1-1-1992

An Arabidopsis serine/threonine kinase homologue with an epidermal growth factor repeat selected in yeast for its specificity for a thylakoid membrane protein

Bruce D. Kohorn
Duke University

Steven Lane
Duke University

Tracy A. Smith
Duke University

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 10989–10992, November 1992
Cell Biology

An *Arabidopsis* serine/threonine kinase homologue with an epidermal growth factor repeat selected in yeast for its specificity for a thylakoid membrane protein

(light-harvesting chlorophyll-binding protein/GAL4 fusion protein)

BRUCE D. KOHORN*, STEVEN LANE, AND TRACY A. SMITH

Botany Department, Duke University, Durham, NC 27706

Communicated by Charles S. Levings III, July 31, 1992 (received for review April 23, 1992)

ABSTRACT **A number of molecules have recently been described that effect the correct transport and assembly of cytoplasmically synthesized proteins to cellular membranes. To identify proteins that bind or modify other proteins during the process of membrane translocation, we developed a yeast selection scheme that employs the yeast transcriptional activator GAL4. This selection facilitates the isolation of cDNAs that encode proteases and binding proteins for known target peptide sequences. We report the isolation of an *Arabidopsis* cDNA encoding a polypeptide that can interact with the amino terminus of a light-harvesting chlorophyll a/b-binding protein (LHCP), a cytoplasmically synthesized protein that is integral to the chloroplast thylakoid membrane. The cDNA was selected in yeast from an *Arabidopsis* expression library for its ability to inhibit a transcriptional activator GAL4–LHCP fusion protein, but not inhibit native GAL4 protein. The LHCP amino-terminal sequences included in the fusion protein are known to regulate LHCP biogenesis and function. The *Arabidopsis* cDNA encodes a 595-amino acid protein with at least two functional domains, one with similarity to the family of protein-serine/threonine kinases and another that contains an epidermal growth factor repeat. The identification of an EGF repeat in *Arabidopsis* indicates that the motif is conserved between the plant and animal kingdoms. Hybridization studies indicate that this gene is likely to be present in other genera of plants. Its mRNA is detected in green leaves but not in other plant tissues or in etiolated plants. The specificity in yeast and the expression pattern in plants together are suggestive of a role for this protein kinase in the assembly or regulation of LHCP.**

The translocation of membrane proteins from their site of synthesis to a target membrane often involves specific receptors, chaperonin molecules, and proteases (1). Few interactions have been characterized in detail for chloroplast thylakoid membrane proteins, such as the light-harvesting chlorophyll a/b-binding protein (LHCP), that are synthesized in the cytoplasm. While there are suggestions that LHCP binds a cytoplasmic factor (2) and a chloroplast envelope receptor (3), is processed by a stromal peptidase (4), and subsequently binds at least two stromal factors (5) that are required for a productive interaction with thylakoid membrane proteins (6), none of these molecules has been isolated.

To identify those proteins that modify or bind LHCP, we developed a selection scheme for the isolation of cDNAs that encode proteases and binding peptides for known target peptide sequences (7). This scheme was first tested using an isolated protease and a defined peptide target. Introduction of an 18-amino acid cleavage signal for tobacco etch virus (TEV) protease into the central portion of the GAL4 transcription factor did not alter the ability of the transcription factor to promote galactose-dependent growth of yeast. Cells that expressed the TEV protease in addition to the GAL4 fusion protein were unable to grow on galactose and could be selected in the presence of a suicide substrate, 2-deoxygalactose. In principle, targeted inactivation could also be brought about by binding factors that cause steric hindrance or modification of the GAL4 fusion protein.

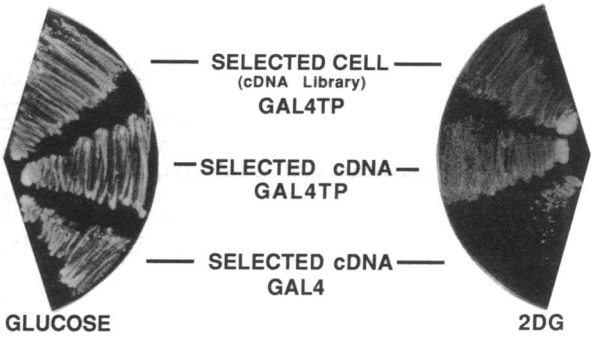

FIG. 1. Selection of LHCP-specific cDNA in yeast. The original 2-deoxygalactose (2DG)-selected cell is shown after growth on glucose (top left) or 2DG/glucose (top right). cDNA *pro25* was isolated and transformed into YJ0/pSB32GAL4TP (middle left and right) or YJ0/pSB32GAL4 (bottom left and right).

