

# Post-genomics of microsporidia, with emphasis on a model of minimal eukaryotic proteome: a review

Catherine Texier, Damien Brosson, Hicham El Alaoui, Guy Méténier and Christian P. Vivarès

Equipe de Parasitologie Moléculaire et Cellulaire, Laboratoire de Biologie des Protistes, UMR CNRS 6023, Université Blaise Pascal (Clermont 2), Complexe Universitaire des Cézeaux, 63177–Aubière Cedex, France

Key words: Microsporidia, *Encephalitozoon*, intracellular parasite, post-genomics, protein size, deletions, proteome analysis

**Abstract.** The genome sequence of the microsporidian parasite *Encephalitozoon cuniculi* Levaditi, Nicolau et Schoen, 1923 contains about 2,000 genes that are representative of a non-redundant potential proteome composed of 1,909 protein chains. The purpose of this review is to relate some advances in the characterisation of this proteome through bioinformatics and experimental approaches. The reduced diversity of the set of *E. cuniculi* proteins is perceptible in all the compilations of predicted domains, orthologs, families and superfamilies, available in several public databases. The phyletic patterns of orthologs for seven eukaryotic organisms support an extensive gene loss in the fungal clade, with additional deletions in *E. cuniculi*. Most microsporidian orthologs are the smallest ones among eukaryotes, justifying an interest in the use of these compacted proteins to better discriminate between essential and non-essential regions. The three components of the *E. cuniculi* mRNA capping apparatus have been especially well characterized and the three-dimensional structure of the cap methyltransferase has been elucidated following the crystallisation of the microsporidian enzyme Ecm1. So far, our mass spectrometry-based analyses of the *E. cuniculi* spore proteome has led to the identification of about 170 proteins, one-quarter of these having no clearly predicted function. Immunocytochemical studies are in progress to determine the subcellular localisation of microsporidia-specific proteins. Post-translational modifications such as phosphorylation and glycosylation are expected to be soon explored.

## INTRODUCTION

Sequencing the small genome (2.9 Mbp) of *Encephalitozoon cuniculi* Levaditi, Nicolau et Schoen, 1923, a monokaryotic microsporidian parasite infecting mammals, has revealed a complement of ~2,000 protein-coding genes that are densely distributed over 11 small chromosomes (Katinka et al. 2001). Some metabolic potentialities deduced from gene annotation have been recently reviewed, mainly as regard to energy metabolism, lipid biosynthesis, protein trafficking (Vivarès and Méténier 2004) and a Fe-S cluster synthesis pathway that is thought to take place within a mitochondrion-derived compartment or “mitosome” (Méténier and Vivarès 2004, Van der Giezen and Tovar 2004).

Although sharing a common intracellular parasitic status and a basic invasion strategy involving spore polar tube extrusion, microsporidia are highly diversified species. Undoubtedly, comparative genomic studies are of great interest to examine this diversity at the gene level. Genome sequence surveys from *Vittaforma corneae* (Mittleider et al. 2002) infecting humans and *Antonosporea* (formerly *Nosema*) *locustae* (Slamovits et al. 2004) infecting grasshoppers, have yet revealed some important differences in gene content compared to *E. cuniculi* (see review by Keeling et al. 2005). A striking example of divergence is provided by the recent demonstration that *A. locustae* spores express a class II photo-

lyase, an enzyme involved in the repair of UV-induced DNA damage (Slamovits and Keeling 2004). Indeed, no photolyase homologue is encoded in *E. cuniculi* genome and no class II photolyase has been found in fungi, the organisms viewed as the closest relatives of microsporidia. The “*Nosema locustae* Genome Project” from the Marine Biological Laboratory at Woods Hole (funded by NSF award number 0135272) has released numerous contigs at the site <http://jbpc.mbl.edu/Nosema/Nannotation.html>. New interesting data are expected from large-scale analyses of this genome.

In the three last years, bioinformatic and/or experimental studies have used *E. cuniculi* gene sequences with different objectives such as the identification of additional orthologous proteins, the evaluation of gene losses during evolution and the exploration of the function of some key-enzymes (Table 1). Regrettably, very little is still known about the characterisation of microsporidia-specific proteins that may play crucial roles in parasitic processes of development and host invasion.

Here, we will review major features of the proteome of *E. cuniculi* and some published experimental data relevant to protein functionality. Mass spectrometry studies aimed to identify proteins expressed during spore differentiation, are in progress in our laboratory and will be finally evoked.

## THE REDUCED COMPLEXITY OF THE PREDICTED PROTEOME OF *ENCEPHALITOZOON* *CUNICULI*

Statistical analyses for completed genomes and predicted proteomes of several prokaryotic and eukaryotic organisms including *E. cuniculi*, are available in the EBI Integr8 web site (<http://www.ebi.ac.uk/integr8/EBI-Integr8-HomePage.do>). The *E. cuniculi* non-redundant proteome is represented by only 1,909 potential “proteins” (polypeptide chains, more exactly), which is much less than in the free-living yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (6,224 and 4,929 proteins, respectively) or in the pathogenic filamentous fungus *Ashbya* (or *Eremothecium*) *gossypii* with a 9.2-Mbp genome (4,713 proteins). The number of proteins with InterPro matches, i.e. with signatures characteristic for known families, domains or functional sites, attains only 1,294 in *E. cuniculi*, contrasting with 3,713 in *A. gossypii* and 3,366 in the apicomplexan parasite *Plasmodium falciparum* 3D7. A reduced diversity of the potential microsporidian proteins sharing common features with known proteins is also evident through their distribution at the superfamily level (Madera et al. 2004, <http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY>). Only 364 superfamilies with an assignment are distinguished in *E. cuniculi*. This is far from the numbers in two *Plasmodium* species (566 in *P. falciparum* and 567 in *P. yoelii*) and in various fungi (from 620 in *Candida glabrata* to 774 in *Fusarium graminearum*).

