G&BM
Tome IV
Iasi, 2003

IZOLATION AND CHARACTERIZATION OF cDNA CLONE CODING FOR LOTUS JAPONICUS NODULE CARBONIC ANHYDRASE

DANIELA COTZUR[1], PANAGIOTIS KATINAKIS[2]

Key words: *Lotus japonicus*, clone, carbonic anhydrase (CA), root nodule

Abstract: To study the expression of genes coding for enzymes involved in carbon metabolism, we were concerned with the isolation and characterization of cDNA clone coding for carbonic anhydrase. The nucleotide sequence of the cDNA clone encoding *Lotus japonicus* carbonic anhydrase was determined. The deduced amino acid sequence of this clone was aligned with other homologous carbonic anhydrase sequences, derived from different organisms, in order to identify the conserved regions.

THE AIM OF INVESTIGATIONS

The aim of the study was to describe the isolation and characterization of a cDNA clone encoding for a *Lotus japonicus* nodule specific β-type carbonic anhydrase, whose predicted amino acid sequence is conserved in plant kingdom.

INTRODUCTION

The interaction of the soil bacteria *Mesorhisobium, Azorhizobium*, *Sinorhizobium* and *Bradyrhizobium* with the root system of leguminous plants leads to the formation of a specialized plant organ, the root nodule. During the course of nodule initiation and development a number of plant genes, referred to as nodulin genes, are induced (Handberg and Stougaard, 1992; Mylona et al., 1995). Recently, however, the number of nodulin genes whose expression is also found in tissues other than nodules has increased considerably. For example, one of the nodulin genes expressed at very early stages of nodule initiation (ENOD40) is also expressed in roots, stems and embryos (Papadopoulou et al., 1996, Fang and Hirsh, 1998, Flemetakis et al., 2000 and 2002). Homologues of several nodulin genes including leghemoglobin, NOD93, NOD35 and ENOD40 have been found in legume and non-legume plants (Kouchi et al., 1999, Marchfelder et al., 1997, Reddy et al., 1998).

The presence of a non-photosynthetic related carbonic anhydrase activity in the plant fraction of nodules was first reported by Atkins in 1974. Moreover, recent studies also verified the presence of nodule carbonic anhydrase activity in a number of legumes including *Pisum sativum, Vicia faba* and *Lupinus angustifolius* (Atkins et al., 2001). Although carbonic anhydrase transcripts have been identified in the nodules of soybean (Kavroulakis et al., 2000) and alfalfa (Coba de la Pena et al., 1997), the physiological role of carbonic anhydrase is not yet clear.

Advances in DNA sequencing technology have resulted in the generation of a large amount of nucleotide sequence data in a short period of time.

An important method to obtain information on genes expressed in an organism is the systematic sequencing of cDNA clones, generally known as the EST (Expressed Sequence Tag) approach (Boguski, 1995; Claverie, 1995). It is the fastest way of gene discovery in eukaryotes. EST serves as markers for genes

expressed by a certain cell type or tissue under specific environmental (culture) conditions, and are used for the discovery of full length cDNA or genomic clones, discovery of new genes, recognition of exon/intron boundaries and development of genetic maps. Furthermore, ESTs with no homology to known proteins may provide the first clues of new proteins (Baxevanis et al., 1998).

Genome analysis in higher eukaryotes was initially carried out by cataloguing only the expressed portions of the genome by developing anonymous partial cDNA sequences, EST (expressed sequence tags), rather than genome sequencing. A 300–500 bp –long sequence thus obtained is sufficient to identify a gene by similarity search against the public databases. The specificity of expression as well as expression levels of genes can be estimated by generating EST sequences from different organs or developmental stages and comparing the frequency of the particular sequence in the dataset. As a contribution to genome analysis, cDNA structural information is indispensable for the precise assignment of genes on a genomic sequence, usually done based on similarities to known genes and predictions by computer programs.

To make a catalogue of genes expressed in *L. japonicus* L. and understand biological processes specific to legume plants, large-scale EST analyses have been performed.

With the aim of understanding the genetic system related to legume-specific biological processes, recently, large numbers of expressed sequences tags (ESTs) from *Lotus japonicus* nodules have been deposited in public databases and analyzed by DNA arrays for transcriptome analysisThe sequence information and search results of these Lotus japonicus clones, generated in different studies, are available at the web sites: http://www.kazusa.or.jp/en/plant/lotus/EST and http://www.agowa. Similarity search was performed against the public EST database using the BLAST program. A BLAST score of more than 80 is generally regarded as a significant match (Pearson, 1997).

This enables identification of genes coding for enzymes involved in diverse metabolic pathways, whose expression is induced during nodule formation and functioning (Asamizu et al., 2000).

