

Update section

Sequence

PCR cloning of the full-length cDNA for the seed protein canavalin from the jack bean plant, *Canavalia ensiformis*

Joseph D. Ng, Timothy Stinchcombe, Tzu-Ping Ko, Edwin Alexander and Alexander McPherson
Department of Biochemistry, University of California at Riverside, Riverside, CA 92521, USA

Accepted 13 August 1991

Key words: *Canavalia ensiformis*, canavalin, cDNA clone, polymerase chain reaction

Canavalin is the major storage protein in jack bean (*Canavalia ensiformis*) belonging to the classical vicilin fraction. To obtain detailed genetic and structural information on the reserve proteins of legumes a full-length cDNA for canavalin was generated by use of the polymerase chain reaction (PCR) [1]. A single-stranded cDNA template synthesized from total jack bean RNA by reverse transcriptase [2] was used with primer pairs selected from the 5' and 3' sequence of the cDNA coding for the vicilin protein from *Canavalia gladiata* [3]. The primers were 5' primer 5'-ATCATCCCCTCACACTGCAATACCA-3' and the 3' primer 5'-AGAGAGAAAAGAAAA-CATGATGGT-3'. The reaction conditions included the standard PCR buffer [1] using 25 pmol of each primer in a 100 μ l volume. PCR was carried out using a DNA Thermal Cycler (Ericomp) for 90 s at 96 °C initial denaturation, followed by 30 cycles for 30 s at 55 °C, 2 min at 72 °C, 15 s at 96 °C. A final extension step was performed for 2 min at 55 °C and 7 min at 72 °C. The fragment obtained was directly sequenced by Sanger's dideoxy method [4] with the Sequenase kit version 2.0 (United States Biochemical, Cleveland, OH) following the directions supplied by the manufacturer with slight modifications.

The direct sequencing of the PCR-cloned cDNA shows a very similar nucleotide and de-

duced amino acid sequence to the vicilin protein isolated from *C. gladiata* seeds [3]. Only 5 nucleotides differ: C to G at base 505, C to T at base 889, C to A at base 1058, A to C at base 1166 and A to T at base 1213. There are two base changes that are silent while the other three generate the amino acid substitutions; asparagine to lysine at residue 161 (AAC to AAG), leucine to isoleucine at residue 346 (CTC to ATC) and asparagine to histidine at residue 382 (AAC to CAC).

We previously demonstrated by amino acid sequence homology searches using *C. gladiata* that canavalin contains an internal sequence homology which indicates that its gene was evolved by duplication and mutation of an ancestral genetic domain [5]. We further showed by comparison with other seed reserve proteins whose sequences were known, that this gene structure is a common feature of all of the vicilin class proteins. In addition, this same ancestral genetic domain was found to comprise a portion of the legumin class proteins as well [5].