The updated version of the NCBI COG database contains a compilation of 4,852 clusters of orthologs derived from seven sequenced eukaryotic genomes (*Homo sapiens*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *S. cerevisiae*, *S. pombe* and *E. cuniculi*) and named “eukaryotic orthologous groups” or KOGs (Tatusov et al. 2003, <http://www.ncbi.nlm.nih.gov/COG/new/shokog.cgi>). About 60% of *E. cuniculi* proteins are assigned to KOGs. A massive gene loss in the fungal clade, with additional elimination in *E. cuniculi*, has been inferred from the reconstruction of parsimonious evolutionary scenarios using either a subset of KOGs (Tatusov et al. 2003) or the complete collection (Koonin et al. 2004). The common ancestor of *E. cuniculi* and two yeast species is predicted to contain 3,048 KOGs, and the branch leading to the microsporidian would be characterized by 586 gene gains and up to 1,969 gene losses! This dramatic elimination concerns 269 genes of unknown function, 241 genes with only a general functional prediction and various genes assigned to all major functional categories: 167 for post-translational modification, protein turnover and chaperones, 122 for translation, 116 for intracellular trafficking and secretion, 110 for energy production and conversion, among others.

Strikingly, the *E. cuniculi* proteome shows a tendency to the shortening of individual protein sequences, in relation with a high frequency of intragenic deletions.

The average protein length is close to 360 amino acid (aa) residues and only 6 sequences encompass 2,000 aa (Katinka et al. 2001). Numerous proteins are frequently smaller than most eukaryotic counterparts. Typical examples of compacted proteins are offered by two conserved members of the AAA ATPase family: cytoplasmic dynein heavy chain (Vivarès et al. 2002) and midasin (Garbarino and Gibbons 2002). The latter is a nuclear protein encoded by the largest open reading frame in *S. cerevisiae*. Yeast midasin (4,910 aa; ~560 kDa) is smaller than human ortholog (5,596 aa; ~632 kDa) but *E. cuniculi* midasin has been prone to a severe reduction (2,832 aa; ~324 kDa). As illustrated in Fig. 1A, so far the microsporidian protein is the smallest, compared to those from animals, fungi or protists. Analysis by multiple sequence alignment has revealed the reduction of all five domains at different extents (Garbarino and Gibbons 2002). Taking *S. cerevisiae* midasin as a reference, the authors have provided the following degrees of reduction: 90% for the N-terminal domain, 19% for the AAA domain retaining all six tandem AAA protomers, 50% for the lengthy linker domain as for the acidic D/E-rich domain, and 20% for the C-terminal M-domain retaining all three MIDAS motifs. This is consistent with the critical roles of AAA- and M-domains, and indicates the non-essential character of large parts of the three other domains. We have examined the case of ribonucleotide reductase (RNR), a ubiquitous enzyme required to provide the 2'-deoxyribonucleotide precursors for both DNA replication and DNA repair. Like in all other eukaryotes, the *E. cuniculi* enzyme can be assigned to the oxygen-dependent class Ia and should exist as a tetramer  $\alpha_2\beta_2$ . The large ( $\alpha$ ) and small ( $\beta$ ) subunits contain 768 aa and 325 aa, respectively. These correspond to the lowest RNR lengths among eukaryotic organisms whereas most fungi display the highest values (Fig. 1B). The large subunit of *E. cuniculi* RNR has a reduced C-terminal domain, justifying a length close to R1 chain in *Escherichia coli* (761 aa) or in several eukaryote viruses. Conversely, the small subunit has a reduced N-terminal domain and is closer in size to *E. coli* R2F chain (319 aa).

After assignment of conserved domains from the NCBI CDD database to orthologous sets of eukaryotic proteins, Koonin et al. (2004) have considered that “...among the unicellular eukaryotes, *Encephalitozoon* had by far the least complex domain architectures, which reflects the general reduction in this intracellular parasite”. The number of orthologous groups in which the average number of domains is greater in *E. cuniculi* ranges only from 10 (vs *Homo sapiens*) to 19 (vs *S. pombe*). In contrast, the number of groups in which the domain number is lower in *E. cuniculi* varies from 332 (vs *H. sapiens*) to 140 (vs *S. pombe*). Thus, the *E. cuniculi* proteome appears as a package of compact proteins containing a significant proportion of orthologs with simplified domain organisation.

**Table 1.** A list of studies on the characterisation of *Encephalitozoon cuniculi* gene sequences or products, in a three-year period (2002–2004).

