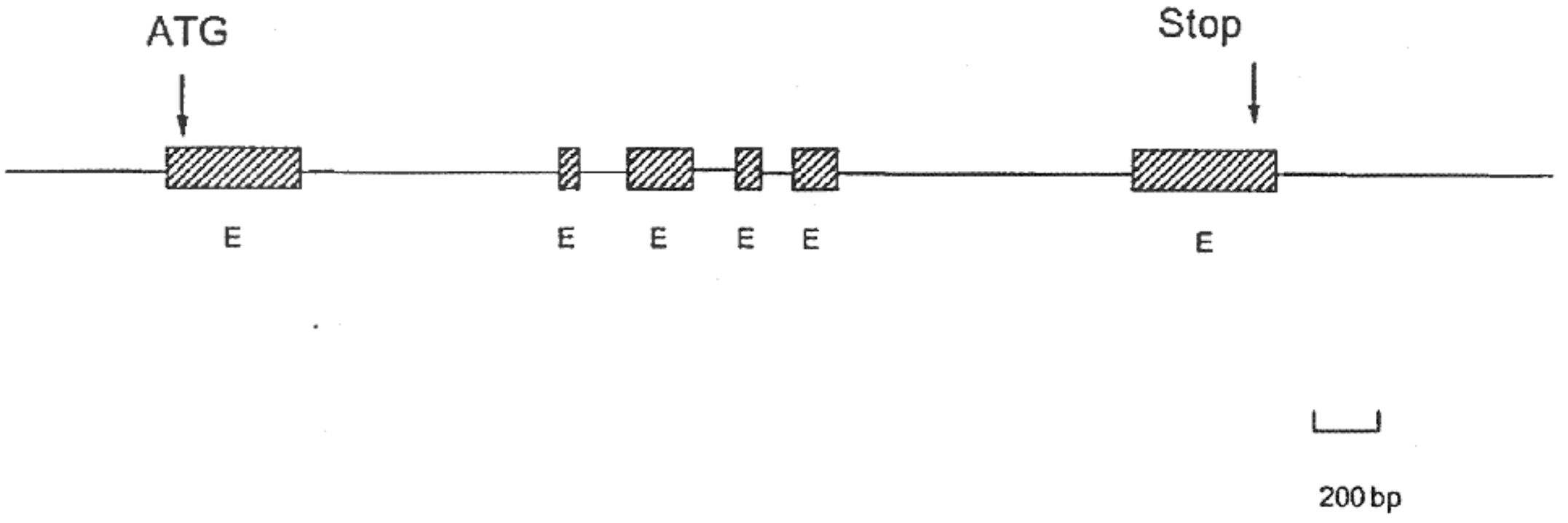

Figure 1. Schematic organization of the millet *opaque-2*-like gene. The hatched boxes represent the exons and their respective numbers (E1-6), the lines correspond to the 5' UTR, five introns and the 3' terminal.

Gene organization: The gene is organized into six exons interrupted by five introns (Figure 1) and encodes a protein of 426 amino acids (Figure 2). In Figure 2 the alignment of the amino acid sequence of the pearl millet O2 protein with that from six grass species is shown. In analogy to the situation found in maize, coix and sorghum, it is possible to identify functional domains that roughly correspond to the exons.

Exon 1 contains an "Acidic Domain" between D^{37} and T^{89}, the theoretical pI-value of which is 3.38. This conserved acidic region could be, in analogy to the maize situation [4], responsible for the protein's overall biological activity. Considering the degree of conservation within the 50 amino acids of this exon, one can infer that the 10 core amino acids are functionally the most important. This exon also contains a "Nuclear Localization Signal" between A^{105} and L^{127} with an extremely high degree of identity in all the genes analyzed. *DNAMAN* analysis of exon 2 did not highlight any identities or regions with potential biological functions. Exon 3 contains the start of another "Nuclear Localization Signal" at M^{209}, extending and covering almost the entire exon 4 until M^{240}. Obviously, exon 4 also represents the proteins "Basic Domain". Interestingly, in exon 3 of all species analyzed there is a highly conserved region (only two substitutions in rice and maize in the wobble position) coding for four serine residues.