We have used this scheme to isolate a cDNA† encoding a protein with specificity for the amino terminus of the thylakoid membrane protein LHCP. The LHCP region chosen for the GAL4 fusion protein includes the 36-residue chloroplast transit peptide that is necessary for targeting to the chloroplast (8) and should therefore interact with both cytoplasmic and chloroplast proteins. We also included the amino-terminal 18 amino acids of the mature LHCP, as this region (*i*) when included with the transit peptide, increases the efficiency of chloroplast import (8) and (*ii*) contains a threonine whose phosphorylation is correlated with a reduction in the channeling of light energy into photosystem II (9). Moreover, the stromal processing peptidase (4), to date not isolated, cleaves between these 36- and 18-amino acid regions. Thus the selection scheme using the GAL4–transit peptide (GAL4TP) fusion could potentially reveal proteases, receptors, chaperonins, or kinases with specificity for functionally defined LHCP sequences.

MATERIALS AND METHODS

cDNA Selection. *Saccharomyces cerevisiae* strain YJ0/pSB32GALTP (*LEU2*) was transformed (7) with an *Arabidopsis* total plant cDNA expression library in pYES (*URA3*;

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: LHCP, light-harvesting chlorophyll a/b-binding protein; EGF, epidermal growth factor.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. L04999).

ref. 10). Over 50,000 Ura3$^+$ Leu2$^+$ transformants were isolated, subjected to a second selection with 2-deoxygalactose, and then picked to a gridded, glucose master plate. Cells that had suffered a mutation in the *GAL4TP* fusion gene were distinguished from those having functional *GAL4TP* and a selected cDNA in the following manner: the master plate was replica plated two successive times to media containing leucine and 5-fluoroorotic acid, the latter of which selects for the loss of the *URA3* plasmid carrying the cDNA (11). We confirmed that these cells were indeed *ura*$^-$ and then tested for their ability to grow on 2-deoxygalactose. One of the 200 *ura*$^-$ colonies was unable to grow on 2-deoxygalactose, and the remaining 199 were therefore discarded as putative *gal4* mutants. The cDNA plasmid was isolated from this one strain