<i>E. cuniculi</i> proteins or RNAs	Main findings	References
Kinesin-like proteins (KLPs 2, 3, 4, 5 and 6)	– Expression and localisation of five <i>E. cuniculi</i> GFP-KLPs in the protist <i>Leishmania major</i>	Dubessay et al. 2004*
Protein kinase A, catalytic subunit (ECU07_0520; EcPKAc)	– Recombinant EcPKAc is inactive in phosphorylating a usual synthetic substrate of PKAs. Two gene copies in <i>E. intestinalis</i>	Equinet et al. 2004*
RNA polymerase II subunit A C-terminal domain phosphatase (ECU07_0890; EcFcp1)	– EcFcp1 is a monomeric enzyme that has CTD phosphatase activity <i>in vitro</i> and can replace yeast Fcp1 <i>in vivo</i>	Hausmann et al. 2004*
Orthologous protein clusters (see COG database)	– Scenarios of loss and emergence of genes during eukaryotic evolution, based on comparative analyses of KOGs from seven eukaryotes including <i>E. cuniculi</i>	Krylov et al. 2003 Tatusov et al. 2003 Koonin et al. 2004
Signal recognition particle RNA	– An SRP RNA sequence is present on chromosome X	Rosenblad et al. 2004
ATP/ADP translocase (4 genes)	– All ATP/ADP translocases would be of deep origin	Amiri et al. 2003
Methionine aminopeptidase-2 (ECU10_0750, EcMetAP2)	– Homology modelling; potential binding sites of EcMetAP2 for Co <sup>2+</sup> , fumagillin and TNP-470 – TNP-470 inhibits the activity of EcMetAP2 <i>in vitro</i>	Bontems et al. 2003* Weiss et al. 2003*
Homeobox (12 genes)	– Clustering of two homeobox genes is suggestive of a mating-type locus (ECU10_1470, EcHd1; ECU10_1480, EcHd2)	Burglin 2003*
Midasin (ECU08_1900)	– Large AAA ATPase characterized by a strong size reduction in <i>E. cuniculi</i>	Garbarino and Gibbons 2002
20S proteasome (14 genes)	– Seven $\beta$ -type subunits are assigned to $\beta$ 1– $\beta$ 7 subtypes but $\alpha$ subunits are highly divergent	Gille et al. 2003
tRNAs (46 genes)	– Comparative study of 60 genomes to detect tRNA paralogs	Xue et al. 2003
Pot1-like protein (ECU05_1050)	– Similar to a single-stranded telomeric DNA-binding protein	Baumann et al. 2002
ABC transporters (13 genes)	– Only four families of ABC proteins are distinguished	Cornillot et al. 2002*
mRNA 5'-triphosphatase (ECU04_1550, EcCet1)	– EcCet1 is shown to be an homodimeric metal-dependent RNA triphosphatase and lacks a non-essential N-terminal extension	Hausmann et al. 2002*
mRNA guanylyltransferase (ECU09_0400, EcCeg1)	– EcCeg1 exhibits guanylyltransferase activity <i>in vitro</i> and can sustain the growth of yeast <i>ceg1<math>\Delta</math></i> mutants	Hausmann et al. 2002*
mRNA cap (guanine-N7) methyltransferase (ECU10_0380, Ecm1)	– Ecm1 can sustain the growth of yeast <i>abd1<math>\Delta</math></i> mutants – Resolution of crystal structures of this enzyme in complex with S-adenosylmethionine, S-adenosylhomocysteine and the cap guanylate	Hausmann et al. 2002* Fabrega et al. 2004*
ORMDL (ECU11_1150)	– Assignment to a novel family of ER membrane proteins	Hjelmqvist et al. 2002
NARF-like protein (ECU05_0970)	– Similar to nuclear prelamin A recognition factor and Fe-hydrogenases	Horner et al. 2002
Ribosomal proteins	– Comparative analysis in sequenced genomes; <i>E. cuniculi</i> lacks S21e, L14e, L28e and L38e homologs	Lecompte et al. 2002
Polar tube protein 3 precursor (ECU11_1440, PTP3)	– Identification and characterisation of a microsporidia-specific protein associated with the polar tube	Peuvel et al. 2002*
Ferredoxin (ECU07_0600)	– Assignment to type-II mitochondrial ferredoxin	Seeber 2002

\*Stars indicate reports dealing with only microsporidial genes.

Most gene gains presumably arose from duplication-divergence events. However, a small set of KOGs represented in *E. cuniculi* but absent in the two yeast species, has led to propose two explanations: (1) loss of ancestral archaeo-eukaryotic genes in yeasts, for the orthologs that are also found in all archaea but not in bacteria, and (2) acquisition of genes by *E. cuniculi* through horizontal gene transfer, for those showing high similarities to some bacterial orthologs, e.g. 3-methyl-adenine DNA glycosylase and ATP/ADP translocase (Tatusov et al.

2003, Koonin et al. 2004). Horizontal gene transfer has been also invoked to justify the presence of class II catalase and photolyase in *A. locustae* (Fast et al. 2003, Slamovits and Keeling 2004). As non-mitochondrial ATP/ADP translocases (ATP-importing) are distributed in plant chloroplasts, microsporidia and intracellular parasitic bacteria (*Rickettsia*, Chlamydiaceae), a relatively recent acquisition by microsporidia and plants via independent horizontal gene transfers from bacteria has been hypothesized (Koonin et al. 2004, Méténier and

Vivarès 2004). An alternate hypothesis has been proposed in a phylogenetic study supporting a very ancient origin for all mitochondrial and non-mitochondrial (“plastid/parasite”) ATP/ADP translocases (Amiri et al. 2003). The “plastid/parasite” gene would have been present in a bacterial ancestor of mitochondria and *Rickettsia*, and then transferred into the nuclear genome of the early eukaryotic cell. This gene would have been subsequently lost in various eukaryotic lineages, except in plants for cytosol-to-plastid ATP transport and in microsporidia for host ATP import. Whatever the evolutionary scenario, discovering the membrane location and substrate specificity of the four ATP/ADP translocase paralogs in *E. cuniculi* will be of interest for the knowledge of host-microsporidia energetic relationships.