Exon 4 contains the "Basic Domain" that enables the interaction between the target and the regulator and the codon for the first leucine of the "ZIP" region that enables the homo- and /or hetero-dimerization of the protein in order to be functional. The remaining "Leucine Zipper Domain" resides in exon 5. Exon 5 also contains a conserved region coding for the last 17 amino acids inclusively the codons for the two final leucines and a highly conserved DNRVL motif, which, so far, has not yet been mentioned for *opaque-2*-like proteins. For exon 6 so far no biological activity has been reported and it has almost no similar regions with the exons of the six other species.

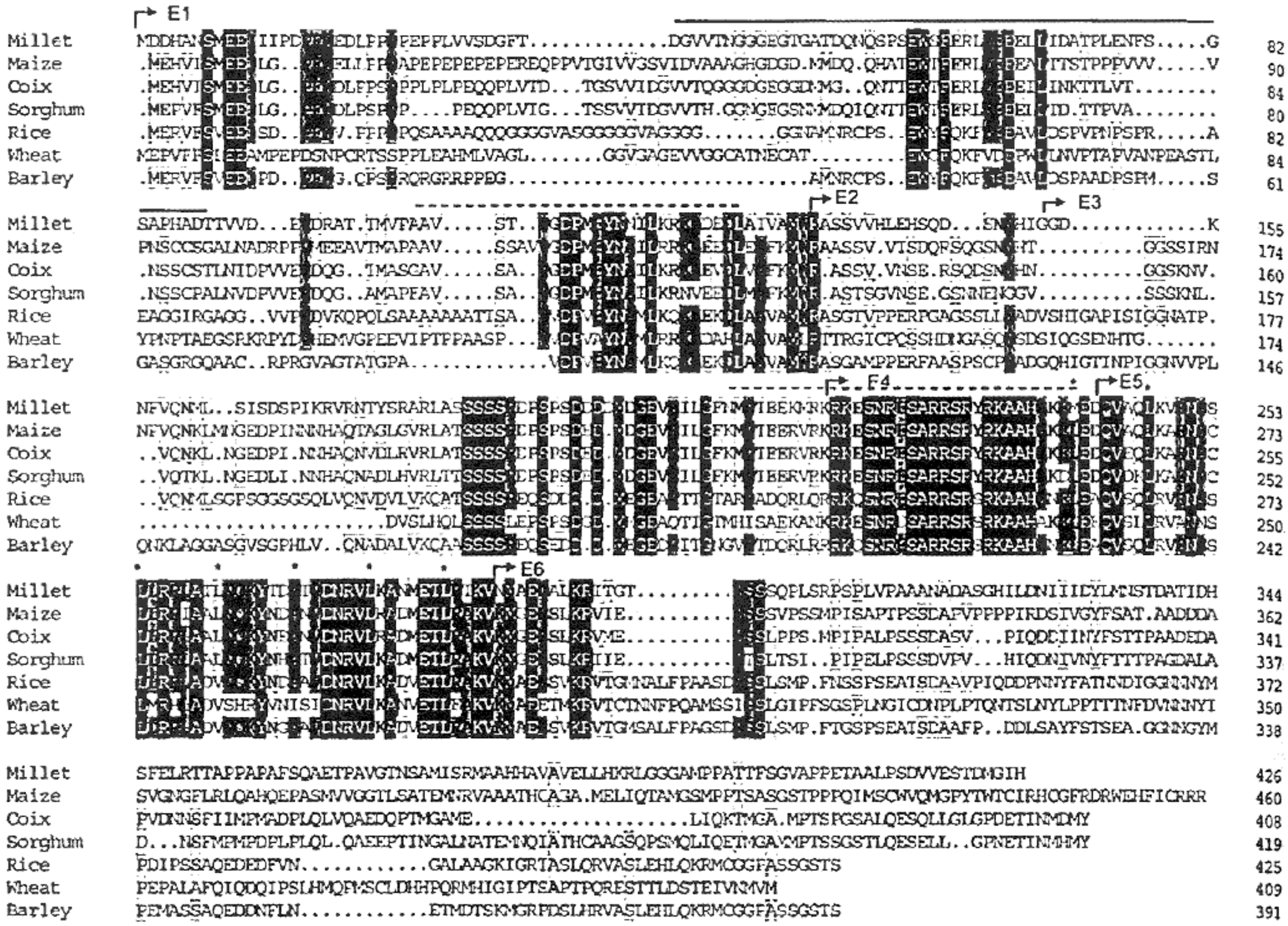

Figure 2. Alignment of opaque related proteins from seven grass species. E1 to 6: exons 1 to 6. Decreasing degrees of identities are displayed by decreasing shades from black to white. The solid line indicates the active domain described for O2. The dashed lines indicate the 1st and 2nd nuclear localization signal (NLS). Note that the 2nd NLS corresponds also to the basic region. (*) indicate the Leu residues of the leucine-zipper domain.

Protein Structure and Modeling: The basic and the Leu-ZIP region were submitted to structural analyses. In particular, we focalized on the region between glutamic acid (E^{214}) at the beginning of the basic region and lysine (K^{276}) at the end of the leucine-zipper. The single chain model is shown in figure 3A and in figure 3B a part of the dimerized Leu-ZIP region is presented.

The b-ZIP region is characterized by four to seven leucine residues spaced by six amino acids. In the case of pearl millet, the N-terminal ZIP-amino acid is actually a methionine, which, most probably, arose due to a substitution of the triplet's first nucleotide T with an A, thus changing TTG (L) to ATG (M). This substitution does not impair the dimerization process since this amino acid, for steric and electrostatic reasons, participates in the contacts necessary for chain interaction as shown in figure 3B.

