ELSEVIER

FEMS Microbiology Letters 252 (2005) 215–222

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

DING proteins are from *Pseudomonas*

Alan P. Lewis *, Daniel Crowther

Bioinformatics Discovery and Analysis, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage SG1 2NY, UK

Received 2 June 2005; received in revised form 30 August 2005; accepted 30 August 2005

First published online 15 September 2005

Edited by M.Y. Galperin

Abstract

DING proteins have been described as animal and plant proteins with potential biomineralisation, receptor or signalling roles that have been characterised by an N-terminal DINGGG-sequence. However, these sequences have only ever been identified as either N-terminal peptides or partial cDNA sequences, and have yet to be detected in any of the many genomic animal and plant genomes now available. Microbial relatives of the DING proteins have been described, which appear to be periplasmic phosphate-binding proteins. Recently, full-length *Pseudomonas aeruginosa* UCBPP-PA14 and *Hypericum perforatum* genes have been sequenced that show high homology to the published DING protein N-terminal sequences, and small peptides previously identified in conjunction with the peptide sequencing of DING proteins can also be mapped to regions across these full-length sequences. Searching with these sequences identifies other plant and animal cDNA fragments in the public nucleotide databases, and, additionally, an unordered rat genomic contig that contains a DING-like sequence on a small fragment. Analysing the codon usage of these DNA sequences identifies all of these sequences as of *Pseudomonas* origin, suggesting that DING proteins do not exist in eukaryotes, but instead are potentially due to microbial contamination or infection.

Keywords: DING protein; *Pseudomonas*; Synonymous codon usage; ABC-type phosphate transport system; Periplasmic component

1. Introduction

DING proteins are so named due to their characteristic DINGGG- N-terminal sequence [1]. The first DING protein was initially identified as a 40 kDa rheumatoid arthritis (RA) autoantigen, synovial stimulatory protein (SSP), by N-terminal sequencing of tryptic peptides derived from protein purified from the synovial fluid of RA patients and from the conditioned medium of cultured RA synovial fibroblasts [2]. A 40 kDa protein with an identical N-terminus was independently isolated from human skin fibroblast cytoplasmic extracts and from the conditioned medium of cultured fibroblasts using hirudin-agarose affinity columns, and was initially named the hirudin-sensitive fibroblast proteinase (HSFP) [3]. Further internal tryptic peptides identified from this protein shared no homology to any other known proteins. A 860 bp cDNA was isolated using degenerate and non-degenerate PCR primers derived from internal peptides [3], and was annotated as a putative partial cDNA sequence of the DING gene. However, BLAT searching [4] of the human genome identifies this sequence as a partial cDNA splice variant of the human dishevelled, dsh homolog 3 (*Drosophila*) gene (DVL3; data not shown) [5]. The N-terminal peptide of a 39 kDa protein, crystal adhesion inhibitor (CAI), secreted from human kidney epithelial cells and blocking calcium oxalate monohydrate (COM) crystal adhesion to the cell surface, showing identity to the SSP and HSFP peptides, has also been described [6].

* Corresponding author. Tel.: +44 1438 763625; fax: +44 1438 764231.

E-mail address: alan.p.lewis@gsk.com (A.P. Lewis).

0378-1097/$22.00
doi:10.1016/j.femsle.2005.08.047

The N-terminal sequence of a 40 kDa DING protein has been identified from human urinary stones and gallstones [7]. This sequence, however, shows some amino acid differences when compared to the SSP/HSFP/CAI peptides, suggesting that there is potentially more than one DING protein identified in human.

DING proteins have also been reported in other animals. Peptide sequencing of N-terminal fragments has identified a 40 kDa turkey DING protein, lipid-free polysaccharide-binding protein (LFPBP-40), in air sac fluid [8], and also a 40 kDa DING protein in rat neurones described as a cotinine receptor [9]. Additionally, a number of unpublished observations describing the identification of DING proteins in plants have been revealed in a DING protein review [10].