### SOME INSIGHTS INTO THE FUNCTIONALITY OF *ENCEPHALITOZOON CUNICULI* PROTEINS

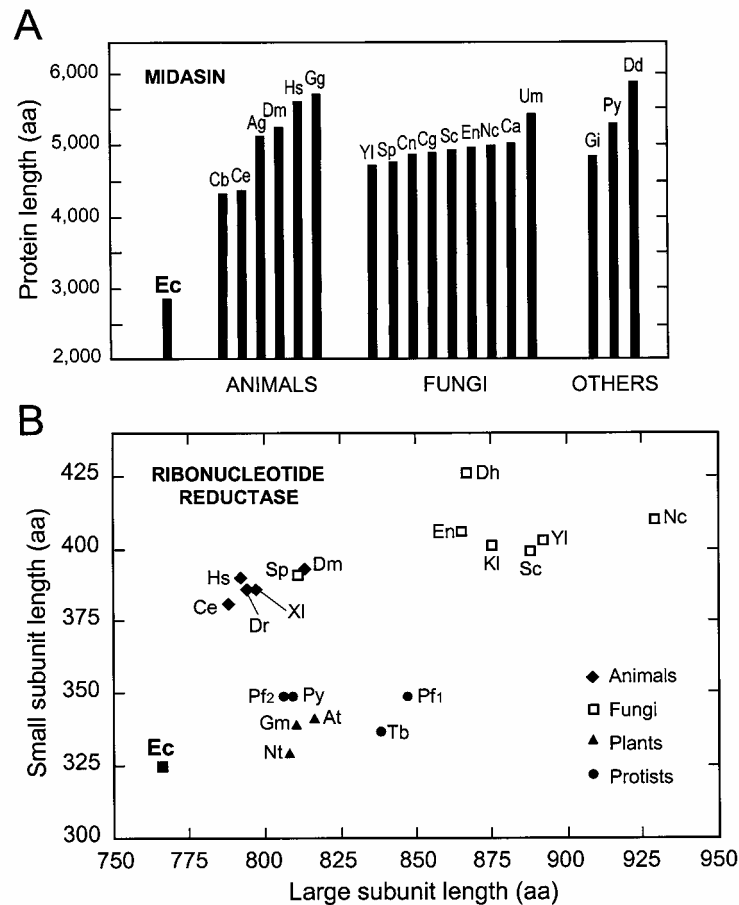
When examining the list of reports in Table 1, it is clear that the deepest post-genomic experimental investigations of *E. cuniculi* proteins deal with the biochemical and genetic dissection of the three enzymes responsible for mRNA capping: mRNA 5'-triphosphatase (EcCet1), RNA guanylyltransferase (EcCeg1) and RNA (guanine-N7) methyltransferase (Ecm1). One reason of the interest in the use of *E. cuniculi* sequences was to determine whether the triphosphatase component was of cysteine-phosphatase or metal-dependent type. Another reason was to test the functionality of three typically compacted microsporidian enzymes using yeast expression plasmids. Lacking the large N-terminal extensions observed in fungal homologs, EcCet1 contains only 221 aa. The purified recombinant enzyme can hydrolyze ATP in the presence of  $Mn^{2+}$  or  $Co^{2+}$  ions, and the replacement of conserved glutamate residues required for metal coordination with alanine residues abolishes its activity, revealing a metal-dependent triphosphatase like in fungi, DNA viruses and *Plasmodium* (Hausmann et al. 2002). Intermediate in size between *Chlorella* virus and *S. pombe* guanylyltransferases, EcCeg1 (364 aa) has been demonstrated to be specific for GTP and active in cap formation *in vivo* in *S. cerevisiae* cells deleted for the endogenous enzyme. Likewise, Ecm1 without N-terminal extensions and reduced to 298 aa (*vs* 476 aa in *H. sapiens* and 402 aa in *S. pombe*), has been verified to be active by complementation of yeast mutants without endogenous cap methyltransferase (Hausmann et al. 2002). In summary, the components of the *E. cuniculi* mRNA capping apparatus provide a nice illustration of eukaryotic enzymes having preserved just a little more than the minimum catalytic domains. The X-ray structure of guanine-N7-methyl transferase has been successfully determined for the first time after co-crystallisation of Ecm1 with three different ligands, thus completing the suite of protein structures for mRNA capping enzymes (Fabrega et al. 2004). This confirms that the use

of minimal *E. cuniculi* orthologs can help to elucidate the three-dimensional (3-D) structure of important enzymes and to identify critical regions for ligand binding.

Because of its somewhat surprising results, we will also mention a recent study on the expression of green fluorescent protein (GFP)-fused versions of *E. cuniculi* genes in an original heterologous system represented by the parasitic flagellate *Leishmania major* (Dubessay et al. 2004). The selected genes were those coding for a family of six kinesin-like proteins (EcKLPs) and five of these were successfully expressed in *Leishmania*. Three EcKLPs exhibited a diffuse cytoplasmic localisation. The surprise arose from the very curious but precise localisation of the two other GFP-fused proteins. EcKLP3 was confined at the posterior end of the *Leishmania* cell, a region characterized by the termination of subpellicular microtubules, while EcKLP6 was targeted to the tip of the flagellum. Even if these two opposite locations cannot be extrapolated to the *Encephalitozoon* cell, it seems likely that the different EcKLPs play specific roles in intracellular transport processes required for microsporidian life. The comparison with data of immunolocalisation in *E. cuniculi* cells is expected to be informative.