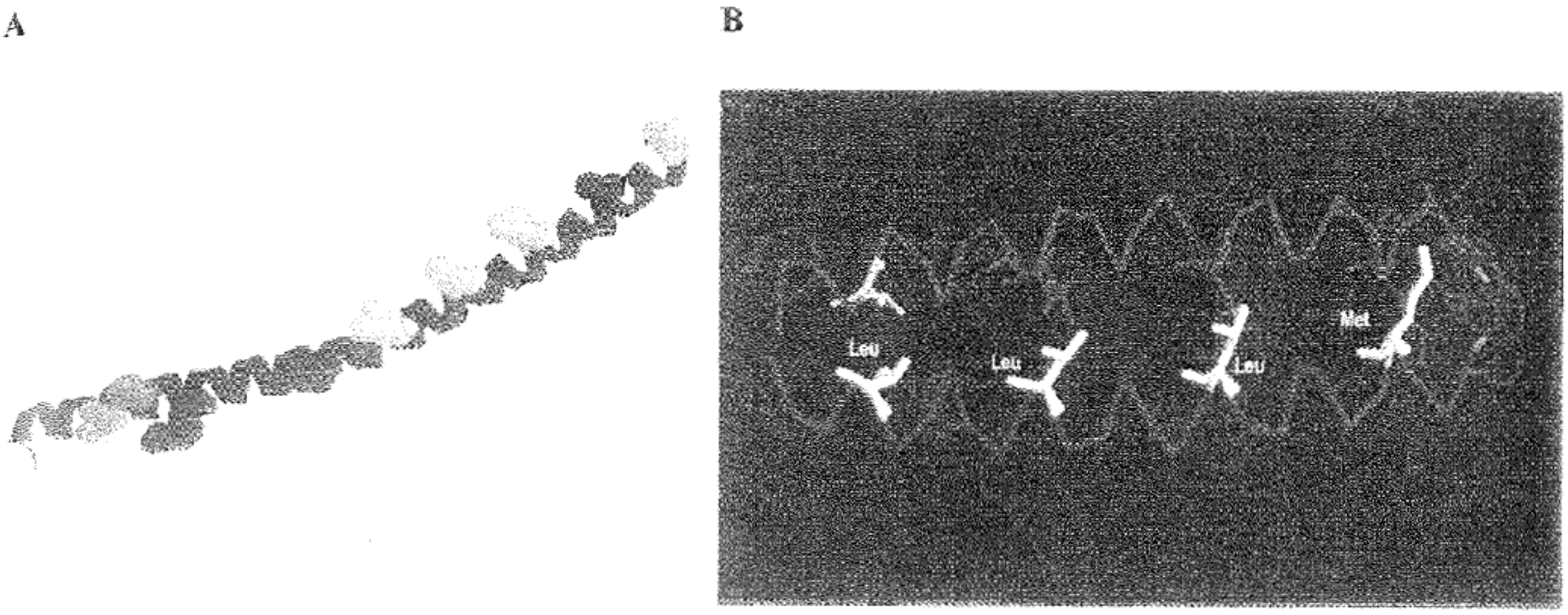

Figure 3. 3-D structure of the b-ZIP region of the pearl millet opaque-2-like protein. A) b-ZIP monomer. The methionine, leucines and alanine involved in dimerization are highlighted in red, yellow and purple. In the basic region the conserved amino acids asparagine, serine and arginine are in green, ocre and pink, respectively. B) Close up of the dimerized N-terminal of the zip region. The interactions of the two methionines and six leucines are shown. The residue's Van der Waals forces are also shown.

In addition, the sixth leucine is substituted by an alanine. Again it is reasonable to assume that this substitution does not affect the physical-chemical properties of the molecule since this substitute is also a hydrophobic amino acid and thus the biological character of the surface involved in dimerization is maintained. Of special interest is the fact that the amino acids between leucine ^{5} and leucine ^{7} are highly conserved, in particular with respect to the DNRVL motif in all the opaque-2 proteins investigated here. Since this region delineates the end of the zipper region, this conservation might be essential for the three-dimensional arrangement during dimerization. The pearl millet opaque-2 zipper region spans 48 amino acids and it is well known, from studies on other bZIP proteins [13], that the longer this region, the higher the stability of the dimers.

