

# *Eimeria tenella*: Identification of secretory and surface proteins from expressed sequence tags

Christian Klotz<sup>a</sup>, Richard J. Marhöfer<sup>b</sup>, Paul M. Selzer<sup>b</sup>,  
Richard Lucius<sup>a</sup>, Thomas Pogonka<sup>a,\*</sup>

<sup>a</sup> Department of Molecular Parasitology, Humboldt University, Berlin, Germany

<sup>b</sup> Intervet Innovation GmbH, Drug Discovery/BioChemInformatics, Schwabenheim, Germany

Received 8 October 2004; received in revised form 29 March 2005; accepted 14 April 2005

Available online 3 June 2005

## Abstract

To identify new vaccine candidates, *Eimeria tenella* expressed sequence tags (ESTs) from public databases were analysed for secretory molecules with an especially developed automated in silico strategy termed DNAsignalP. A total of 12,187 ESTs were clustered into 2881 contigs followed by a blastx search, which resulted in a significant number of *E. tenella* contigs with homologies to entries in public databases. Amino acid sequences of appropriate homologous proteins were analysed for the occurrence of an N-terminal signal sequence using the algorithm signalP. The resulting list of 84 entries comprised 51 contigs whose deduced proteins showed homologies to proteins of apicomplexan parasites. Based on function or localisation, we selected candidate proteins classified as (i) secreted proteins of Apicomplexa parasites, (ii) secreted enzymes, and (iii) transport and signalling proteins. To verify our strategy experimentally, we used a functional complementation system in yeast. For five selected candidate proteins we found that these were indeed secreted. Our approach thus represents an efficient method to identify secretory and surface proteins out of EST databases.

© 2005 Elsevier Inc. All rights reserved.

**Index descriptors and abbreviations:** *Eimeria tenella*; Apicomplexa; Secretory proteins; AMA1, apical membrane protein 1; AN, accession number; bp, base pair; CN, contig number; ER, endoplasmic reticulum; EST, expressed sequence tag; Lag1, longevity assurance homolog 1; MIC, micronema protein; PAP, phosphatidic acid phosphatase; RT-PCR, reverse transcribed-polymerase chain reaction; SPATR, secreted protein with altered thrombospondin repeat; SRS, sequence retrieval system; TRAP, thrombospondin-related adhesive protein; TSR, thrombospondin type 1 repeat; TTC, 2,3,5-triphenylterazolium chloride

**Keywords:** Apicomplexa; Coccidia; *Eimeria tenella*; Chicken; Host–parasite interaction; Secretory proteins; Surface proteins; Expressed sequence tags; In silico analysis

## 1. Introduction

Protozoa of the genus *Eimeria* are intracellular parasites in the intestine of many domestic animals causing coccidiosis, a severe infection particularly of poultry (Fernando, 1990). These infections are responsible for

high economical losses (Williams, 1999) and have to be controlled by permanent medication with anti-coccidia. In addition, live vaccines have become more important in the last years (Chapman et al., 2002; Jenkins, 1998). The considerable increase of drug resistance (Chapman, 1993) and the high costs of live vaccines have made it necessary to develop additional control options like subunit vaccines. One starting point for the development of such vaccines, and also for the identification of new drug targets, is the investigation of mole-

\* Corresponding author. Fax: +49 30 20936051

E-mail address: [thomas.pogonka@rz.hu-berlin.de](mailto:thomas.pogonka@rz.hu-berlin.de) (T. Pogonka).

cules that are part of the host–parasite interface. Such molecules interact during invasion and intracellular development of the parasites with host cell components (Menard, 2001). From rational considerations the expected molecules should be secreted or localised on the surface of the parasite.