A protein that has drawn the attention of workers in the field of the chemotherapy of microsporidiosis is methionine aminopeptidase type 2 (MetAP2). Fumagillin or its synthetic derivative TNP-470 has been shown to bind covalently to mammalian MetAP2 (Griffith et al. 1998, Liu et al. 1998). These compounds are efficient antimicrosporidian agents and it is conceivable that microsporidian MetAP2 also represents fumagillin/TNP-470 target. An explanation is offered by the lack of MetAP1 gene in the *E. cuniculi* genome (Katinka et al. 2001). The effect of a drug inhibiting the unique MetAP2 might be sufficient to strongly alter the viability of microsporidia. Data from two different approaches support specific interactions between *E. cuniculi* MetAP2 (EcMetAP2) and fumagillin/TNP-470: (1) *in silico* modelling of EcMetAP2 3-D structure indicates the conservation of critical residues for these interactions (Bontems et al. 2003), and (2) a purified recombinant EcMetAP2 is strongly inhibited by TNP-470 (Weiss et al. 2003). Note, however, that depletion of MetAP2 by small interfering RNA failed to inhibit the proliferation of endothelial cells and to modify their sensitivity to fumagillin (Kim et al. 2004). It cannot be excluded that other fumagillin targets exist in such cells, and perhaps also in microsporidia.

Investigating microsporidia-specific proteins is a major challenge. Two polar tube proteins (PTP1, PTP2), extracted in the presence of a thiol-reducing compound, are encoded by two neighbouring genes in *Encephalitozoon* species (Delbac et al. 1998, 2001, Keohane et al. 1998). A large-sized protein (PTP3; 1,256 aa) has been subsequently demonstrated to be associated with the polar tube of *E. cuniculi* (Peuvel et al. 2002). Chemical



**Fig. 1.** Lengths of midasin (**A**) and of ribonucleotide reductase subunits (**B**) in *Encephalitozoon cuniculi* (Ec) and other eukaryotic organisms. Species abbreviations are: Ag – *Anopheles gambiae*; At – *Arabidopsis thaliana*; Ca – *Candida albicans*; Cb – *Caenorhabditis briggsae*; Ce – *Caenorhabditis elegans*; Cg – *Candida glabrata*; Cn – *Cryptococcus neoformans*; Dd – *Dictyostelium discoideum*; Dh – *Debaryomyces hansenii*; Dm – *Drosophila melanogaster*; Dr – *Danio rerio*; En – *Emericella nidulans*; Gg – *Gallus gallus*; Gi – *Giardia intestinalis*; Gm – *Glycine max*; Hs – *Homo sapiens*; Kl – *Kluyveromyces lactis*; Nc – *Neurospora crassa*; Nt – *Nicotiana tabacum*; Pf<sub>1</sub> – *Plasmodium falciparum* 3D7; Pf<sub>2</sub> – *Plasmodium falciparum* Dd2; Py – *Plasmodium yoelii*; Sc – *Saccharomyces cerevisiae*; Sp – *Schizosaccharomyces pombe*; Tb – *Trypanosoma brucei*; Um – *Ustilago maydis*; Xl – *Xenopus laevis*; Yl – *Yarrowia lipolytica*.

cross-linking experiments support the involvement of the three proteins in the formation of a large multimeric complex. In contrast with the two other PTPs, PTP3 has no cysteine residues, indicating that this protein cannot interact via disulfide linkages. The most abundant PTP is PTP1 (~70% of the mass of the polar tube) and evidence for O-mannosylation of this protein has been recently provided (Xu et al. 2004). Sequence data about spore wall proteins are restricted to one *E. cuniculi* protein (SWP1, Bohne et al. 2000) and two related *Encephalitozoon intestinalis* proteins having divergent C-terminal domains (SWP1 and SWP2, Hayman et al. 2001), all localised to the outer wall layer (exospore). Two *E. cuniculi* proteins associated with the inner wall layer (endospore) are currently investigated in our laboratory (Delbac F., unpublished data).

It should be stressed that a significant conservation of gene order exists between two distantly related micro-

sporidians such as *E. cuniculi* and *Antonospora locustae* (Slamovits et al. 2004). The compilation of the groups of neighbouring genes common to the two species should provide a good basis to further investigate undefined proteins that are encoded by homologous genes within similar chromosomal contexts, shared by at least two different genera of the microsporidian phylum.

#### FIRST MASS SPECTROMETRY-BASED PROTEOMICS APPROACHES AND PERSPECTIVES

An important area of proteomics research implies the use of mass spectrometry (MS) as a powerful tool to identify small amounts of proteins from complex mixtures, frequently after separation by 2-D gel electrophoresis procedures. Even in the case of species cultured *in vitro*, there are very few reports of 2-D patterns of microsporidian proteins, mainly because of difficulties in the development of reproducible protocols for the ex-

traction of proteins from spores equipped with a resistant wall and for the fractionation of intracellular stages including wall-lacking meronts. However, an interesting approach has been realized with *E. intestinalis* maintained in monkey kidney E6 cells, consisting in the comparison of 2-D protein profiles of four kinds of samples: purified *E. intestinalis* spores, E6 cells with estimated infection levels of either ~50% or ~90%, and control uninfected E6 cells (Moura and Visvesvara 2001). Of more than 200 protein spots resolved in mid-sized gels, 95 were found to be immunogenic on the basis of their reactivity with a rabbit anti-*E. intestinalis* serum, some of these spots providing potential markers for either merogonial or sporogonial stages. In another study, whole spores and soluble fractions of three *Encephalitozoon* species and *Brachiola algerae* were directly analyzed by matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) MS, providing species-specific spectral markers within the molecular size range of 2–8 kDa (Moura et al. 2003).