MATERIALS AND METHODS

Isolation of cDNA clones for Lotus japonicus specific β-type carbonic anhydrase

In a previous study a cDNA clone coding for *Glycine max* nodule carbonic anhydrase (GmCA1) has been isolated and characterized at the Agricultural University of Athens, Department of Agricultural Biotechnology (Kavroulakis et al., 2000). For the isolation of *L. japonicus* clones homologues to GmCA1, two αZAPII *L japonicus* cDNA libraries from 9- and 21-day-old nodules (kindly provided by Dr. J. Stougaard, University of Aarhus, Denmark) were screened by plaque hybridization using ^{32}P-dCTP-labeled insert of the GmCA1 cDNA clone as a probe. Hybridization was performed under low stringency conditions at 35ºC in the presence of 25% formamide. Plaque purification and in vivo excision of the pBlueScriptSK^{+} phagemid from the positive αZAPII recombinant bacteriophages were performed according to the standard protocols (Stratagene, La Jolla, CA, U.S.A.). The nucleotide sequence of the cDNA clones, giving positive hybridization signal, was determined by the dideoxi chain termination method (Sanger et al., 1997, Sambrook et al., 1989).

The complete nucleotide sequence of the cDNA clone coding for β-type carbonic anhydrase was translated into amino acid sequence and compared with the published sequences in the GenBank database, using the BLAST program at the ExPASy National Centre for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/BLAST/).

The deduced amino acid sequence of these clones were aligned with other homologous carbonic anhydrase sequences, derived from different organisms, in order to identify theconserved regions.

RESULTS AND DISCUSSIONS

Isolation of cDNA clones for Lotus japonicus specific β-type carbonic anhydrase

In order to isolate a cDNA clone coding for *L.japonicus* nodule carbonic anhydrase, a αZAPII cDNA library prepared from 21-days-old nodules was screened under low stringency conditions using a previously published *Medicago sativa* MsCA1 cDNA clone (Coba de la Pena et al., 1997) as heterologous probe.

Characterization of cDNA clone coding for β-type carbonic anhydrase

An CA-β-type cDNA clone (designated Lj CA1) contained an open reading frame of 1088 base pairs encoding a polypeptide of 263 amino acids. The nucleotide sequences corresponding to the longest open reading frame (ORF) are shown in Fig. 1.

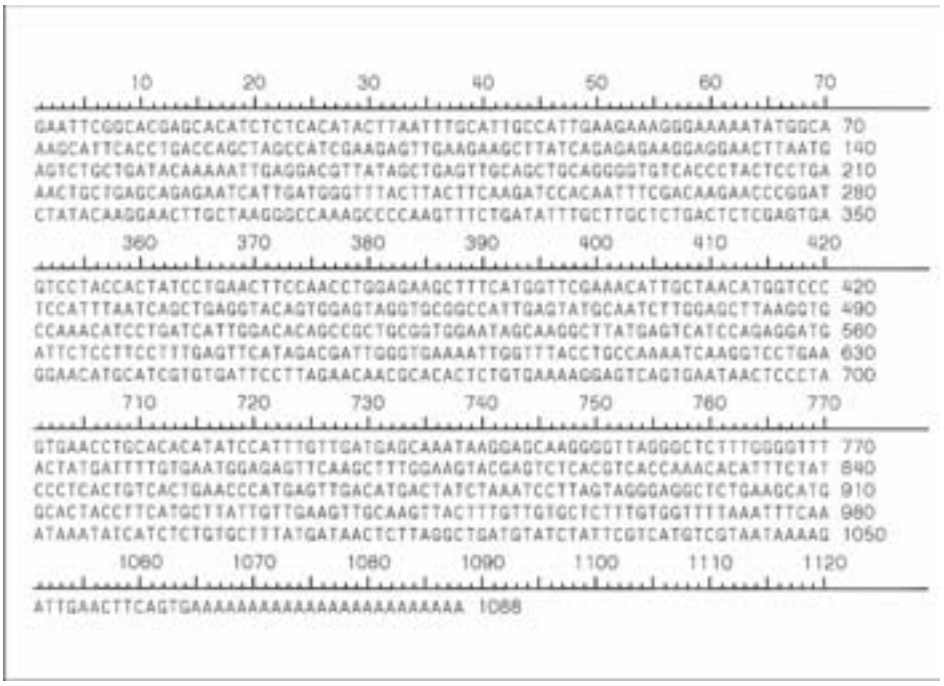

Figure 1: The nucleotide sequences of the cDNA clone encoding *Lotus japonicus* L. β-carbonic anhydrase

The largest cDNA clone, designated LjCA1 contained an open reading frame (ORF) of 263 amino acids, including the starting methionine, indicating that LjCA1 represents a full-length cDNA clone.