Expression Pattern: The expression pattern of bZIP observed on Western blots of proteins extracted from the three stages of seeds is shown in Figure **4**. An immunological reaction is observed as a double band in the 50 kDa region in young seeds (arrow, lane 1), while in seeds close to maturity as well as in leaves almost no reaction occurs (lanes 2 - 4). The molecular mass of the double band of about 50 kDa corresponds well with the 46.3 kDa calculated on the basis of the predicted amino acid composition for O2. The fact that there is a double band in the expected molecular weight region is most probably due to the presence of phosphorylated and non-phosphorylated molecules of the opaque-2 like protein, a situation also described for opaque-2 of maize by Ciceri *et al.* [14].

Besides these reactions, various other positive bands appear in the lower molecular weight region which are likely to be the result of the anti-bZIP antibody reacting with other regulatory proteins with conserved basic and leucine zipper regions.

RT-PCR: The expression pattern of the millet *O2*-like gene described above was confirmed by RT-PCR using total RNA from seeds at different developmental stages (Figure **5**). Stage 1 represents the

youngest seeds; stage 2. seeds at an intermediate stage of development and stage 3, seeds close to maturity. Stage 1 seeds are small. white, soft and moist, stage 2 seeds are still small and white but harder to the touch and stage 3 seeds are mature in size, of grayish color and hard to the touch. It can be seen in figure 5 that the expression pattern follows the scheme: 1>2>3: there is a progressive decrease in the expression of the millet *O2*-like gene in ail stages.