Even though DING proteins have been isolated on multiple occasions in multiple laboratories, as yet no cDNA or finished genomic DING sequences have been identified. It has been suggested that this discrepancy could be accounted for by low transcript abundance, mRNA instability or rapid turnover, and by gaps still existing in the human genome [10]. Using sequence analysis, we show here that the DING proteins are from *Pseudomonas* species, and do not actually exist in plant or animal genomes.

2. Materials and methods

2.1. Sequence analysis

Sequence databases were searched using the BLAST algorithm [11] with default parameters. Comparison and alignment of peptides to full-length protein sequences was made using a combination of Dotter [12] and manual alignment. Phylogenetic analysis of protein sequences was performed with Phylo_win [13] using the Neighbour Joining method [14], distances being calculated using the PROTDIST program [15] utilising a Dayhoff PAM matrix, and with the "pairwise gap removal" parameter selected to handle the large number of gaps in the DING sequences constructed from short peptides. Synonymous codon usage of nucleotide sequences was analysed using the EMBOSS [16] program, syco, using an averaging window of 30, and a minimum common codon value of 0.15. This program compares synonymous codon usage of DNA across all three forward frames with selected codon usage tables to generate a Gribskov statistic plot [17]. Species-specific codon usage tables were obtained from EMBOSS [16], apart from the *Leishmania* table, which was taken from the Codon Usage Database [18]. Codon usage tables for the individual sequences were generated using the EMBOSS [16] program, cusp, and codon usage tables were compared using the EMBOSS [16] program, codcmp.

3. Results and discussion

3.1. Full-length DING proteins

A full-length DING protein from *Pseudomonas aeruginosa* UCBPP-PA14 (RefSeq Accession No. ZP_00138283; periplasmic component of the ABC-type phosphate transport system, PstS) and an almost full-length cDNA from *Hypericum perforatum* (St. John's Wort; GenBank Accession No. AY866430; p27SJ, annotated as being able to inhibit HIV-1 gene expression and replication) have recently been deposited in the public databases. The mammalian N-terminal DING peptides described above exhibit similarity to a range of microbial proteins annotated as periplasmic phosphate-binding proteins or alkaline phosphatases, which lack the specific DINGGG-motif and show levels of identity to the DING peptide in the range of 30–50%. The *Pseudomonas* and *Hypericum* sequences, however, possess the DINGGG-motif and have levels of identity to DING greater than 85% (similar to that seen between the different animal DING proteins).

The internal peptide fragments from the human [3,6] and turkey [8] DING proteins that previously had shown no similarity to other known protein sequences were compared to the full-length *P. aeruginosa* DING sequence. It was found that the majority of these peptides could be reliably aligned to regions of the *P. aeruginosa* sequence, confirming this full-length sequence as a DING protein (Fig. 1).

Sequence database searching using the *P. aeruginosa* DING protein identified a human EST from the TIGR Human Cancer Gene database (TIGR Accession No. TC105368), two overlapping unfinished unassigned *Leishmania major* Friedlin contigs (Sanger Accession Nos. LM23_BIN_Contig2072 and LM16_BIN_Contig2206; merged and annotated as LM23_16 for the remainder of this paper), and a series of partial protein sequences from a variety of plants (Fig. 1); potato (*Solanum tuberosum*; GenBank Accession No. AY741549 and GenBank Accession No. AY227747), mouse-ear cress (*Arabidopsis thaliana*; GenBank Accession No. AY224598) and common tobacco (*Nicotiana tobacum*; GenBank Accession No. AY227749). Additionally, a rat unfinished genomic sequence (GenBank Accession No. AC135282.3) was also identified (Fig. 1) showing high homology (greater than 80% at the DNA level in particular regions) to the *P. aeruginosa* sequence. This 225 kb working draft sequence consisted of 2 unordered pieces, a large 223 kb piece and a small 1.5 kb piece, this smaller piece containing the region matching the DING protein. On further analysis, this sequence appeared to represent a potential DING pseudogene, exhibiting both a variety of unusual indels (compared to the DING and DING-like sequences) and a frameshift. A DING