Comparative analysis of the deduced amino acids sequence from LjCA1 clone with previously characterized nodule carbonic anhydrase polypeptides (Fig.2), revealed that LjCA1 exhibits 77.4% and 80.2% similarity with *M. sativa* and *G. max* nodule carbonic anhydrases respectively. Lower percentage of similarity, around 45%, was observed with carbonic anhydrases proteins derived from photosynthetic plant tissues (Fig.2). Like all nodule carbonic anhydrases characterized so far, LjCA1 does not posses the N-terminal transient peptide that is responsible for the translocation of the photosynthesis-related carbonic anhydrases proteins into the chloroplasts (Fig.2). Apart from the high overall sequence similarity to other plant carbonic anhydrases, the *L japonicus* homologue posses all the amino acids that where proposed to participate in the active site of the *Pisum sativum* β-carbonic anhydrase (Kimber and Pai, 2000). Residues Cys^{90}, His^{150} and Cys^{153} of the LjCA1 can be assigned as the equivalents of Cys^{60}, His^{220} and Cys^{223} of *P. sativum* carbonic anhydrase residues thought to act as the zinc ligants. Similarly, residues Ile^{114}, Tyr^{135} and Phe^{109} can be assigned as the residues that collectively present a continuous hydrophobic surface in the binding pocket, while residues Tyr^{135}, Gln^{81}, Gly^{154} and Asp^{92} form the bottleneck in the hydrophilic channel that in the only access to the active site of β-carbonic anhydrase.

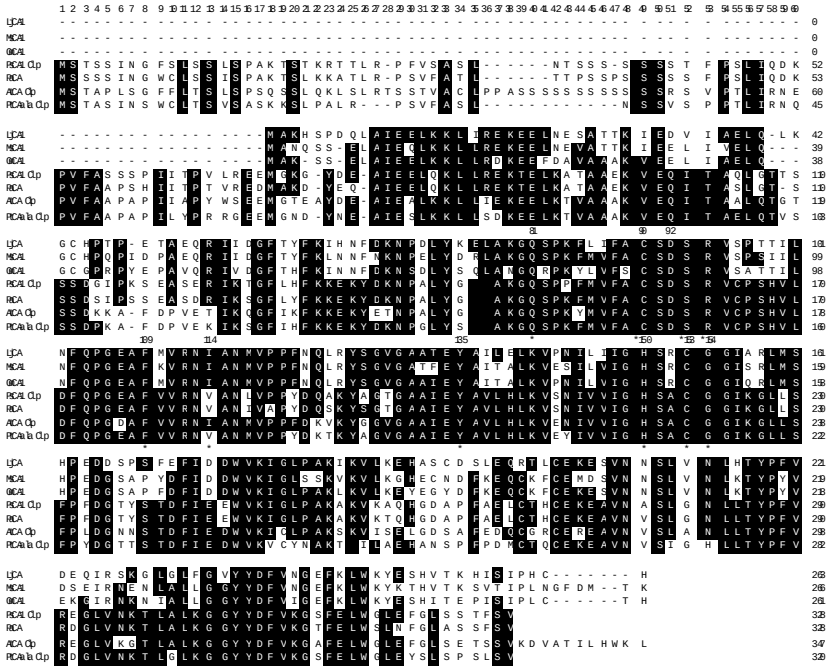

Figure 2. Comparison of the deduced amino acid sequences of *Lotus japonicus* nodule carbonicanhydrase (LjCA1) with various carbonic anhydrases from legume and non-legume plants.

Organism symbols (sequences shown) and GenBank database accesion numbers are as follows: GmCA1 *Glycine max* nodule CA (AJ239132); MsCA1, *Medicago sativa* nodule CA (X93312); PsCA Clp, *Pisum sativum* chloroplastic CA (P17067); PaCA, *Phaseolus aureus* CA (Q9XQB0); AtCA Clp, *Arabidopsis thaliana* chloroplastic CA (AAL16228); PtCA1 Clp *Populus tremula x Populus tremuloides* chloroplastic CA (U55837). Black-shaded boxes represent conserved amino acids, while dashes represent gaps in the alignment. Amino acids in the proposed active site are marked with an asterisk. Sequences were aligned using the Clustal method with PAM250 residue weight table (Altschul et al., 1990).

CONCLUSIONS

As *Lotus japonicus* is one of the two model legumes used to study the function of genes involved in symbiosis, we set out to characterize the structure of the β-carbonic anhydrase as the first step towards elucidating their function.

The deduced amino acid sequence of β-carbonic anhydrase clone was aligned with amino acid sequence carbonic anhydrase of different organisms in order to identify the conserved regions within these sequences.

Comparative analysis of the deduced amino acids sequences of this clone with other known plant carbonic anhydrases amino acids sequences revealed a high level of sequence conservation.

BIBLIOGRAPHY

Altschul, S.F., Gish, W., Miller, W., Myers, E. W. and Lipman, D.J., 1990. Basic local alignement search tool. J. Mol. Biol. 215: 403-410.