Within the context of the exploitation of *E. cuniculi* genome sequence and with the help of a laboratory specialised in protein identification by automated MS techniques, we have undertaken a first large-scale study of the proteome representative of the latest *E. cuniculi* sporogonial stages (immature and mature spores) (Texier et al. 2004). Approximately 350 Coomassie blue-stained spots, in amounts accessible to MALDI-TOF and MS/MS analyses, were resolved with molecular masses between 10 and 250 kDa and within the pI range of 3–10, after 2-D electrophoresis in large-sized gels. Independently, whole protein extracts were subjected to trypsin digestion, then liquid chromatography coupled to MS/MS identification, revealing especially proteins with multiple transmembrane domains. More than 170 different proteins have been assigned, at the present time. Their distribution among large functional categories is consistent with some major requirements for the protection against environmental stresses. Several components of the translation machinery were also identified. Whether this reflects a late protein synthesis during sporogony or a pool of essential factors that will be used after sporoplasm discharge for initiating merogony, or both, is an open question.

About 25% of identified proteins have no clearly predicted function, suggesting that a large proportion of *E. cuniculi* genes encoding the so-called “hypothetical proteins” can be expressed. Some of these genes were cloned and corresponding recombinant proteins were used for the production of polyclonal antibodies in order to determine the localisation of the antigenic proteins at the ultrastructural level. Currently, we have obtained evidence for four proteins targeted to the spore envelope. One of these proteins has a predicted domain characteristic for the polysaccharide deacetylase family and was found to be localised at the endospore/plasma membrane interface (Brosson D., Kuhn L., Prensier G., Vivarès C.P. and Texier C., manuscript in preparation). This is reminiscent of the localisation of chitin deacetylase near the inner face of the cell wall (periplasmic space) in the mycelium of the zygomycete *Absidia coerulea* (Gao et al. 1995). Given the chitin richness of the microsporidian endospore, it seems very likely that the *E. cuniculi* enzyme acts as a chitin deacetylase, hydrolyzing the N-acetamido groups of N-acetyl-D-glucosamine residues in chitin chains and producing chitosan. Thus, the construction of a highly rigid cell wall during sporogony presumably depends on the combined functioning of chitin synthase and chitin deacetylase.

A usual characteristic of protozoan genomes is the placement of gene families encoding surface proteins near the ends of multiple chromosomes. Putative gene families associated with subtelomeric regions of *E. cuniculi* chromosomes are currently tested for a possible expression. In a next future, we expect to evaluate the subsets of microsporidian proteins subjected to major post-translational modifications such as phosphorylation and glycosylation (“phosphoproteome” and “glycoproteome”), in complement of specific studies on proteins required for cell cycle progression or for spore wall and polar tube organisation. As already stressed in a previous review (Vivarès and Méténier 2004), it is important to test whether *E. cuniculi* is really incapable of N-glycosylation. Thorough biochemical investigations of carbohydrate content and glycosidic linkages are therefore required.

## REFERENCES

- AMIRI H., KARLBERG O., ANDERSSON S.G.E. 2003: Deep origin of plastid/parasite ATP/ADP translocases. *J. Mol. Evol.* 56: 137–150.
- BAUMANN P., PODELL E., CECIL T.R. 2002: Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing. *Mol. Cell. Biol.* 22: 8079–8087.
- BOHNE W., FERGUSON D.J., KOHLER K., GROSS U. 2000: Developmental expression of a tandemly repeated, glycine- and serine-rich spore wall protein in the microsporidian pathogen *Encephalitozoon cuniculi*. *Infect. Immun.* 68: 2268–2275.
- BONTEMPS F., LE FLOCH P., DUFFIEUX F., BIDERRE C., PEYRET P., LALLEMAND J.Y. 2003: Homology modeling and calculation of the cobalt cluster charges of the *Encephalitozoon cuniculi* methionine aminopeptidase, a potential target for drug design. *Biophys. Chem.* 105: 29–43.
- BURGLIN T.R. 2003: The homeobox genes of *Encephalitozoon cuniculi* (Microsporidia) reveal a putative mating-type locus. *Dev. Genes Evol.* 213: 50–52.