Taken together, these results suggest that in pearl millet the expression pattern of the opaque-2-like protein is quite similar to that described for opaque-2 of maize and also to an orthologous protein (BLZ1) of barley where the expression occurs early in the endosperm development [15,16].

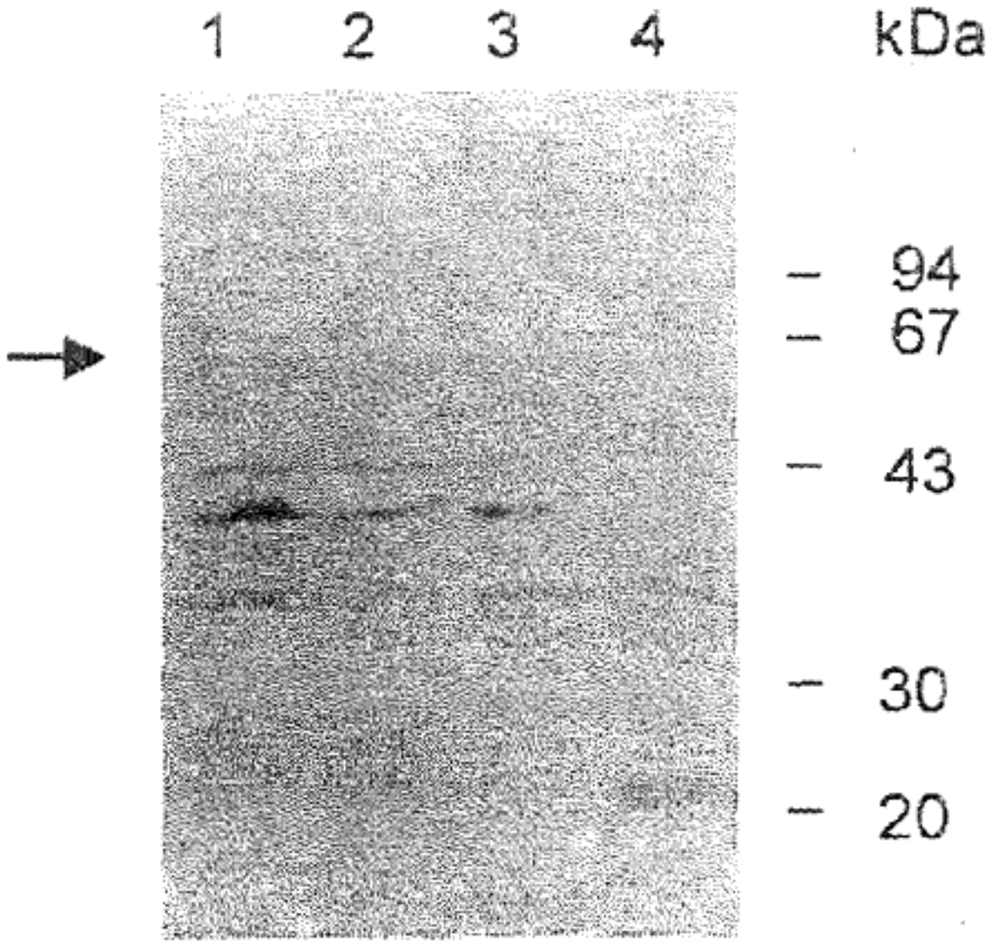

Figure 4. Western blot of protein extracted from three stages of seed development Lane 1: stage 1; lane 2: stage 2: lane 3: stage 3; lane 4: leaves. The arrow indicates the O2-like protein

