

Inteins in pathogenic fungi: a phylogenetic tool and perspectives for therapeutic applications

Raquel Cordeiro Theodoro, Eduardo Bagagli/+

Departamento de Microbiologia e Imunologia, Instituto de Biociências, UNESP-Botucatu, Distrito de Rubião Junior s/n, 18618-000 Botucatu, São Paulo, Brasil

Inteins or “internal proteins” are coding sequences that are transcribed and translated with flanking sequences (exteins). After translation, the inteins are excised by an autocatalytic process and the host protein assumes its normal conformation and develops its expected function. These parasitic genetic elements have been found in important, conserved proteins in all three domains of life. Most of the eukaryotic inteins are present in the fungi kingdom and the PRP8 intein is one of the most widespread inteins, occurring in important pathogens such as Cryptococcus neoformans (varieties grubii and neoformans), Cryptococcus gattii, Histoplasma capsulatum and Paracoccidioides brasiliensis. The knowledge of conserved and non-conserved domains in inteins have opened up new opportunities for the study of population variability in pathogenic fungi, including their phylogenetic relationships and recognition or diagnoses of species. Furthermore, inteins in pathogenic fungi should also be considered a promising therapeutic drug target, since once the autocatalytic splicing is inhibited, the host protein, which is typically vital, will not be able to perform its normal function and the fungal cell will not survive or reproduce.

Key words: Intein - PRP8 species - pathogenic fungi - therapeutic target

The modern synthesis of natural selection with the Mendelian inheritance theory explains how natural selection can interfere with gene frequencies in different populations. But how does natural selection operate? This has been the theme of several discussions among important evolutionists. Some believe that it operates over a group or the individual and others assume that natural selection acts at the molecular level, operating over genes. The latter hypothesis has been consistently defended by Richard Dawkins (1989), in his book entitled “The Selfish Gene”, where he proposed that all living organisms are machines for gene survival and replication, such that natural selection will obviously favour those genes that increase the fitness of their hosts. Parasitic genetic elements, such as inteins, are excellent examples of how natural selection occurs at the molecular level. Inteins are considered parasitic genes because they are found in coding genes, with which they are transcribed and translated and, apparently, do not contribute to the organism’s fitness. The inteins undergo a post-translational excision by an autocatalytic process and the host protein assumes its normal conformation, by a peptide ligation, and can develop its normal function (Fig. 1) (Cooper & Stevens 1995).

The first described intein was found in the vacuolar ATPase gene from *Sacharomyces cerevisiae*. The central region of this gene presented high similarity to some

endonucleases (Hirata et al. 1990). Interestingly, the ATPase protein showed a molecular weight of 70KDa, which was less than expected and corresponded to the weight without the central region. In 1990, Kane et al. demonstrated that the insertion (corresponding to an endonuclease) was present in the mRNA and that all the mRNA was translated. Therefore, the 70kDa protein was a result of post-translational processing.

Inteins are coding sequences whose translated products can range from 134-608 amino acids. They excise themselves from the host protein and join the flanking protein sequences (exteins) with a peptide ligation, thus generating a functional protein (Gogarten et al. 2002). There are two intein types, mini-inteins and full-length inteins, both of which present a splicing domain containing four blocks of more or less well-conserved amino acid sequences. Blocks A and B constitute the N-terminal splicing domain and blocks F and G constitute the C-terminal splicing domain. In addition, full-length inteins present an endonuclease domain that can perform a homing function [homing endonuclease gene (HEG)] with the potential to make the intein a mobile genetic element, resulting in occupation of empty alleles and duplication of the parasitic genetic element (Liu 2000). The transfer of an intein to a cognate and empty allele considerably increases the frequency of HEG⁺ alleles in the population through super-Mendelian inheritance (Burt & Koufopanou 2004). Therefore, in a few generations, the HEG intein may be fixed in the population. There are different endonuclease types that are encoded by the HEG domain, but endonucleases from the LAGLIDADG family are the most frequent (Perler 2000). The HEG is divided into blocks C, D, E and H, which can be found in some introns that also have a “homing” capacity (Belfort et al. 1995).

