Volume 260, number 1, 127-130 FEB 08027 January 1990

Structure of the gene encoding concanavalin A from *Canavalia gladiata* and its expression in *Escherichia coli* cells

Daisuke Yamauchi and Takao Minamikawa

Department of Biology, Tokyo Metropolitan University, Fukazawa, Setagaya-ku, Tokyo, 158 Japan

Received 31 October 1989

We have cloned and sequenced the gene encoding concanavalin A (Con A) from *Canavalia gladiata*. The sequence covers the whole transcribed region as well as the 5'- and 3'-untranscribed sequences. The coding sequence lacks introns. Con A expressed in *Escherichia coli* cells was purified by Sephadex G-50 affinity chromatography. The precursor of Con A expressed in *E. coli* undergoes a peptide cleavage and ligation in the same way as that synthesized during seed maturation.

Concanavalin A; Gene structure; Legume lectin; Carbohydrate binding activity; Processing; Peptide ligation; (*Canavalia gladiata, Escherichia coli*)

1. INTRODUCTION

Lectins, carbohydrate binding proteins, are commonly found in seeds of Leguminosae and accumulate during seed maturation. The structures of lectin genes have been determined in soybean [1], French bean [2] and pea [3]. Con A, a seed lectin of the *Canavalia* species, is synthesized by a unique biosynthetic mechanism in maturing seeds. The precursor to Con A which is glycosylated in the endoplasmic reticulum receives post-translational processing, which includes the ligation of cleaved polypeptides [4–8].

In our previous work, we isolated a cDNA clone for Con A, pCONA1, from *C. gladiata* [9] and determined the nucleotide sequence [10]. In the present paper, we describe the isolation and sequencing of the Con A gene from *C. gladiata* and its expression in *Escherichia coli* cells harboring pCONA1.

2. MATERIALS AND METHODS

2.1. *Construction of a genomic library and screening for Con A clones*

Nuclear DNA from *C. gladiata* was isolated as described previously [10]. The DNA was partially digested with *Mbo*I and size-fractionated through a 10–40% sucrose gradient. DNA fragments, 10–20 kb in size, were ligated to the λDASH-*Bam*HI digested arm (stratagene) and packaged in vitro. Approximately 2×10^6 independent recombinant phages obtained were screened with ^{32}P-labeled cDNA for Con A, pCONA1 [9]. Subclones were constructed in pUC118 [11] and mp18/19 [12] for sequencing.

2.2. *S1 nuclease mapping and primer extension*

The S1 nuclease mapping and the primer extension were performed as described by Sakurai et al. [13]. The total RNA fraction of immature embryos collected at 40 days after flowering was prepared as described elsewhere [14].

2.3. *Purification of Con A from* E. coli

From cells of the *E. coli* strain JM109 harboring pCONA1 grown in L-broth containing 50 μg/ml ampicillin and 1 mM isopropyl-β-D-thiogalactopyranoside at 37°C overnight, Con A was purified by affinity chromatography as described by Stubbs et al. [15] except that Sephadex G-50 and glucose were used instead of Sephadex G-75 and α-methyl-D-mannopyranoside.

2.4. *Analysis of the amino acid sequence*

The M_r 30000 polypeptide of Con A from *C. gladiata* seeds or *E. coli* cells was separated by SDS-PAGE [16] and used for the determination of amino acid sequences.

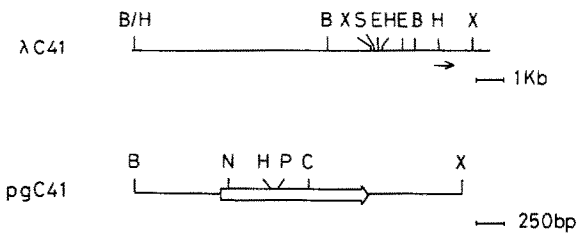

Fig.1. Restriction maps of the Con A gene (λC41) and genomic subclone (pgC41). A fine arrow and a thick open arrow denote the direction and region of transcription of the Con A gene, respectively. The restriction enzyme sites indicated are: B, *Bgl*II; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; N, *Nco*I; P, *Pvu*II; S, *Sal*I and X, *Xba*I.

Correspondence address: D. Yamauchi, Department of Biology, Tokyo Metropolitan University, Fukazawa, Setagaya-ku, Tokyo 158, Japan

Abbreviations: Con A, concanavalin A; PAGE, polyacrylamide gel electrophoresis

The nucleotide sequence data reported here will appear in the EMBL nucleotide sequence database under the accession number X16041.