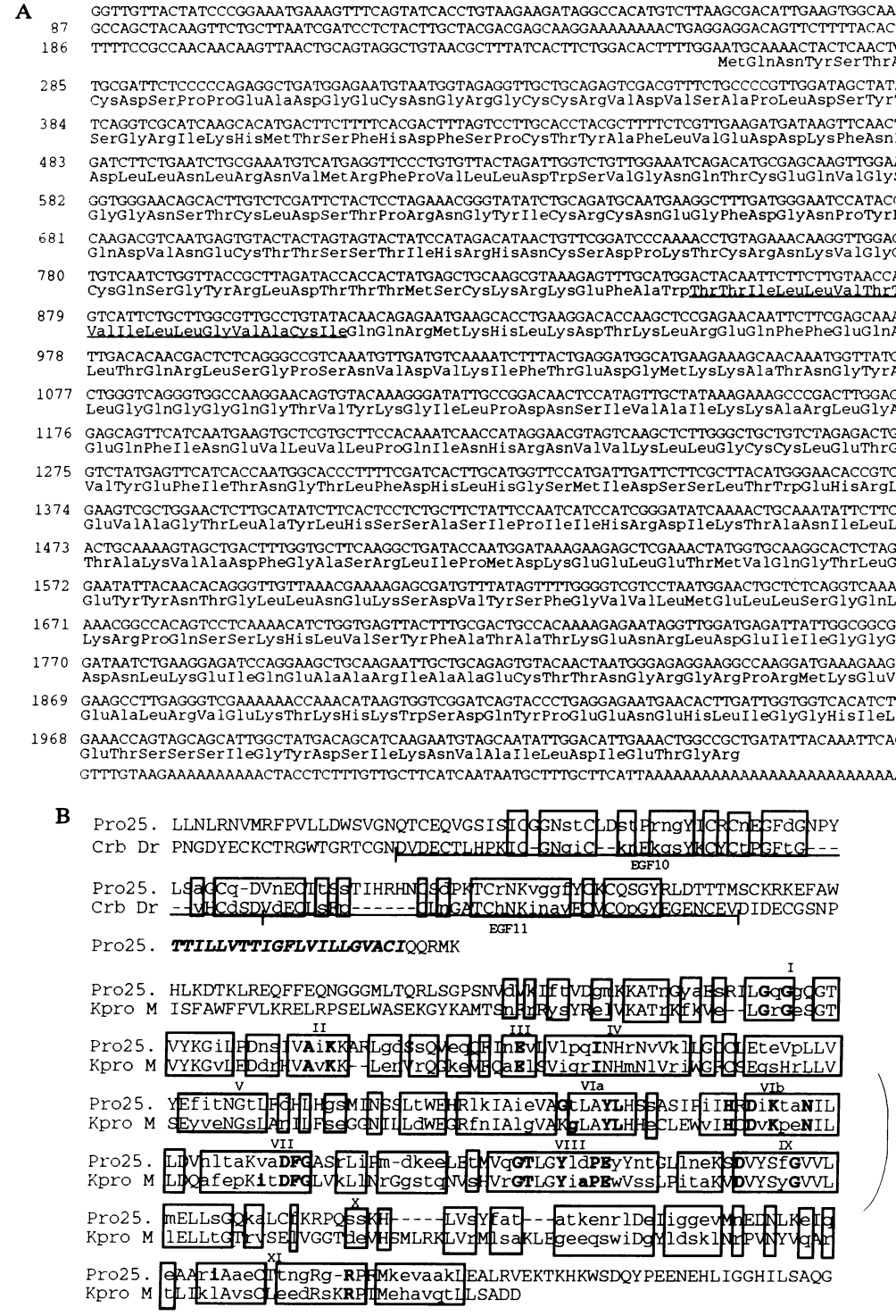

FIG. 2. (*A*) DNA and predicted amino acid sequence for cDNA *pro25*. Underlining indicates the predicted transmembrane region. (*B*) PRO25 contains two epidermal growth factor (EGF) repeats and homology with protein kinases. Amino acid residues 81–572 of PRO25 are shown on the top line. Residues 628–733 of the Crumbs protein EGF repeats 10 and 11 (17) (Crb Dr) are shown below PRO25. Residues 490–817 of a putative receptor serine/threonine kinase from maize (Kpro M; ref. 18) are shown below PRO25, carboxyl-terminal to the predicted membrane-spanning region (boldface italics). Boxes indicate conserved regions, wherein uppercase letters show identity and lowercase letters show similarity. Roman numerals indicate the 11 conserved regions containing the residues (boldface amino acid symbols) characteristic of protein kinases (19).

by total yeast cell DNA preparation and amplification in *Escherichia coli*. cDNA pro25 and pSB32GAL4TP were transformed back into the strain that had been selected on 2-deoxygalactose and subsequently cured of all plasmids.

Additional Methods. RNA was isolated as described (12) and electrophoresed in denaturing agarose gels (13). DNA was isolated as described (12), and nucleic acid hybridization conditions were according to Church and Gilbert (14). DNA sequence was determined from single-stranded templates by using Sequenase according to the manufacturer's instructions (United States Biochemical). Analysis of proteins by denaturing gel electrophoresis was as described (15).

RESULTS AND DISCUSSION

The coding region for the 36-residue LHCP transit peptide and the first 18 amino acids of the mature LHCP were fused in frame between codons 236 and 237 of *GAL4* (7), creating *GAL4TP*. *GAL4TP* was cloned into the leucine complementing yeast/*E. coli* shuttle vector pSB32 and introduced into a *gal4* deletion mutant of *S. cerevisiae*, YJ0 (*ura3-52, leu2-3,112 gal4Δgal80Δ*; refs. 7 and 16). Yeast that express GAL4TP grow on glucose and galactose, but not in the presence of the suicide substrate 2-deoxygalactose, indicating that the GAL4TP fusion protein functions as a transcription factor (7). We introduced an *Arabidopsis* cDNA library (in pYES) into YJ0/*GAL4TP* yeast and selected for growth in the presence of 2-deoxygalactose. Of 200 colonies that were obtained, only one yeast isolate required a cDNA for growth on 2-deoxygalactose (Fig. 1); the remaining 199 had most likely suffered null mutations in *GAL4TP*. This cDNA plasmid (termed pRO25 and having cDNA insert *pro25*) was isolated and retransformed into the selected YJ0 containing only *GAL4TP*. Fig. 1 shows that the addition of pRO25 permits YJ0/GAL4TP to grow on 2-deoxygalactose/glucose as well as glucose. The effect of pRO25 is specific to the LHCP sequences, as the presence of pRO25 permits neither YJ0/*GAL4* (Fig. 1) nor YJ0 (data not shown) to grow on 2-deoxygalactose/glucose. A randomly picked cDNA that was transformed into YJ0/*GAL4TP* will not permit growth on 2-deoxygalactose/glucose (data not shown). Thus, the effect of pRO25 is specific to the GAL4TP fusion.