+ Corresponding author: bagagli@ibb.unesp.br

Received 21 July 2008

Accepted 13 March 2009

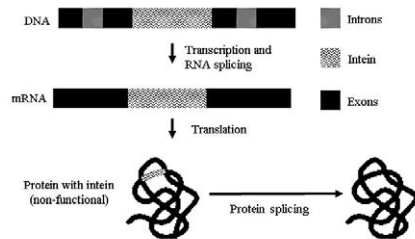


Fig. 1: scheme of intein splicing from the host protein (extein) after its translation.

Inteins have been detected in all three domains of life: eukaryotes, bacteria and archaea, according to the intein database InBase (<http://www.neb.com/neb/inteins.html>). They are usually found at conserved sites of housekeeping proteins that have vital functions in the cell, such as metabolic enzymes, DNA and RNA polymerases, proteases and others (Liu 2000). For this reason, any mutation that disrupts the functionality of the splicing domain can be fatal for the cell. In this manner, natural selection operates at the molecular level: although the inteins do not appear to promote any improvement in the fitness of the organism, their localization at important and thus conserved sites can be the reason for their survival in genomes, since they recognize specific and conserved genetic niches in the genome for the homing process.

Horizontal transfer of the intein coding sequence and a small portion of the flanking sequence readily accounts for the discontinuous intein dispersion in corresponding insertion points, since inteins have an established homing mechanism (Dassa & Pietrokovski 2008). This horizontal transfer, as well as independent loss of inteins from different lineages, contributes to their large and sporadic phylogenetic distribution.

When inteins occupy exactly the same site, they are referred to as “allelic inteins”, even though they are from different species (Butler et al. 2006). Today, more than 130 inteins are known to be distributed among approximately 34 host proteins with distinct functions (Perler 2000). Some inteins are present in distantly related organisms. For example, the DnaB intein is present in a photosynthetic cyanobacterium and the heterotrophic marine thermophile *Rhodothermus marinus* (Liu & Hu 1997). An intein can also parasitise functionally different proteins that display structural similarity, as in the case of the intein TopA, which is present in topoisomerase I and reverse gyrase (Chute et al. 1998). Inteins are inherited vertically when alleles with and without inteins are assembled during sexual reproduction, enabling the homing process. This was documented through the invasion of an intein-containing allele into the *vma* allele without the intein in *S. cerevisiae*. After meiosis, more than 50% of the *vma*-1 alleles had the intein (Gimble & Thorner 1992).

The distribution and population dynamics of the VMA intein seem to follow a model of invasion, fixation, degeneration, loss and reinvasion. Once the intein

is established in the population, there is no selection for endonuclease functionality, such that this domain can become degenerate through mutations, deletions and genetic drift. These empty sites will be reoccupied again according to the cyclic model (Burt & Koufopanou 2004) (Fig. 2). However, a recent analysis of HEG inteins suggested that the homing endonuclease domain can persist over long evolutionary times due higher recombination frequency, complex population structures preventing the fixation of HEG, a balance between the spreading of HEG and its possible decrease in relation to the host's fitness, or a function of HEG that increases the fitness of its host (Gogarten & Hilario 2006).

It is also known that some inteins could have evolved from a parasitic to a functional role (Fig. 2) through positive selection. One example is the endonuclease HO, which is responsible for mating-type switching, between the sexual types α and a , in *S. cerevisiae* (Liu 2000). This endonuclease is very similar to the VMA1 intein from *S. cerevisiae* and, although inactive, it also has conserved regions related to splicing activity. This is a strong indication that HO has evolved from a VMA intein (Gogarten et al. 2002, Haber & Wolfe 2005).

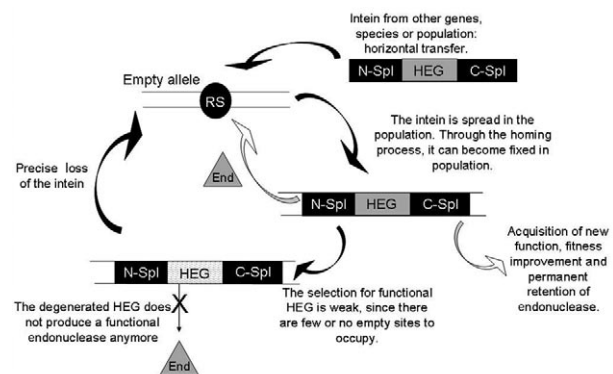


Fig. 2: homing cycle of the VMA intein (modified from Gogarten & Hilario 2006). C-Spl: splicing domain C-terminal; End: endonuclease encoded by the HEG domain; HEG: homing endonuclease gene; N-Spl: splicing domain N-terminal; RS: endonuclease recognition site.