Published by Elsevier Science Publishers B.V. (Biomedical Division)

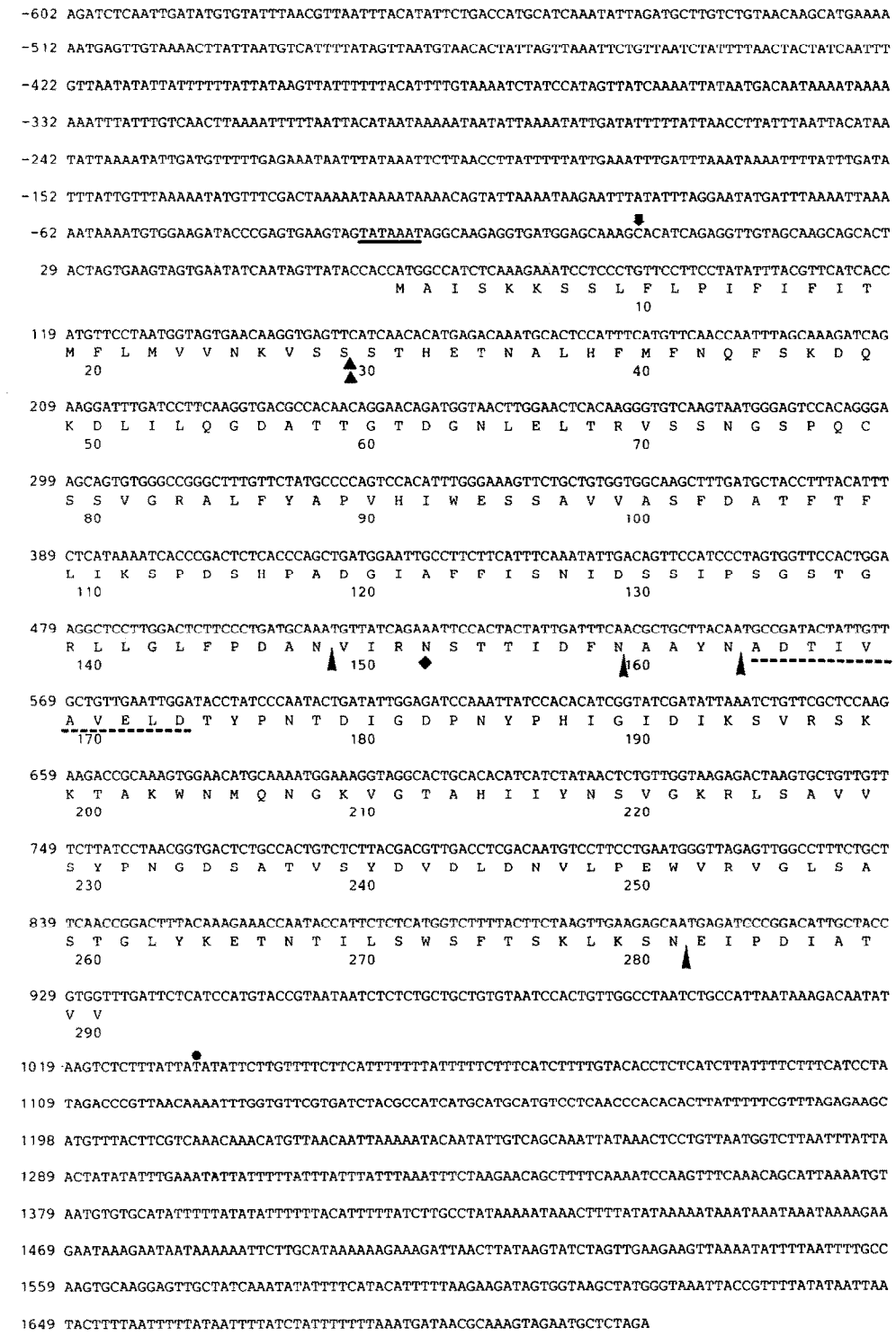

Fig.2. The nucleotide sequence of pgC41 and the deduced amino acid sequence. The nucleotide sequence is numbered in a 5′ to 3′ direction beginning with one of the major transcription initiation sites (+1), which is indicated by an arrow. Amino acid numbers are given below the amino acid sequences. The poly(A) addition site is indicated by a dot. The TATA box is underlined. The 10-amino-acid N-terminus of Con A from seeds, which is identical to that of Con A from *E. coli* cells, is indicated by a broken line. Double triangles and a diamond indicate the co-translational processing site and Asn residue attached to the oligosaccharide chain, respectively [1,4]. Arrowheads indicate the sites of post-translational processing [3,4].

3. RESULTS AND DISCUSSION

From a genomic phage library constructed with DNA from *C. gladiata*, we isolated three positive clones by plaque hybridization; these clones showed similar restriction maps. λC41 carrying a 13 kb fragment was chosen for detailed characterization (fig.1). The 2.3 kb

*Bgl*II-*Xba*I fragment of λC41 was subcloned into pUC118, designated pgC41 and further mapped with several enzymes (fig.1).

