Poster H-90

PCR primer design for 2,4-D degradation genes using BLOCKS and the CDD database

Authors:
VIANEZ JUNIOR, J. L. S. G (*Departamento de Microbiologia Geral, LBSBM, UFRJ*)
MACRAE, A (*Departamento de Microbiologia Geral, LBSBM, UFRJ*)

Short Abstract: Conserved regions within genes were identified after sequential, protein structural and functional analysis and used for PCR primer design of 6 genes involved in the biodegradation of the globally important pesticide 2,4-D. The six primer pairs designed are intended to be used as tools to study the capacity of the local microbial community to biodegradate this herbicide.

Long Abstract:

INTRODUCTION

2,4-Dichlorophenoxyacetic acid (2,4-D) is one of the most used herbicide in the world. In Brazil, 2,4-D is the most widely used herbicide in pastures to control invasive weeds (1). This compound has a short lifetime in soil, due to degradation by microorganisms, that have been show capable of fully degrade this compound. The TFD degradation pathway of the 2,4-D is well established (2), and the key enzymes involved in the process are well known. There are several databases focused in metabolism pathways, including the pathways of several xenobiontics. This, together with the ever increasing availability of nucleotide sequences in public databases has made it possible to develop in silico studies to guide the experimental work in the bioremediation field. The goal of this study was to develop primers, based on knowledge of protein structure using the BLOCKS (3) method, to detect genes involved in the degradation pathway of 2,4-D in environmental samples. The end use of the primers is as tools to study the capacity of the local microbial community of a soil to biodegradate this herbicide.

MATERIALS AND METHODS

The degradation pathway of the 2,4-D can be found at the Biodegradation Database (4). Full length nucleotide and protein sequences coding for the enzymes (genes tfdA, tfdB, tfdC, tfdD, tfdE and tfdF) were retrieved from Genbank (5) and aligned with clustalw (6). The alignment was submitted to the Alignment Processor, in the BLOCKS server to extract short fragments of conserved amino acid sequences, from which a physical map of the localization of the BLOCKS was obtained. Representative sequences from each alignment were submitted to the Conserved Domain Database – CDD – (7) in order to obtain structural/functional information about relevant regions of the sequences. The basic idea behind this procedure is that the blocks extracted from the alignments that match conserved domains found at CDD, represent short stretches of amino acids involved in key structures for the correct function of the enzyme, and are more likely to be conserved even among distantly related sequences. The chosen blocks were submitted to COnsensus-DEgenerate Hybrid Oligonucleotide Primer – CODEHOP – (8), using the default values. Finally, the primer pairs were evaluated for possible primer-primer interactions and melting temperature using the software oligo analyzer (9). The specificity of the selected primers was accessed

by making local alignments of the primers with intended gene sequences and related genes.

RESULTS:

For the tfdA protein sequences, a total of 8 blocks were extracted from the alignment. From these only 2 blocks were consistent with the CDD database and were submitted to CODEHOP. The blocks were from region 37-85, which represents the TauD, TfdA family (Taurine catabolism dioxygenase) and region 291-331, which represents both Clavaminic acid synthetase (CAS) and TauD,TfdA families (primers tfdA320F: CGCCAACCAGTTCTGGCATAGC and tfdA510R: AAGCGGGAATTCAGCGCGTAGT). From the tfdB sequences seven blocks were retrieved and five of these blocks were a significant match against CDD database and satisfied the conditions in CODEHOP (positions 58-113, 113-168,223-263,263-318 and 318-373). Those blocks were associated with a monooxygenase family and a FAD binding domain (primers tfdB360F: GCC AAGTACCTCATTGGAGCGG and tfdB820R: TTGGATGGAGGTGTTGGAGCC). tfdC amino acid sequences yielded a total of eight blocks, of which three satisfied the desired criteria (positions 171-203,253-291 and 381-402). The blocks were associated with the dioxygenase family (primers tfdC260F: TTGTGGACGGCAGGCTGAAGA and tfdC650R: CCAGTCGTCGTCCACCCAGT). From the tfdD sequences, eight blocks were extracted, and three produced significant matches against CDD database and satisfied CODEHOP default values (positions 63-75,215-270 and 270-286). All blocks were associated with the Mandelate racemase family (primers tfdD190F: GGA GTTCCGAGTCTGCGGAGA and tfdD620R: CCAGATCATGGCGGTCTGCTC). Analysis of the tfdE sequences revealed five blocks, from which only two were a consistent matches against CDD database and CODEHOP search (positions 23-62 and 123-154). The matches were associated with the Dienelactone hydrolase family (primers tfdE100F: GTCGTGATCGCCCAGGAGATT and tfdE590R TTTCACATCGAAGACGAACGCCTG) . The tfdF sequences produced a total of eight blocks, and only two satisfied the desired criteria. The blocks were associated with Fe-ADH (iron-containg alcohol dehydrogenase) and EutG (alcohol dehydrogenase class IV) families (primers tfdF340Foward: ATCGTCATCCCCACCACCTATGCC and tfdF730R CCAGGGTGTGGCACA GTTTGTGA).

DISCUSSION

Regardless of the application for which a primer is designed, several criteria must be considered in the process of primer design. In this specific case, it is crucial that the primers are designed in a way that can capture as much genetic diversity as possible, since several microorganisms are involved in the process. To design optimized primers, consensus sequences based on probabilistic computations of aligned sequences, aided by experimental knowledge (structural an functional information) should be used. In this study we used available tools in public domain that are consistent with this more conscientious strategy of primer design. It is valid to warn that when using this approach, one can design a primer that is not specific to the gene of interest, given the weight given to functional domains. To avoid this problem, the selected primers were checked against related gene sequences from public databases. Since the blocks extracted from the sequences are long enough, one can carefully exploit various positions within the blocks in such a way that avoids those kind of false positives. Nevertheless the regions from gene sequences obtained this way represent conserved sequences of amino acids, determined based on sequential, functional and structural information and represent good candidate regions to be exploited to primer design.

REFERENCES

1.Nunes, S.G. 2001. Controle de plantas invasoras em pastagens cultivadas nos Cerrados. Embrapa Gado de Corte, documentos 117.

2.Vallaeys, T et al. 1996. The metabolic pathway of 2,4-dichlorophenoxyacetic acid degradation involves different families of tfdA and tfdB genes according to PCR-RFLP analysis. FEMS Microbiol Ecol 20, 163–172

3.J.G. Henikoff et al. 2000. Increased coverage of protein families with the blocks database servers, Nucl. Acids Res. 28:228-230.

4.Ellis L.B.M. et al. 2006. The University of Minnesota Biocatalysis/Biodegradation Database: The First Decade. Nucleic Acids Research 34: D517-D521

5.Benson DA. et al. 2006. GenBank. Nucleic Acids Res. Jan 1;34

6.Thompson JD. et al. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. Nov 11;22(22):4673-80.

7.Marchker-Bauer A. et al. 2005. CDD: a Conserved Domain Database for protein classification.
Nucleic Acids Res. Jan 1;33(Database issue):D192-6.

8.Rose TM. et al. 2003. CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design.Nucleic Acids Res. Jul 1;31(13):3763-6.

9.Integrated DNA Technologies, Inc.Oligo Analyzer 3.0 (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/)