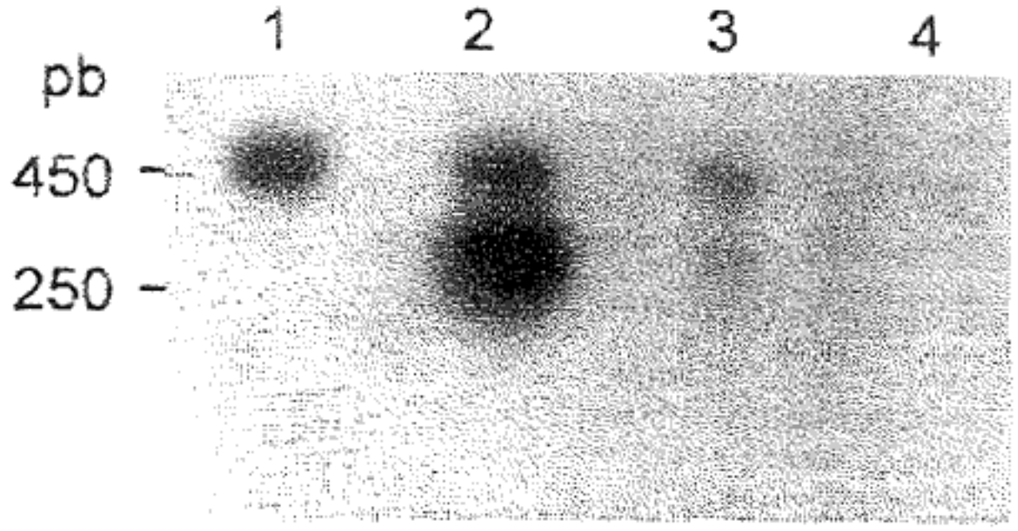

Figure 5. Pearl millet opaque-2 expression in seeds of three different developmental stages. mRNA from the different stages were analyzed by RT-PCR. Lane 1: stage 2 w/o reverse transcriptase as a control; Lane 2: stage 1; Lane 3: stage 2 and Lane 4: stage 3.

Gene Tree and Phylogenetics: We used the *DNA*MAN software to compare the amino acid sequences of pearl millet, coix, sorghum, rice. maize. barley and wheat opaque-like proteins and established a "Gene Tree" (figure 6A) based on the identities of the coding regions [17]. The tree shows a clear-cut divergence corresponding to the subfamilies *Panicoideae* on the one hand and *Poideae/Erhartoideae* on the other. Interestingly, the "molecular" tree diverges from the classification system established in the "Catalogue of New World Grasses" (figure **6B**) in the sense that the rice bZIP and barley blz-1 proteins seem to be closer related to each other and farther apart from wheat transcriptional activator SPA. The comparison of the individual exons of the seven species on the amino acid level (Figure **2**) reveals a gradient from pearl millet over maize. coix. sorghum. rice. barley down to wheat with respect to exon identities. The most conserved exons are exons 4 >5> 3; identities in exons 1, 2 and 6 are less prominent.