The sequence of the Con A coding and flanking regions is presented in fig.2. A comparison between sequences of pgC41 and pCONA1 [10] revealed complete agreement and showed that the Con A gene lacks introns as reported for other legume lectin genes [1–3]. The transcription initiation sites were determined with S1 nuclease mapping and primer extension analysis, and several bands were detected by both methods (fig.3). These bands may be ascribed to multiple transcription initiation sites with the majority located in nucleotides 62–64 upstream from the initiation codon. The 64th nucleotide upstream from the initiation codon is number +1. The transcription initiation sites of the soybean lectin gene [1] and phytohemagglutinin genes [2] were determined by S1 nuclease mapping and several protected fragment bands were also detected on the gels. These results suggest that there are multiple transcription initiation sites in legume seed lectin genes. The region upstream of the TATA box of the Con A gene shows low homology to other lectin genes, and no consensus sequences appear in these 5′-flanking regions of legume seed lectin genes including the Con A gene.

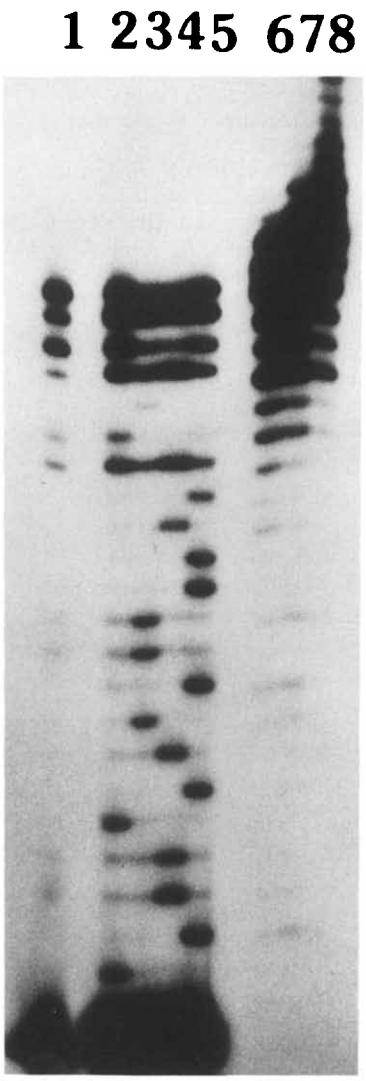

Fig.3. Determination of the transcription initiation sites of Con A mRNA. The primer extension analysis is shown in lanes 1–5. The 5′-end labeled primer (position +23 to +63) was hybridized with the total mRNA of immature embryos at 47°C for 16 h and extended by reverse transcriptase. Lane 1, primer extension; lanes 2, 3, 4, and 5 are G, A, T and C reactions, respectively. The S1 mapping analysis is shown in lanes 6–7. The 5′-end labeled probe (position −71 to +63) was hybridized with total RNA at 37°C and digested with various concentrations of S1 nuclease. Lanes 6, 7 and 8 reflect 500, 150 and 50 units/ml, respectively.

In a previous paper [9], we constructed a cDNA expression library from total poly(A)$^+$ RNA of immature *C. gladiata* embryos by using an *E. coli* expression vector pKEN602 [17], and selected a cDNA clone for Con A, pCONA1, from the library by the direct immunological screening of the colonies. When proteins extracted from *E. coli* cells harboring pCONA1 were analyzed by protein blotting with the anti-Con A antibody, it reacted specifically to a polypeptide with M_r 30 000 (data not shown). About 250 μg of purified Con A, which had carbohydrate binding activity was obtained from 250 ml of culture. This corresponded to about 0.15% of unbound proteins eluted from the DEAE cellulose column [15]. The Con A gave four polypeptide bands with an approximate M_r of 30 000, 21 000, 17 000, and 11 000 as determined by SDS-PAGE (fig.4). This