- CORNILLOT E., MÉTÉNIER G., VIVARÈS C.P., DASSA E. 2002: Comparative analysis of sequences encoding ABC systems in the genome of the microsporidian *Encephalitozoon cuniculi*. FEMS Microbiol. Lett. 210: 39–47.
- DELBAC F., PEUVEL I., MÉTÉNIER G., PEYRETAILLADE E., VIVARÈS C.P. 2001: Microsporidian invasion apparatus: identification of a novel polar tube protein and evidence for clustering of *ptp1* and *ptp2* genes in three *Encephalitozoon* species. Infect. Immun. 69: 1016–1024.
- DELBAC F., PEYRET P., MÉTÉNIER G., DAVID D., DANCHIN A., VIVARÈS C.P. 1998: On proteins of the microsporidian invasive apparatus: complete sequence of a polar tube protein of *Encephalitozoon cuniculi*. Mol. Microbiol. 29: 825–834.
- DUBESSAY P., PAGÈS M., DELBAC F., BASTIEN P., VIVARÈS C., BLAINEAU C. 2004: Can heterologous gene expression shed (a torch) light on protein function? Trends Biotechnol. 22: 557–559.
- EQUINET L., BAPTESTE E., THELLIER M., OUARZANE-AMARA M., VIVARÈS C.P., DESPORTES-LIVAGE I., DOERIG C. 2004: The genes encoding cAMP-dependent protein kinase catalytic subunit homologues of the microsporidia *Encephalitozoon intestinalis* and *E. cuniculi*: molecular characterisation and phylogenetic analysis. Parasitol. Int. 53: 277–285.
- FABREGA C., HAUSMANN S., SHEN V., SHUMAN S., LIMA C.D. 2004: Structure and mechanism of mRNA cap (guanine-N7) methyltransferase. Mol. Cell 13: 77–89.
- FAST N.M., LAW J.S., WILLIAMS B.A., KEELING P.J. 2003: Bacterial catalase in the microsporidian *Nosema locustae*: implications for microsporidian metabolism and genome evolution. Eukaryot. Cell 2: 1069–1075.
- GAO X.D., KATSUMOTO T., ONODERA K. 1995: Purification and characterization of chitin deacetylase from *Abisidia coerulea*. J. Biochem. (Tokyo) 117: 257–263.
- GARBARINO J.E., GIBBONS I.R. 2002: Expression and genomic analysis of midasin, a novel and highly conserved AAA protein distantly related to dynein. BMC Genomics 3: 18.
- GILLE C., GOEDE A., SCHLOETELBURG C., PREISSNER R., KLOETZEL P.M., GOBEL U.B., FROMMEL C. 2003: A comprehensive view on proteasomal sequences: implications for the evolution of the proteasome. J. Mol. Biol. 326: 1437–1448.
- GRIFFITH E.C., SU Z., NIWAYAMA S., RAMSAY C.A., CHANG Y.H., LIU J.O. 1998: Molecular recognition of angiogenesis inhibitors fumagillin and ovalicin by methionine aminopeptidase 2. Proc. Natl. Acad. Sci. USA 95: 15183–15188.
- HAUSMANN S., SCHWER B., SHUMAN S. 2004: An *Encephalitozoon cuniculi* ortholog of the RNA polymerase II carboxyl-terminal domain (CTD) serine phosphatase Fcp1. Biochemistry 43: 7111–7120.
- HAUSMANN S., VIVARÈS C.P., SHUMAN S. 2002: Characterization of the mRNA capping apparatus of the microsporidian parasite *Encephalitozoon cuniculi*. J. Biol. Chem. 277: 96–103.
- HAYMAN J.R., HAYES S.F., AMON J., NASH T.E. 2001: Developmental expression of two spore wall proteins during maturation of the microsporidian *Encephalitozoon intestinalis*. Infect. Immun. 69: 7057–7066.
- HJELMQVIST L., TUSON M., MARFANY G., HERRERO E., BALCELLS S., GONZALEZ-DUARTE R. 2002: ORMDL proteins are a conserved new family of endoplasmic reticulum membrane proteins. Genome Biol. 3: R27.
- HORNER D.S., HEIL B., HAPPE T., EMBLEY T.M. 2002: Iron hydrogenases – ancient enzymes in modern eukaryotes. Trends Biochem. Sci. 27: 148–153.
- KATINKA M.D., DUPRAT S., CORNILLLOT E., MÉTÉNIER G., THOMARAT F., PRENSIER G., BARBE V., PEYRETAILLADE E., BROTTIER P., WINCKER P., DELBAC F., EL ALAOUÏ H., PEYRET P., SAURIN W., GOUY M., WEISSENBAACH J., VIVARÈS C.P. 2001: Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414: 450–453.
- KEELING P.J., FAST N.M., LAW J.S., WILLIAMS B.A.P., SLAMOVITS C.H. 2005: Comparative genomics of microsporidia. Folia Parasitologica 52: XXX–XXX.
- KEOHANE E.M., ORR G.A., ZHANG H.S., TAKVORIAN P.M., CALI A., TANOWITZ H.B., WITTNER M., WEISS L.M. 1998: The molecular characterization of the major polar tube protein gene from *Encephalitozoon hellem*, a microsporidian parasite of humans. Mol. Biochem. Parasitol. 94: 227–236.
- KIM S., LAMONTAGNE K., SABIO M., SHARMA S., VERSACE R.W., YUSUFF N., PHILLIPS P.E. 2004: Depletion of methionine aminopeptidase 2 does not alter cell response to fumagillin or bengamides. Cancer Res. 64: 2984–2987.
- KOONIN E.V., FEDOROVA N.D., JACKSON J.D., JACOBS A.R., KRYLOV D.M., MAKAROVA K.S., MAZUMDER R., MEKHEDOV S.L., NIKOLSKAYA A.N., RAO B.S., ROGOZIN I.B., SMIRNOV S., SOROKIN A.V., SVERDLOV A.V., VASUDEVAN S., WOLF Y.I., YIN J.J., NATALE D.A. 2004: A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 5: R7.
- KRYLOV D.M., WOLF Y.I., ROGOZIN I.B., KOONIN E.V. 2003: Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res. 13: 2229–2235.
- LECOMPTE O., RIPP R., THIERRY J.-C., MORAS D., POCH O. 2002: Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucleic Acids Res. 30: 5382–5390.
- LIU S., WIDOM J., KEMP C.W., CREWS C.M., CLARDY J. 1998: Structure of human methionine aminopeptidase-2 complexed with fumagillin. Science 282: 1324–1327.
- MADERA M., VOGEL C., KUMMERFELD S.K., CHOTHIA C., GOUGH J. 2004: The SUPERFAMILY database in 2004: additions and improvements. Nucleic Acids Res. 32 Database issue: D235–239.
- MÉTÉNIER G., VIVARÈS C.P. 2004: Genomics of microbial parasites: the microsporidian paradigm. In: R.P. Hirt and D.S. Horner (Eds.), Organelles, Genomes and Eukaryote