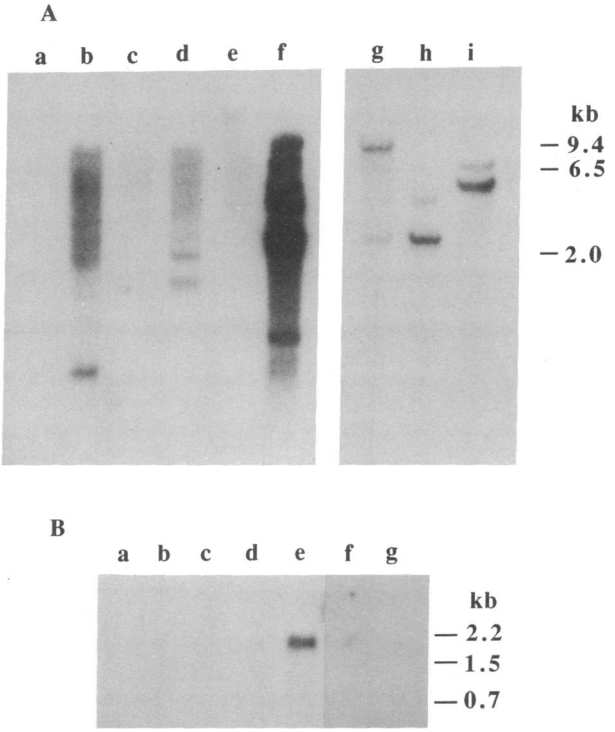

FIG. 3. *pro25* is found in other plant genera and is expressed in light-grown leaves. (*A*) The entire *pro25* cDNA sequence shown in Fig. 2*A* was labeled with ^{32}P and hybridized at high stringency (65°C) to *Eco*RI-, *Hin*dIII-, or *Bam*HI-digested *Arabidopsis thaliana* var. Columbia total DNA (lanes *g–i*, respectively) or at low stringency (45°C) to *Hin*dIII-digested DNA of *Chlamydomonas reinhardtii* (lane a), spinach (Krogers grocery) (lane b), tobacco (SR1) (lane c), pea ('Alaska'), (lane d), maize ('Black Mexican Sweet') (lane e), or *A. thaliana* var. Columbia (lane *f*). (*B*) Total RNA isolated from *A. thaliana* var. Columbia roots (lane a), stems (lane b), flowers (lane c), siliques (lane d), leaves (lane e), 1-week dark-adapted seedlings (lane f), and dark-grown seedlings (lane g) was run in a denaturing agarose gel, blotted, and hybridized to labeled *pro25* cDNA. *A* and *B* are autoradiographs. kb, Kilobases.

The DNA sequence of *pro25* contains a single, large open reading frame capable of encoding a 595-amino acid polypeptide (Fig. 2*A*). A hydrophobic region suggestive of a transmembrane domain is found at amino acids 198–218 (underline, Fig. 2). A search of the Swiss-Prot data base (January 1992) reveals several interesting homologies (Fig. 2*B*). Amino acids 224–540 have homology with the family of protein kinases, and a putative maize receptor kinase (Kpro M) shows the highest (33%) identity (18). Each of the 11 catalytic sites characteristic of protein kinases are present (roman numerals above boxes and bold amino acid symbols, Fig. 2*B*), and identities in kinase subdomains VI and VIII suggest that *pro25* encodes a serine/threonine kinase (19). It is intriguing that LHCP has an amino-terminal threonine residue whose phosphorylation is correlated with LHCP function (9), but further experiments will be necessary to determine whether this threonine, which is present in the GAL4TP fusion, is the target of the putative kinase. Amino acids 110–183 of *pro25*, amino-terminal to the transmembrane domain (in bold italics, Fig. 2*B*), have 35% identity to the EGF repeats 10 and 11 of the Crumbs protein of *Caenorhabditis elegans* (ref. 17; EGF10 and EGF11, Fig. 2*B*). This identification of an EGF repeat in *Arabidopsis* indicates that the motif has been conserved between the plant and animal kingdoms.