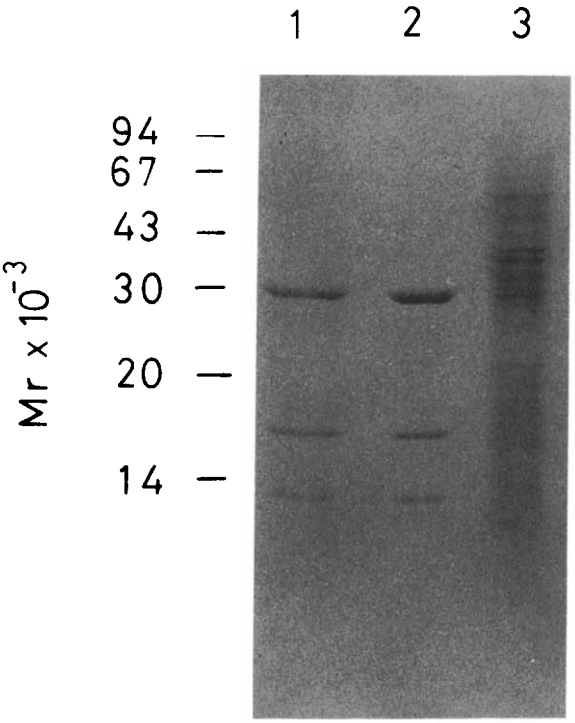

Fig.4. Comparison of SDS-PAGE patterns of Con A from seeds and *E. coli* cells. The gel was stained with Coomassie brilliant blue R. Lanes: (1) Con A from seeds, (2) Con A from *E. coli* cells, (3) unbound proteins from the DEAE cellulose column.

subunit composition is identical to that from seeds. There appeared to be no difference between the two Con A preparations, when compared by two-dimensional gel electrophoresis (data not shown). N-terminus amino acid sequences for both of the M_r 30000 polypeptides corresponded to amino acids 164–173 of the amino acid sequence deduced from the nucleotide sequence of pCONA1 [14] or pgC41 (fig.2). We thus conclude that Con A expressed in *E. coli* cells is identical to mature Con A from seeds.

In developing seeds, the glycosylated precursor of Con A (amino acids 30–290) is processed to two smaller polypeptides (amino acids 30–148 and 160–281) which are ligated at amino acids 281 and 30 to produce mature Con A (fig.2) [4,6,7]. The present results indicate that the precursor form of Con A is cleaved and ligated to produce mature Con A in *E. coli* cells.

Bowles et al. [6] proposed that the precursor form of Con A produced in *C. ensiformis* probably has a conformation close to that of the mature molecule. It is likely that in *E. coli* cells the precursor form of Con A, which also has a conformation close to that of mature Con A, is cleaved by proteases and the cleaved peptides are ligated to produce mature Con A having carbohydrate binding activity. In the present studies, however, we were unable to detect the occurrence of the precursor form of Con A in *E. coli* cells. This bacterial system for Con A synthesis offers an approach to the study of the relationship between the primary structure of Con A and its carbohydrate binding activity.

Acknowledgements: This work was partly supported by a grant-in-aid (no. 63622002) from the Ministry of Education, Science and Culture of Japan. We thank Dr S. Izumi and Dr H. Sakurai of Tokyo Metropolitan University for their advice on the plaque hybridization and determination of the transcription initiation sites. We are grateful to Dr H. Hattori of the National Institute for Basic Biology for determination of the N-terminus amino acid sequences of Con A from seeds and *E. coli* cells.

REFERENCES

[1] Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983) Cell 34, 1023–1031.
[2] Hoffman, L.M. and Donaldson, D.D. (1985) EMBO J. 4, 883–889.
[3] Kaminski, P.A., Buffard, D. and Strosberg, A.D. (1987) Plant Mol. Biol. 9, 497–507.
[4] Carrington, D.M., Auffret, A. and Hanke, D.E. (1985) Nature 313, 64–67.
[5] Herman, E.M., Shannon, L.M. and Chrispeels, M.J. (1985) Planta 165, 23–29.
[6] Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R. and Burgess, J. (1986) J. Cell Biol. 102, 1284–1297.
[7] Chrispeels, M.J., Hartl, P.M., Strum, A. and Faye, L. (1986) J. Biol. Chem. 261, 10021–10024.
[8] Faye, L. and Chrispeels, M.J. (1987) Planta 170, 217–224.
[9] Yamauchi, D., Nakamura, K., Asahi, T. and Minamikawa, T. (1988) Eur. J. Biochem. 170, 515–520.
[10] Yamauchi, D., Nakamura, K., Asahi, T. and Minamikawa, T. (1989) Plant Cell Physiol. 30, 147–150.
[11] Vieira, J. and Messing, J. (1987) Methods Enzymol. 153, 3–11.
[12] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.
[13] Sakurai, H., Fujii, T., Izumi, S. and Tomino, S. (1988) J. Biol. Chem. 263, 7876–7880.
[14] Yamauchi, D. and Minamikawa, T. (1987) Plant Cell Physiol. 28, 421–430.
[15] Stubbs, M.E., Carver, J.P. and Dunn, R.J. (1986) J. Biol. Chem. 261, 6141–6144.
[16] Laemmli, U.K. (1970) Nature 227, 680–685.
[17] Nakamura, K., Iwasaki, Y. and Hattori, T. (1986) Gene 44, 347–351.