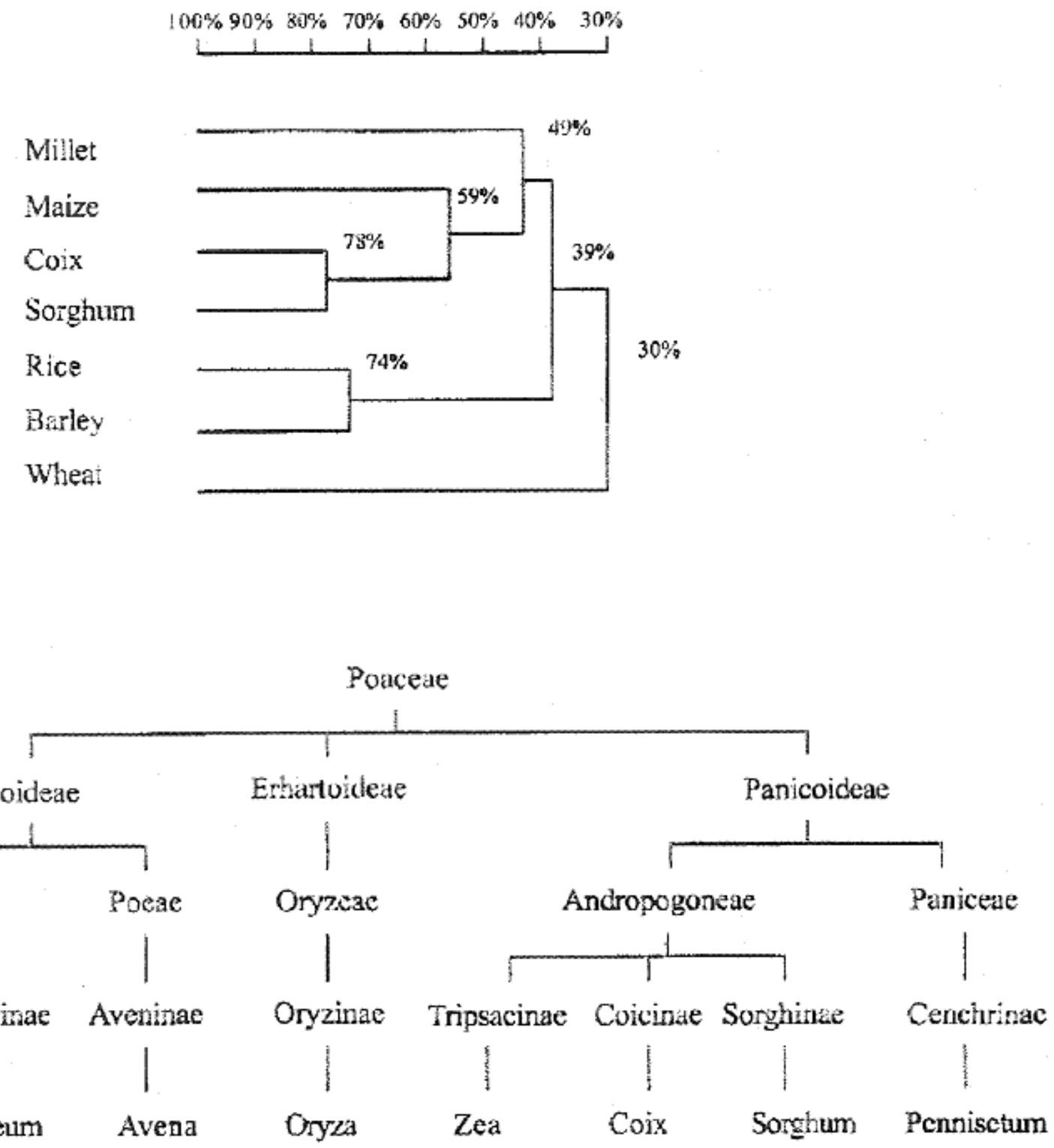

Figure 6. (A) Homology tree of O2-like proteins of seven grass species. The following protein sequences from the NCBI «Entrez protein» database were used: S42493, opaque 2 protein from *Coix lacrima-jobi*; S06022, regulatory protein-O2 from maize [33]; S56073, opaque-2 protein from sorghum; T04477, blz-1 protein from barley; T06767, probable transcription factor SPA from wheat [34]; BAA36492, bZIP protein from rice [35] and AAM53650, opaque-2-like protein from pearl millet (this paper).
B) Systematic classification of the *Poaceae* family according to the "Catalogue of New World Grasses" (2001).

For almost 30 years the consensus has been that transcriptional regulators perform a crucial role in evolution and that morphological changes are primarily the result of changes in regulatory genes [18,19]. The latter authors suggested that most evolutionary changes originate from modifications in regulatory genes rather than from modifications of structural genes themselves.

In this context, the opaque-2-like millet gene described here, the first identified in the important grass tribe Paniceae, could, in comparison with orthologous genes from other species, contribute to the phylogenetics analyses of these subfamilies, which so far is still largely based on morphological criteria. In particular in the light of the recognized importance of mutations in regulatory genes for morphological evolution, it is hoped that the comparison of individual opaque-2 exons of the seven grass species, could shed some light on the phylogenetic interrelationship between these species.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Brazilian National Research Council (CNPq).

