

ARCELIN SEED STORAGE PROTEINS FROM PHASEOLUS VULGARIS

This application is a continuation of application Ser. No. 07/180,404, filed Apr. 12, 1988, now abandoned.

TECHNICAL FIELD

The present invention is directed to plants and plant proteins. More particularly, the present invention is directed to arcelin seed storage proteins from the *Phaseolus vulgaris*, nucleic acid sequences encoding such proteins, and the use of such nucleic acid sequences.

BACKGROUND

There are two commonly known seed storage proteins in the cultivated common bean, *Phaseolus vulgaris*: phytohemagglutinin (PHA, also referred to as bean lectin) and phaseolin. Recently, wild forms of *Phaseolus vulgaris* L., indigenous to Middle and South America, have been found to contain a novel family of proteins previously unreported in the common bean. These proteins, named arcelins, have subunit molecular weights similar to lectin proteins or intermediate to phaseolin and lectin proteins, and occur in the globulin-2 protein fraction. Four electrophoretic variants, or isoproteins, have been observed and designated arcelin-1, -2, -3 and -4. See Romero Andreas et al. (1986) *Theor Appl. Genet.* 72:123-128; Osborn et al. (1986) *Theor. Appl. Genet.* 71:847-855.

The presence of arcelin in wild beans has been correlated with resistance to two bruchid beetle species. It is not known, however, whether this resistance is attributable to arcelin in whole, or even in part. See Osborn et al. *supra*; Schoonhoven et al. (1983) *J. Econ. Entomol.* 76:1255-1259.

Other seed storage proteins have been cloned and expressed in heterologous plants. For example, a sample for phaseolin protein from *Phaseolus vulgaris* (French bean) has been cloned and expressed in heterologous plants under the control of its own promoter and heterologous promoters. See, e.g., Murai et al. (1984) *Science* 222:476; Segupta-Gopalan et al. (1985) *Proc. Natl. Acad. Sci. USA* 82:3320; EPO Pub. No. 126,546; EPO Pub. No. 122,791. Heterologous plants have also been transformed by the gene for the corn storage protein, zein. Matzke et al. (1984) *EMBO J.* 3:1525-1531; Messing in *Genetic Engineering* 6:1-46 (W. J. Rigby ed. 1987). The Brazil nut 2S storage protein has also been expressed in heterologous plants. See copending U.S. Pat. App. Ser. No. 065,303, filed Jun. 19, 1987.

SUMMARY OF THE INVENTION

The present invention provides nucleic acid sequences, particularly DNA sequences, encoding arcelin. It has also been determined that the arcelin proteins alone are toxic to bean bruchid pests, and the transfer of the gene encoding arcelin to bean cultivars results in insect resistance. Thus, the present invention provides nucleic acid sequences which are useful in the genetic transformation of plants to improve nutritional value and to introduce insect resistance.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 shows the nucleotide sequence and derived amino acid sequence of an arcelin cDNA (arcelin-1) from plasmid pAR1-11. The 265 amino acid open reading frame is shown. The third ATG is presumed to be the initiation codon. The mature protein N-terminal

sequence, as determined by Edman degradation, is underscored.

FIG. 2 shows a comparison of the nucleotide-derived amino acid sequences of PHA-L (first line, Hoffman et al., 1985, *EMBO J* 4:883), PHA-E (second line, id.), arcelin-1 (third line) and a "lectin-like" protein (fourth line, Hoffman et al., 1982, *Nucl. Acids Res.* 10:7819).

DETAILED DESCRIPTION OF THE INVENTION

In addition to the techniques described below, the practice of the present invention will employ conventional techniques of molecular biology, microbiology, recombinant DNA technology, and plant science, all of which are within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Maniatis et al., *Molecular Cloning: A Laboratory Manual* (1982); *DNA Cloning: Volumes I and II* (D. N. Glover ed. 1985); *Oligonucleotide Synthesis* (M. J. Gait ed. 1984); *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1985); *Transcription and Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Animal Cell Culture* (R. I. Freshney ed. 1986); *Plant Cell Culture* (R. A. Dixon ed. 1985); *Propagation of Higher Plants Through Tissue Culture* (K. W. Hughes et al. eds. 1978); *Cell Culture and Somatic Cell Genetics of Plants* (I. K. Vasil ed. 1984); Fraley et al. (1986) *CRC Critical Reviews in Plant Sciences* 4:1 (hereinafter *Plant Sciences*); *Biotechnology in Agricultural Chemistry: ACS Symposium Series 334* (LeBaron et al. eds. 1987); *Genetic Engineering*, Vol. 6 (Rigby ed., 1987).

In describing the present invention, the following terminology will be used in accordance with the definitions below.

A "replicon" is any genetic element (e.g., plasmid, cosmid, chromosome, virus, etc.) that behaves as an autonomous unit of DNA replication in vivo; i.e., capable of replication under its own control within a cell.

A "vector" is a replicon, such as a plasmid, cosmid, or bacteriophage, to which another DNA segment may be attached so as to bring about replication of the attached segment, or to allow its introduction into a cellular host.

A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in either single- or double-stranded form. The double-stranded form may, of course, be implied by the context of the molecule's environment (e.g., a chromosome). When in the double-stranded form, the molecule will usually be in its normal, double-stranded helix. The term "DNA molecule" is not limited to any particular tertiary form of DNA. Thus, the term includes double-stranded DNA found, inter alia, in linear DNA molecules, viruses, plasmids, and chromosomes. When discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction (left to right) along the nontranscribed (anti-sense) strand of DNA (i.e., the strand having a sequence homologous to the mRNA). If both strands are shown, the anti-sense strand will be on top.

A DNA "coding sequence" is a DNA sequence which is transcribed and translated into a polypeptide in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by the start codon at the 5'