- Phylogeny. An Evolutionary Synthesis in the Age of Genomics. CRC Press, Boca Raton, pp. 207–236.
- MITTLEIDER D., GREEN L.C., MANN V.H., MICHAEL S.F., DIDIER E.S., BRINDLEY P.J. 2002: Sequence survey of the genome of the opportunistic microsporidian pathogen, *Vittaforma corneae*. J. Eukaryot. Microbiol. 49: 393–401.
- MOURA H., OSPINA M., WOOLFITT A.R., BARR J.R., VISVESVARA G.S. 2003: Analysis of four human microsporidian isolates by MALDI-TOF mass spectrometry. J. Eukaryot. Microbiol. 50: 156–163.
- MOURA H., VISVESVARA G.S. 2001: A proteome approach to the host-parasite interaction of the microsporidian *Encephalitozoon intestinalis*. J. Eukaryot. Microbiol. 48 (Suppl.): 56S–59S.
- PEUVEL I., PEYRET P., MÉTÉNIER G., VIVARÈS C.P., DELBAC F. 2002: The microsporidian polar tube: evidence for a third polar tube protein (PTP3) in *Encephalitozoon cuniculi*. Mol. Biochem. Parasitol. 122: 69–80.
- ROSENBLAD M.A., ZWIEB C., SAMUELSSON T. 2004: Identification and comparative analysis of components from the signal recognition particle in protozoa and fungi. BMC Genomics 5: Article 5.
- SEEBER F. 2002: Eukaryotic genomes contain a [2Fe-2S] ferredoxin isoform with a conserved C-terminal sequence motif. Trends Biochem. Sci. 27: 545–547.
- SLAMOVITS C.H., FAST N.M., LAW J.S., KEELING P.J. 2004: Genome compaction and stability in microsporidian intracellular parasites. Curr. Biol. 14: 891–896.
- SLAMOVITS C.H., KEELING P.J. 2004: Class II photolyase in a microsporidian intracellular parasite. J. Mol. Biol. 341: 713–721.
- TATUSOV R.L., FEDOROVA N.D., JACKSON J.D., JACOBS A.R., KIRYUTIN B., KOONIN E.V., KRYLOV D.M., MAZUMDER R., MEKHEDOV S.L., NIKOLSKAYA A.N., RAO B.S., SMIRNOV S., SVERDLOV A.V., VASUDEVAN S., WOLF Y.I., YIN J.J., NATALE D.A. 2003: The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4: 41.
- TEXIER C., BROSSON D., KUHN L., GARIN J., VIVARÈS C. 2004: Proteomics of the microsporidian *Encephalitozoon cuniculi*: application to the identification of new spore wall proteins. J. Eukaryot. Microbiol. 51 (Suppl.): 28A.
- VAN DER GIEZEN M., TOVAR J. 2004: Mitosomes, hydrogenosomes and mitochondria: variations on a theme? In: R.P. Hirt and D.S. Horner (Eds.), Organelles, Genomes and Eukaryote Phylogeny. An Evolutionary Synthesis in the Age of Genomics. CRC Press, Boca Raton, pp. 289–308.
- VIVARÈS C.P., GOUY M., THOMARAT F., MÉTÉNIER G. 2002: Functional and evolutionary analysis of a eukaryotic parasitic genome. Curr. Opin. Microbiol. 5: 499–505.
- VIVARÈS C.P., MÉTÉNIER G. 2004: The microsporidia genome: living with minimal genes as an intracellular eukaryote. In: D.S. Lindsay and L.M. Weiss (Eds.), Opportunistic Infections: *Toxoplasma*, *Sarcocystis* and Microsporidia. World Class Parasites. Vol. 9. Kluwer Academic Publishers, Boston, pp. 215–242.
- WEISS L.M., ZHOU G.C., ZHANG H. 2003: Characterization of recombinant microsporidian methionine aminopeptidase type 2. J. Eukaryot. Microbiol. 50 (Suppl.): 597–599.
- XU Y., TAVKORIAN P.M., CALI A., ORR G., WEISS L.M. 2004: Glycosylation of the major polar tube protein of *Encephalitozoon hellem*, a microsporidian parasite that infects humans. Infect. Immun. 72: 6341–6350.
- XUE H., TONG K.-L., MARCK C., GROSJEAN H., WONG J.T. 2003: Transfer RNA paralogs: evidence for genetic code-aminoacid biosynthesis coevolution and an archaeal root of life. Gene 310: 59–66.

Received 8 December 2004

Accepted 2 February 2005