

# Nucleotide and primary sequence of a major rice prolamine

Woo Taek Kim and Thomas W. Okita

*Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340, USA*

Received 11 January 1988; revised version received 2 March 1988

A recombinant cDNA clone encoding a major rice seed storage prolamine was isolated by antibody screening of a cDNA  $\lambda$ gt 11 library. This clone contained a single open reading frame encoding a putative rice prolamine precursor ( $M_r$  17 300). In contrast to other cereal prolamines, the primary sequence of the rice prolamine was devoid of any major tandem repetitive sequences, a feature prevalent in all cereal prolamines studied to date. No significant homology was detected between the rice prolamine and other cereal prolamines, indicating that the rice gene evolved from unique ancestral DNA segments.

Rice; Storage protein; Prolamine; cDNA

## 1. INTRODUCTION

Most cereal grains synthesize as major storage proteins of endosperm tissue prolamines with their characteristic solubility in alcohol-water solutions. In contrast to this, the prolamines of the rice grain comprise only 5–10% of the total grain protein which instead consists of up to 80% of glutelins [1,2]. The prolamine fraction obtained by differential extraction procedures of rice endosperm was found to contain a major component of 13–14 kDa [3–5] and immunoprecipitation of in vitro translation products using partially purified rice prolamine antibody revealed the synthesis of a 16 kDa precursor form presumably containing a signal peptide [5]. As the first step to study the molecular aspects of these genes, we have isolated and sequenced a near full-length cDNA clone. The derived primary sequence of this protein is unique among cereal prolamines as it lacks significant tandem repetitive sequences, a feature displayed by all cereal prolamine genes isolated to date.

*Correspondence address:* T.W. Okita, Institute of Biological Chemistry, Washington State University, Pullman, WA, 99164-6340, USA

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession number Y00747

## 2. MATERIALS AND METHODS

Rice (*Oryza sativa* L. cv. Biggs M-201) was grown in an environment-controlled growth chamber described in detail [5]. Poly(A)<sup>+</sup> RNA was isolated from developing seeds (15 days after flowering) by the method of Reeves et al. [6]. A cDNA library of rice seed poly(A)<sup>+</sup> RNA was constructed using a lambda gt 11 expression vector and clones screened by using a partially purified rice prolamine antibody as described by Huyhn et al. [7]. Northern blot analysis using glyoxal agarose and nick-translated cDNA insert was performed as described in [8]. DNA sequencing was accomplished using the procedures of Henikoff [9].

## 3. RESULTS AND DISCUSSION

In a previous study [5], the specificity of a partially purified rice prolamine antibody was examined by immunoprecipitation of in vitro translation products of rice seed poly(A)<sup>+</sup> RNA as well as Western blot analysis of seed protein extracts. The prolamine antibody exhibited high specificity and showed no cross-contamination with other polypeptides. This antibody was used to screen the rice seed cDNA library constructed in the  $\lambda$ gt 11 expression vector [7]. Several cDNA clones containing inserts of 500–870 bp were isolated and one clone, pProl.7, containing a 870 bp insert, was used for subsequent experiments.

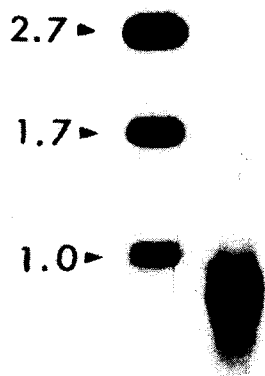


Fig.1. Northern blot analysis of rice seed mRNA. Two  $\mu$ g of poly(A)<sup>+</sup> RNA from 15 DAF rice seeds was glyoxylated and resolved on a 1% agarose gel using glyoxylated *Eco*RI/*Sca*I fragments of pUC19 as size markers. The gel was blotted onto nitrocellulose followed by hybridization to the nick-translated rice prolamine cDNA clone.

Poly(A)<sup>+</sup> RNA obtained from 15 DAF of seed development was analyzed by the Northern blot-hybridization technique using a <sup>32</sup>P-labeled purified cDNA insert as a probe. As shown in fig.1, only a single band of about 900 nucleotides was detected, suggesting the cDNA insert of pProl.7 was near full-length.

DNA sequence analysis showed that the prolamine transcript contained a single long open reading frame encoding 149 amino acids. The prolamine mRNA possesses a relatively long 5'-untranslated region of about 200 bases followed by the 447 nucleotides coding region (fig.2). The 5'-untranslated region was highly enriched for G-C nucleotides and was devoid of any potential AUG translational initiation sites. The 3'-untranslated region contained about 100 bases, and two putative polyadenylation signals (AATAAG and AATAAA) were present. The encoded protein was 17.3 kDa which was in good agreement with the value obtained by SDS-polyacrylamide gel electrophoresis of the prolamine precursor [5]. The first 14 amino acid residues are characteristic of a typical signal peptide containing a penultimate basic amino acid, lysine, followed by 12 hydrophobic amino acids. Cleavage of the signal peptide is likely to occur between alanine (residue 14) and cysteine since this site confirms best to Von Heijne's rules [10]. The deduced mature protein was 15.4 kDa and possess-