The internal amino acid sequence homology is indicated in Fig. 1 by double underscoring. The two homologous segments, found in the amino terminal half and the carboxyl terminal half of the protein, have been shown by X-ray diffraction analysis of the three-dimensional structure of the protein to fold into compact eight-stranded anti-

```

1 ATTTTGTATTTAGTAAACCAATATGGCTTTTTCTGCGCGATTTCCTACTATGGTTATTGCT 60
1 M A F S A R F P L W L L L 13
61 GGGAGTTGTTTTGCTTGCTTCAGTTTCTGCGTCGTTTGGCGACTCGGGACACAGTGGAGG 120
14 G V V L L A S V S A S F A H S G H S G G 33
121 AGAAGCAGAGGACGAGAGTGAAGAGTACACGGGCACAAAATAACCCGTATCTCTTTAGGTC 180
34 E A E D E S E E S R A Q N N P Y L F R S 53
181 CAACAAGTTCTCACTCTCTTCAAGAACCAACACGGTTCTCTCGTCTCCTCCAAGGTT 240
54 N K F L T L F K N O H G S L R L L Q R F 73
241 CAACGAAGACACCGAGA AACTGGAGAATCTTCGAGACTACCGAGTTCTTGAATATTGCTC 300
74 N E D T E K L E N L R D Y R V L E Y C S 93
301 CAAACCAAACACCCCTCCTTCTCCCTCACCCTCCGATTCTGATCTTCTCGTCTTCTCCT 360
94 K P N T L L L P H H S D S D L L V L V L 113
361 CGAGGGACAAGCCATACTTGTTTTTGGTGAACCTTGACGGCAGAGACACTTACAACTTGA 420
114 E G Q A I L V L V N P D G R D T Y K L D 133
421 CCAAGGGCATGCTATCAAATCCAAGCAGGGACCCCTTCTATCTCATTAAACCCAGACAA 480
134 Q G D A I K I O A G T P F Y L I N P D N 153
481 CAACCAGAACCCTCAGAAATATAAGTTCCGCATAACCTTCAGGAGACCCGGGCACAGTCCA 540
154 N O N L R I L K F A I T F R R P G T V E 173
541 GGATTCTCTCTATCTAGCACTAAAAGACTGCCATCCTACCTGAGTGCCTTCAGCAAGAA 600
174 D F F L S S T K R L P S Y L S A F S K N 193
601 TTTTCTAGAGGCTCCTACGATTCGCCATATGACGAGATAGAGCAGACTCTGTTGCCAAGA 660
194 F L E A S Y D S P Y D E I E Q T L L Q E 213
661 AGAACAAGAGGGAGTGATAGTCAAATGCCAAGGATCAGATCCAGGAAATAAGCAAACA 720
214 E Q E G V I V K M P K D Q I Q E I S K H 233
721 TGCCCAATCTAGCTCCAGAAAACACTTTTCCCAAGATAAACCATTTAACTTGAGAAG 780
234 A Q S S S R K T L S S Q D K P F N L R S 253
781 CCGAGACCCCATCTATCCAACAACCTATGGCAAGTTATATGAGATCACTCCAGAGAAAA 840
254 R D P I Y S N N Y G K L Y E I T P E K N 273
841 CTCACAGCTACGGGACTTGGATATCCTCCTCAATTGTTTACAAATGAATGAGGGAGCTCT 900
274 S O L R D L D I L L N C L O M N E G A L 293
901 TTTTGTGCCACTACAATTCAAGGGCCACAGTCATACTGGTGGCTAATGAAGGAAGGC 960
294 F V P H Y N S R A T V I L V A N E G R A 313
961 AGAGTTGAGTTGGTGGTCTAGAACAGCAACAACAGCAAGGATTAGAAAGTATGCAACT 1020
314 E V E L V G L E Q O O O Q G L E S M O ' L 333
1021 GCCGAGGTACGCTGCCACGTTATCTGAAGGCGATATAATCGTAATCCCTCGTCTTTTCC 1080
334 R R Y A A T L S E G D I I V I P S S F P 353
1081 GGTTCGCCCTCAAAGCTGCTTCAGATCTAAATATGGTTGGGATTGGTGTCAATGCTGAAAA 1140
354 V A L K A A S D L N M V G I G V N A E N 373
1141 TAACGAGAGGAACCTTCTTGCAAGTCCACAAAGAGAACCTGATAAGGCAGATACCTAGGCA 1200
374 N E R N F L A G H K E N V I R Q I P R Q 393
1201 AGTGAGTGATCTTACATTCCCTGGATCTGGTGAAGAGGTTGAGGAGTTATTAGAGAATCA 1260
394 V S D L T F P G S G E E V E E L L E N Q 413
1261 AAAGGAATCCTACTTTGTGGATGGTCAGCCTAGGCATATTGACGCTGGTGGAAAGGCTAG 1320
414 K E S Y F V D G Q P R H I D A G G K A R 433
1321 AAGGGCCATCTGCCTAATCTTTTCCGCACCTTTTACTGAATAAACTATCTAAGTTACTA 1380
434 R A H L P N L F R T F Y 453
1381 AATAAATGCTGTAAGCAAAG 1403

```

Fig. 1.

parallel β -barrel secondary structures [6]. These two antiparallel β -barrels, commonly described as 'Swiss roll' barrels, are extremely similar and reflect in a structural sense the genetic duplication seen at the level of the DNA.

The cDNA and amino acid sequence we report here is of immediate value in that it has allowed us to assign amino acid side chains to the three-dimensional structure of the protein. This three-dimensional structural model in turn allows us to identify regions and points where mutations or insertions may be made without probable pertur-

bation of the overall structure and without affecting important physical properties. Through the use of site-directed mutagenesis of the cloned cDNA that we describe here, guided by knowledge of the protein structure, we intend to modify the protein in ways that will enhance its nutritional value. Because canavalin typifies the major reserve proteins of all of those we have so far studied, we are hopeful that this work will have a substantial impact in the long term on the improvement of dietary protein sources for man and his domestic animals.

Acknowledgements

We would like to thank Bob Cudney for his assistance in this project and also Sharon Kochek for her helpful discussions. This study was supported by grants from the National Institute of Health (GM40706) and the National Aeronautics and Space Administration (NAG8-804).

References

1. Mullis KB, Faloona FA, Scharf SJ, Saiki RK, Horn GT, Erlich HA: Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. Cold Spring Harbor Symp Quant Biol 51: 263–273 (1986).
2. Krug MS, Berger SL: First-stranded cDNA synthesis primed with oligo(dT). Meth Enzymol 152: 316–325 (1987).
3. Yamauchi D, Nakamura K, Asahi T, Ninamikawa T: cDNAs for canavalin and concanavalin A from *Canavalia gladiata*. Eur J Biochem 170: 515–520 (1988).
4. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
5. Gibbs PEM, Strongin KB, McPherson A: Evolution of legume seed storage protein: a domain common to legumins and vicilins is duplicated in vicilins. Mol Biol Evol 6: 614–623 (1989).
6. Ko TP, Ng JD, Day J, Greenwood A, Greene M, McPherson A: Cloning, expression and crystallographic analysis of the reserve proteins of the jack bean. In preparation. (1991).