Analysis of genomic and expressed sequence tag (EST) databases provides a key approach to identify new target molecules for the development of drugs and vaccines (Li et al., 2003). As a consequence, the aim of our study was the identification of secretory or surface molecules from *Eimeria tenella* ESTs. As ESTs are often afflicted with errors and in particular incomplete at the 5' end, we used an integrated in silico strategy termed DNAsignalP to identify secretory proteins. This is an automated in silico succession of an EST clustering process (Selzer et al., 2004) followed by a blastx search against a protein database (Gish and States, 1993) and finally an analysis of the top database hit with the algorithm signalP for the presence of an N-terminal signal peptide (Nielsen et al., 1997). In a final conclusion based on analogy, the predicted presence of the signal peptide in the top database hit can be assumed to exist also in the homologous query sequence of the parasite. Our procedure led to a selected number of proteins, which have a homologous partner protein in databases with a clearly predicted signal peptide and yielded to a manageable number of candidates for further studies. Finally, the predicted functional property of secretion was experimentally verified for five selected *Eimeria* cDNAs in yeast cells.

## 2. Materials and methods

### 2.1. Sequence retrieval and in silico signal peptide analysis

*Eimeria tenella* ESTs were identified and downloaded using the Sequence Retrieval System (SRS) at the European Bioinformatic Institute (<http://www.ebi.ac.uk>) applying a query with the organism keyword “*Eimeria tenella*” in the EMBL database (Kulikova et al., 2004). A novel in silico strategy termed DNAsignalP was used for the inspection of the ESTs regarding the potential existence of signal peptides on the respective gene product. ESTs were assembled with the clustering software “Lasergene–Seqman” (DNASTAR, Madison WI, USA) to build sequence contigs (Selzer et al., 2004). The consensus sequences generated out of the contigs were blasted versus the Swissprot and TrEMBL databases (Boeckmann et al., 2003) using the program blastx (Gish and States, 1993). Subsequently, the sequence information for the best matching protein (hit protein) was extracted from the database and used as input for an in silico signal peptide analysis using the signalP algorithm (Nielsen et al., 1997). In case the

existence of a signal peptide is predicted for the hit protein and the significance of the blastx match is above a certain threshold, the existence of a signal peptide can be assumed by conclusion on analogy also for the respective protein of the contig under investigation.

Due to the sheer number of contigs, a manual approach for the signal peptide analysis was not feasible, the workflow described above was integrated as automated workflow in the *LION bioSCOUT* mass analysis tool *DO\_ALL* (LION bioscience AG, Heidelberg, Germany). In addition, all contigs were automatically categorised regarding their potential function using the functional annotation procedure of the bioSCOUT mass analysis tool *DO\_ALL*. Using this customised automated workflow it was possible to analyse the 2881 contigs regarding the presence of signal peptides. For a convenient assessment, the resulting data were transferred to the SRS based *featureBASE* result database system (LION bioscience AG, Heidelberg, Germany). Contigs that represented gene products probably carrying a signal peptide were identified using an SRS query for clear signalP predictions on the hit protein. Because the respective conclusion on analogy is reasonable only if the significance of the according blastx match is above a certain threshold, an *E*-value filter of  $E \leq 10^{-4}$  was applied to the resulting data. The *E*-value (Eq. (1)) is calculated from the length *m* of the query sequence, the overall length *n* of all sequences in the queried database and the bit-score *S'* of the current alignment (Altschul et al., 1997).

$$E = mn2^{-S'}, \quad (1)$$

where *E* is the expectation value, *m* is the length of query sequence, *n* is the overall length of all sequences in the database, and *S'* is the bit-score.

The bit-score *S'* (Eq. (2)) is the alignment score *S* normalised using the substitution matrix dependent Karlin–Altschul parameters  $\lambda$  and *K* (Karlin and Altschul, 1990)

$$S' = \frac{\lambda S - \ln K}{\ln 2}, \quad (2)$$

where *S'* is the bit-score, *S* is the score, and  $\lambda$ , *K* are the Karlin–Altschul parameters.

An *E*-value of 1 assigned to a hit therefore means, that it is expected to find one sequence with a similar score in a database of the current size just by chance. Accordingly, an *E*-value of zero assigned to a hit means, that it is expected to find no sequence with a similar score in a database of the current size by chance. Therefore, the closer the *E*-value is to zero, the more significant the match is. However, because the *E*-value is dependent on the length *m* of the query sequence and the overall length *n* of all sequences of the queried database, database queries with short sequences may result in virtually identical matches with relatively high *E*-values.

Download English Version:

<https://daneshyari.com/en/article/9442923>

Download Persian Version:

<https://daneshyari.com/article/9442923>

[Daneshyari.com](https://daneshyari.com)