TTGCTCCTTC -200

CCGTCCTCCCCGCTTGGGCTCTTGGGCGCCCGTTCCGGGCGCCCCCTCCCTCCTCCCTCCGCGGTACCCGGCCGCTCACTCCTCTGCTGGACCCCC -100

GGCCGCCCCGGGCCGCGCCCCATCCCGGTGCGCGACCCATCGTTACACAGTTCAAGCATTATACAGAAAAATAGAAAGATCTAGTGTCCCGCAGCA -1

ATGAAGATCATTTTCGTCTTTGCTCTCCTTGCTATTGCTGCATGCAGGCCTCTGCCGAGTTTGATGTTTTAGGTCAAAGTTATAGGCAATATCAGC 100

M K I I F V F A L L A I A A C R P L P S L M F L G Q S Y R Q Y Q

TGCAGTCGCCTGTCTGTCTACAGCAACAGGTGCTTAGCCCATATAATGAGTTCGTAAGGCAGCAGTATGGCATAGCGCAAGCCCTTCTTGCAATC 200

L Q S P V L L Q Q Q V L S P Y N E F V R Q Q Y G I A A S P F L Q S

AGCTGCATTTCAACTGAGAAATAACCAAGTCTGGCAACATCAGGCTGGTGGCCAACAATCTCGCTATCAGGACATTAACTTGTTCAGGCCATAGCG 300

A A F Q L R N N Q V W Q H Q A G G Q Q S R Y Q D I N I V Q A I A

TACGAGCTACAACCTCCAGCAATTTGGTGATCTCTACTTTGATCGGAATCAGGCTCAAGCTCAAGCTCTATTGGCTTTTAACGTGCCATCTAGATATG 400

Y E L Q L Q Q F G D L Y F D R N Q A Q A Q A L L A F N V P S R Y

GTATCTACCTAGGTACTATGGTGCACCCAGTACCATTACCACCCTTGGCGGTGTCTTGAATGTGTTTAAACAGTATAGTGGTTCGGAAGTTAAAG 500

G I Y P R Y Y G A P S T I T T L G G V L

ATAAGCTCAGATATCATCATATGTGACATGTGAAACTTTGGGTGATATAAATAGAAATAAAAGTTGCCTTTCATATTT polyA

Fig.2. Nucleotide sequence and deduced amino acid sequence for rice prolamine cDNA clone. Numbers are bp relative to the translation start site. Putative polyadenylation signals are underlined. Arrowhead indicates the possible cleavage site for signal peptide.

Table 1  
Amino acid composition (%)

	pProl.7	[11]	[12]
Ala	8.9	6.6	9.5
Val	5.2	6.5	7.0
Leu	11.9	12.8	12.3
Ile	4.4	5.0	4.4
Pro	5.9	6.1	5.5
Phe	5.2	6.4	4.4
Trp	0.7	0.6	1.6
Met	0.7	0.0	0.8
Lys	0.0	0.0	1.0
His	0.7	trace	1.7
Gly	6.7	5.2	5.1
Thr	2.2	2.2	1.3
Cys	0.7	0.0	trace
Tyr	8.2	9.6	6.4
Asp	6.7	7.6	8.3
Glu	19.3	25.2	19.9

ed a typical prolamine composition, a high mole percentage of glutamine, proline and aromatic amino acids and relatively low levels of charged residues (table 1). The amino acid composition is in good agreement with the known prolamine composition [11,12] confirming the identification of this clone. The primary sequence of the mature prolamine protein was devoid of any large ( $\geq 7$  amino acids) tandemly repeated peptides, a common feature prevalent in all prolamine genes analyzed so far from maize, wheat, barley and rye [13]. The prolamine protein does contain small repetitions dispersed throughout the primary sequence; 12 doublets of amino acids, one triplet, and three consecutive Gln-Ala peptides (at 348 bp). Homology matrix plots revealed that the rice prolamine displayed no significant homology at the DNA or protein level with prolamines of wheat (gliadin) or maize (zein). This result shows

that the rice prolamine is structurally distinct from other cereal prolamines. In view of these differences at both the DNA and protein level, the rice prolamine gene probably evolved from a different ancestral DNA segment responsible for the origin of the prolamine genes of the other major cereals [13].

*Acknowledgements:* This work was supported by a grant from the Rockefeller Foundation. Project 0590 of the College of Agriculture and Home Economics Research Center, Washington State University.

## REFERENCES

- [1] Tecson, E.M.S., Esmana, B.V., Lontok, L.P. and Juliano, B.O. (1971) *Cereal Chem.* 48, 181–186.
- [2] Villareal, R.M. and Juliano, B.O. (1978) *Phytochemistry* 17, 177–182.
- [3] Yamagata, H., Sugimoto, T., Tanaka, K. and Kasai, Z. (1982) *Plant Physiol.* 70, 1094–1100.
- [4] Yamagata, H., Tamura, K., Tanaka, K. and Kasai, Z. (1986) *Plant Cell Physiol.* 27, 1419–1422.
- [5] Krishnan, H.B. and Okita, T.W. (1986) *Plant Physiol.* 81, 748–753.
- [6] Reeves, C.D., Krishnan, H.B. and Okita, T.W. (1986) *Plant Physiol.* 82, 34–40.
- [7] Huyhn, T.V., Young, R.A. and Davis, R.W. (1984) in: *Cloning Technique: A Practical Approach* (Glover, D. ed.) Constructing and Screening cDNA Library in Lambda gt 10 and Lambda gt 11, IRL Press, Oxford.
- [8] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- [9] Henikoff, S. (1984) *Gene* 28, 351–359.
- [10] Von Heijne, G. (1986) *Nucleic Acids Res.* 14, 4683–4690.
- [11] Mandac, B.E. and Juliano, B.O. (1978) *Phytochemistry* 17, 611–614.
- [12] Padhye, W. and Salunkhe, D.K. (1979) *Cereal Chem.* 56, 389–393.
- [13] Kreis, M., Shewry, P.R., Forde, B.G., Forde, J. and Mifflin, B.J. (1985) *Oxf. Surv. Plant Mol. Cell Biol.* 2, 253–317.