Asamizu, E., Nakamura., Y., Sato,, S. and Tabata, S., 2000. Generation of 7137 non-redundant expresses sequence tags from a legume, Lotus japonicus. DNA Res. 7: 127- 130.

Atkins, C.A., 1974. Occurrence and some properties of carbonic anhydrase from legume root nodules. Phytochemistry 13:93-98

Atkins, C.A., Smith, P., Mann, A. and Thumfort, P. 2001. Localization of carbonic anhydrase in legume nodules. Plant, Cell and Environment 24: 317-326

Baxevanis, A.D. and Ouellette, B.F. Francis, 1998 Bioinformatics. A practical guide to the analysis of genes and proteins. Wiley- Interscience. A John Wiley & Sons, INC., Publication, pp. 16-93.

Boguski, M.S., 1995. The turning point in genome research. Trends Biochem Sci. 20: 295-296.

Claverie, J-M.,1995. Exploring the Vast Territory of Uncharted ESTs. In Brown MJ, Thurlby PL (Eds) Genomes, Molecular Biology and Drug Discovery. Academic Press, Long, pp. 55-71

Coba de la Pena, T., Frugier, F., McKhann, H.I., Bauer, P., Brown, S., Kondorosi, A. and Crespi, M.,1997. A carbonic anhydrase gene is induced in the nodule primordium and its cell-specific expression is controlled by the presence of *Rhizobium* during development. Plant Journal 11: 407-420

Fang, Y. and A.M. Hirsh, 1998. Studying early nodulin gene ENOD40 expression and induction by nodulation factor and cytokinin in transgenic alfalfa. Plant Physiol. 116: 53-68

Flemetakis, E., Kavroulakis, N., Qusedvlleg, N.E.M., Spaink, H.S., Dimou, M., Roussis, A., Katinakis, P.,2000. *Lotus japonicus* contains two distinct ENOD40 genes that are expressed in symbiotic, nonsymbiotic and embryonic tissues. Mol. Plant Microb. Inter. 13: 984-987

Flemetakis E., A. Agalou, N. Kavroulakis, M. Dimou, A. Martisikovskaya, A. Slater, H.P. Spaink, A. Roussis and P. Katinakis, 2002. Lotus japonicus gene Ljsbp is highly conserved among plants and animals and encodes a homologue to the mammalian selenium – binding proteins. Molecular Plant – Microbe Interactions, Vol. 15, No.4, p. 390-403

Handberg K. and Stougaard, J., 1992. Lotus japonicus, an autogamous, diploid legume species for classical and molecular genetics. Plant Journal 2: 487-496

Kavroulakis, N., Flemetakis, E., Aivalakis, G. and Katinakis, P., 2000. Carbon metabolism in developing soybean root nodules: the role of carbonic anhydrase. MPMI vol.13, No. 1, p. 14-22

Kimber, M.S. and Pai, E.F., 2000. The active site architecture of *Pisum sativum* β-carbonic anhydrase is a mirror image of that of α- carbonic anhydrases. EMBO J. 19: 1407-1418

Kouchi, H., Takane, K., So, R.B., Ladha, J.K. and Reddy, P.M., 1999. Rice ENOD40: isolation expression analysis in rice and transgenic soybean root nodules 18: 121-129

Marchfelder, A., Binder, S. and Brennicke, A., 1997. A nodulin-35 homologue is encoded in Arabidopsis genome. Trends in Plant Sciences. 2: 167-168

Mylona, P., Pawlowski, K. and Bisseling, T., 1995. Symbiotic nitrogen fixation. Plant Cell. 5: 869-885

Papadopoulou, K., Roussis, A. and Katinakis, P., 1996. Phaseolus ENOD40 is involved in symbiotic and non-symbiotic organogenetic processes: Expression during nodule and lateral root development. Plant Mol. Biol. 30: 403-417

Pearson, W.R., 1997. Identifying distantly related protein sequences. Comput. Appl. Biosci 13: 325-332.

Reddy, P.M., Kouchi, H. and Ladha, J.K., 1998. Isolation, analysis and expression of homologues of the soybean early nodulin gene GmENOD93 (Gm93) from rice. Biochim. Biophys. Acta. 1433: 386-392

Sambrook, J., Fritsch, F.F., Maniatis, T., 1989. *Molecular Cloning. A Laboratory Manual,* 2 nd ed. Cold Spring Harbor laboratory, Cold Spring Harbor,N.Y.

Sanger, F., Nicklen, S. and Coulson, A.R., 1997. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467

1 Corresponding author: "Petru Poni" Macromolecular Chemistry Institute of Iasi, Gr. Gh. Voda 41A, 6600 Iasi, Romania, e-mail:danacotzur@hotmail.com
2 Agricultural University of Athens, Greece