High-stringency hybridization of labeled cDNA to genomic DNA indicates that *pro25* is a single- or low-copy gene in *Arabidopsis* (Fig. 3*A*, lanes *g–i*). Low-stringency hybridization shows that *pro25* is a member of a family of related genes (Fig. 3*A*, lane f), and has a low degree of homology to DNA of other plant genera, notably pea and spinach (Fig. 3*A*, lanes a–e).

We were curious to note that *pro25*, which had been selected in yeast for its specificity for the amino terminus of LHCP, had homology to a membrane-anchored protein kinase. If the polypeptide encoded by *pro25* were associated with LHCP in plants, then we would expect to detect *pro25* mRNA primarily in tissues that also express LHCP and contain chloroplasts. Indeed, like LHCP mRNA, *pro25* mRNA is detected in abundance only in light-grown leaf tissue (Fig. 3*B*, lane e) but not in other plant organs or etiolated seedlings (lanes a–d and g, respectively). *pro25* mRNA is present at low levels in dark-adapted leaves (lane f). The mRNA is estimated to be ≈2 kb, and since the cDNA sequence is 2.165 kb, we conclude that the clone is very likely a full-length representation of the mRNA.

If indeed the *pro25* cDNA encodes a protein (termed PRO25) that interacts with LHCP in plants, then we would expect to find it associated with the chloroplast. While there are no strict consensus sequences for chloroplast targeting signals (8), we cannot identify similarity of PRO25 to any known amino-terminal signal sequences, nor is its amino terminus predicted to be hydrophobic or basic. We have been able to express PRO25 as the expected 70-kDa protein by using SP6 RNA polymerase-derived RNA and wheat germ translation extract. These translation products were incu-

bated with either pea or *Arabidopsis* isolated chloroplasts and small amounts of proteins became resistant to a protease that had been added after the incubation. These data (not shown) suggest an association of PRO25 with the chloroplast, but the levels of this *in vitro* import were insufficient to prove a chloroplast location of PRO25. Support for a chloroplast association of PRO25, therefore, will require an accurate localization in plants by use of an antibody specific to PRO25.

PRO25 represents an unusual plant protein kinase that was selected in yeast for its specificity for a thylakoid membrane protein. It is not clear how PRO25 modifies LHCP in yeast, and further work is required to determine whether the kinase or the amino-terminal domain interact with LHCP. The expression in plants and specificity in yeast are suggestive of some intriguing mechanism of LHCP biogenesis but certainly do not prove a significant role. Description of the function of this protein kinase awaits immunodetection and functional analysis in whole plants.

We thank Audrey Alexander for expert help and Andrea Auchincloss, John Boynton, Nick Gillham, Jim Siedow, Xinnian Dong, and Tai-Ping Sun for helpful discussions. This work was supported by the Pew Charitable Trusts and National Institutes of Health Grant GM39696-01 to B.D.K.

1. Wickner, W. (1988) *Biochemistry* **27**, 1081–1086.
2. Waegemann, K., Paulsen, H. & Soll, J. (1990) *FEBS Lett.* **261**, 89–92.
3. Schnell, D. J., Blobel, G. & Pain, D. (1991) *J. Biol. Chem.* **266**, 3335–3342.
4. Clark, S. E., Oblong, J. E. & Lamppa, G. K. (1990) *Plant Cell* **2**, 173–184.
5. Payan, L. A. & Cline, K. (1991) *J. Biol. Chem.* **112**, 603–613.
6. Fulsom, D. R. & Cline, K. (1988) *Plant Physiol.* **88**, 1146–1153.
7. Smith, T. A. & Kohorn, B. D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5159–5162.
8. von Heijne, G. & Nishikawa, K. (1990) *FEBS Lett.* **278**, 1–3.
9. Allen, J. (1992) *Biochem. Biophys. Acta* **1098**, 275–335.
10. Elledge, S., Mulligan, J., Ramer, S., Spotteswood, M. & Davis, R. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1731–1735.
11. Boke, J. D. (1987) *Methods Enzymol.* **154**, 154–175.
12. Luetwiler, L., Meyerowitz, E. M. & Tobin, E. M. (1986) *Nucleic Acids Res.* **14**, 4051–4064.
13. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
14. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
15. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
16. Salmeron, J. M., Langdon, S. D. & Johnston, S. A. (1989) *Mol. Cell. Biol.* **9**, 2950–2956.
17. Tepass, U., Theres, C. & Knust, E. (1990) *Cell* **61**, 787–799.
18. Walker, J. C. & Zhang, R. (1990) *Nature (London)* **345**, 743–746.
19. Hanks, S. K. & Quinn, A. M. (1991) *Methods Enzymol.* **200**, 38–62.