Inteins in fungi - As already mentioned, inteins are present in the three life domains: eukarya, bacteria and archaea. According to Inbase (Perler 2000), about 60 species from the eukarya domain present inteins in some important genes. Among these species are some amoebae and algae, but the majority of the eukaryotes with inteins are fungi (about 50 species), from distinct phyla (Chytridiomycota, Zygomycota, Basidiomycota and Ascomycota). The majority of inteins in fungi are in the host genes Vacuolar ATPase (H⁺transporting ATP synthase) subunit A (VMA intein) and *prp8* (PRP8 intein). The VMA intein is widespread among yeast fungi, such as *Saccharomyces*, *Zygosaccharomyces* and some *Candida* (all Ascomycota) species, while the PRP8

intein is present in all fungi phyla (Perler 2000), being more abundant among Ascomycetes, some species of the orders Eurotiales, Onygenales and Helotiales, a few Basidiomycetes from the order Tremellales and the chytrid fungus, *Batrachochytrium dendrobatidis* (Inbase). Some important human pathogenic species, such as *Histoplasma capsulatum*, *Paracoccidioides brasiliensis* (Ascomycota, order Onygenales), *Cryptococcus neoformans*, and *Cryptococcus gattii* (Basidiomycota, order Tremellales) are known to have the PRP8 intein (Butler et al. 2006).

PRP8 intein

This intein is located in the *prp8* gene, which codes for the PRP8 protein, the largest and most highly conserved nuclear protein (61% similarity between yeast and human), which occupies a central position in the catalytic core of the spliceosome. It is a component of U5 snRNPs and is implicated in several crucial molecular rearrangements during pre-mRNA splicing (for the removal of introns), a central step in gene expression (Grainger & Beggs 2005).

The PRP8 intein was initially described in a large number of *C. neoformans* (varieties *grubii* and *neoformans*) and *C. gattii* isolates as mini-inteins (without the HEG domain) (Butler et al. 2001, Butler & Poulter 2005). Besides these *Cryptococcus* species, the PRP8 intein can also be found as a full-length intein in *Cryptococcus laurentii*. Since this species is moderately distant from *C. neoformans*, Butler & Poulter (2005) suggested that the mini-intein found in *C. neoformans* and *C. gattii* was derived from horizontal transfer, with *C. laurentii* or another species being the donor.

Butler et al. (2006) searched for PRP8 inteins in a database of genomes and discovered it in many species of Ascomycetes. The authors also compared the PRP8 intein with the VMA intein, with relation to their splicing and HEG domains and HEG functionality and, interestingly, noted that the PRP8 intein can be found as a mini-intein or full-length intein, depending on the fungal species, whereas all known VMA inteins are full-length. In contrast to the VMA intein, the PRP8 intein seems to have a functional HEG domain in most of the host species. Analysis of dS/dN (the ratio between synonymous and nonsynonymous substitutions) also revealed that the HEG domain from PRP8 is saturated with synonymous substitutions when compared to the HEG domain from VMA, suggesting that they are constrained by selection and have been diverging for a longer time than the VMA intein. This is also corroborated by the sporadic occurrence of the PRP8 intein, which is missing in some species closely related to fungi that contain the PRP8 intein. This is the case of *Coccidioides immitis* and *Coccidioides posadasii*, two pathogenic “inteinless” species that are closely related to *Uncinocarpus reesii*, which contains a PRP8 mini-intein. The PRP8 intein appears to have an ancient origin in Ascomycota, and the “inteinless” species indicate that the loss of an intein (by deletion or fixation of an empty allele) could be quite frequent.