REFERENCES

[1] Schmidt, R.J.; Burr, F.A.; Aukerman, M.J. and Burr, B. (**1990**) *Proc. Natl. Acad. Sci. USA, 87*, 46-50.

[2] Vettore, A.L.; Yunes, J.A.; Cord Neto, G. da Silva, M.J.; Arruda, P. and Leite, A. (**1998**) *Plant Mol Biol., 36*, 249-63.

[3] Izawa, T.; Foster, R. and Chua, N.-H. (**1993**) *J. Mol. Biol., 230*.1131-1144.

[4] Schmitz, D.; Lohmer S.; Salamini, F.; Thompson, R.D. (**1997**) *Nucleic Acids Res., 25*, 756-763.

[5] Lohmer, S.; Maddaloni, M.; Motto, M.; Di Fonzo, N.; Hartings, H.; Salamini, F. and Thompson, R.D. (**1991**) *EMBO J., 10*, 617-624.

[6] Pirovano, L.; Lanzini, Hartings, S. H.; Lazzaroni, N.; Rossi, V.; Rama, J.; Thompson, R.D.; Salamini, F. and Motto, M. (**1994**) *J. Mol. Biol., 24*, 515-523.

[7] Chandna, M. and Matta, N.K. (**1990**) *Phytochem., 29*, 3395-3399.

[8] Marcellino, L.H.; Bloch, Jr. C. and Gander, E.S. (**2002**) *Protein and Peptide Letters, 9*, 237-244.

[9] Sali, A. and Blundell, T.L. (**1993**) *J. Mol. Biol., 234*, 779-815.

[10] Neshich, G.; Togawa, R.; Mancini, A.L.; Kuser, P.R.; Yamagishi, M.E.B.; Pappas, Jr. G.; Torres, W.V.; Fonseca e Campos, T.; Ferreira, L.L.; Luna, F.M.; Oliveira, A.G.; Miura, R.T.; Inoue, M.K.; Horita, L.G.; de Souza, D.F.; Dominiquini, F.; Álvaro, A.; Lima, C.L.; Ogawa, F.O.; Gomes, G.B.; Palandrani, J.C.F.; dos Santos, G.F.; de Freitas, E.M.; Mattiuz, A.R.; Costa, I.C.; de Almeida, C.L.; Souza, S.; Baudet, C. and Higa, R.H. (**2003**) *Nucleic Acids Research, 31*, 1-7.

[11] Bompard-Gilles, C.; Rousseau, P.; Rouge, P. and Payan, F. (**1996**) *Structure, 12*, 1441-1452.

[12] Jones, J.D.G.; Dunsmuir, P. and Bedbrook, J. (**1985**) *EMBO Journal, 4*, 2411-2418.

[13] Hurst, H. (**1994**) *Protein Profile, 1*, 138-142.

[14] Ciceri, P.; Gianazza, E.; Lazzari, B.; Lippoli, G.; Genga, A.; Hoschek, G.; Schmidt, R.J. and Viotti, A. (**1997**) *Plant Cell, 9*, 97-108.

[15] Gallusci, P.; Salamini, F.; Thompson, R.D. (**1994**) *Mol. Gen. Genet., 244*, 391-400.

[16] Vicente-Carbajosa, J.; Onate, L.; Lara, P.; Diaz, I. and Carbonero, P. (**1998**) *Plant J., 13*, 629-640.

[17] Nei, M. (**1987**) Columbia University Press, Guildford, Surrey.

[18] Wilson, A.C. (**1975**) Stadler Symp. 7, Columbia, Mo. University of Missouri, 117-134.

[19] Doebley, J. and Lukens, L. (**1998**) *Plant Cell, 10*, 1075-1082.

Received on September 11, 2003, accepted on March 22, 2004.