Besides the *Cryptococcus* species that belonged to the Basidiomycota phylum, inteins in the *prp8* gene were

also described in many Euscomycetes, such as *Aspergillus nidulans*, *Aspergillus fumigatus*, *P. brasiliensis*, *H. capsulatum* and *Botrytis cinera*. In this group, however, there is no evidence of horizontal transfer, since the phylogenetic analysis of these inteins is similar to those using other nuclear sequences, for example, ribosomal sequences, which suggests an ancient origin and vertical transmission among Euscomycetes (Butler et al. 2006). Since only a few *Cryptococcus* species (Basidiomycota, order Tremellales) present the PRP8 intein, of which *C. laurentii* is the only one with a full-length intein (Butler & Poulter 2005) and no other basidiomycetes have been shown to have inteins, it seems reasonable to explain this distribution of the PRP8 intein as the result of horizontal transfer between ascomycetes (donors) and basidiomycetes (receptors). This transfer, according to Poulter et al. (2007), could have occurred during the parasitic phase, in co-infection cases, or during the saprobic phase of these pathogens, when different fungal cells are phagocytosed by microorganisms from soil. In fact, the interaction between pathogenic fungi and phagocytic amoebae from soil has been the theme of several studies that proved the similarity between fungus-macrophage and fungus-amoeba interactions. This proposes that virulence features could have emerged in a saprobic environment and today present a dual use (in both saprobic and host environments) (Casadevall et al. 2003, Steenbergen & Casadevall 2003).

Species concept and use of PRP8 intein for phylogenetic studies in pathogenic fungi

The species concept is one of the central themes of evolution, and although a “species” is considered a concrete group, in contrast to artificial ones (such as genera, order, class, phylum and kingdom), its definition and recognition are considered a challenge, even in the genomic and proteomic era.

Under an evolutionary approach, members of the same species are expected to form a genetic and ecological unit, sharing the same alleles and exploring the same ecological niche. This definition seems to be very simple and direct, but how can we recognize it in nature? The morphological recognition of species for example is based on common phenotypes presented in members of the same species. However, it is typical to group genetically different organisms into one species, mainly when we deal with unicellular organisms, whose phenotypes are restricted to chemical interactions with substrates or a few macro and microscopic features. The biological recognition of species takes into account the interbreeding capacity among members from the same species. However, this recognition is not applicable for asexual populations, and even in sexual ones, it may lead to a conflict between potential and real gene flow (Taylor et al. 2006). There is also an ecological recognition that defines a species as a group of individuals that occupy the same niche, a concept that is closely connected to the biological one, as natural selection favours individuals that interbreed only with others with the same ecological adaptation. It is known that the adaptive value of hybrids could be, in many cases, lower than that of the parental species, such that parental species will present better

competitive adaptations for occupying their niches (Ridley 2006). This concept is also very difficult to apply to microorganisms, whose niche is defined by biochemical interactions with substrates (environmental and host) and is compounded by n dimensions (Hutchinson 1959, Bagagli et al. 2008).

Recently, cryptic speciation was revealed in some important fungal pathogens by phylogenetic recognition of species, which detects genetic divergence among populations through multi-locus sequence typing, by concordance of gene genealogies (Taylor et al. 2000). This method also provides information about the reproductive mode of the different divergent species, even without the isolation of sexual structures. Cryptic species have been detected by gene genealogies in important pathogenic fungi, such as *C. immitis* (separated into 2 species: *C. immitis* and *C. posadasii*) (Koufopanou et al. 2001), *H. capsulatum* (separated into at least 7 species) (Kasuga et al. 2003) and *P. brasiliensis*, which is now separated into at least four previously unrecognized species: S1 (species 1 from Brazil, Argentina, Paraguay, Peru and Venezuela), PS2 (phylogenetic species 2 from Brazil and Venezuela), PS3 (phylogenetic species 3 with 21 isolates from Colombia) (Matute et al. 2006) and Pb01-like (from the central region of Brazil) (Carrero et al. 2008, Teixeira 2008).

Knowledge of the variability among fungal isolates that cause important diseases, including systemic mycoses, is very important not only for population studies but also, and mainly, to correlate different clinical manifestations and distinguish among genotypes of isolates to improve diagnoses and treatment. Once the species are genetically separated, they are supposed to accumulate some morphological differences, which may result in the occupation of new ecological niches or different strategies for survival in the saprobe or host environment. However, if the recognition of cryptic species is difficult and laborious (since the use of many nuclear coding sequences is required), the search for morphological, physiological, ecological and clinical markers correlated with the genetic clustering is even more challenging. Usually, after the discovery of cryptic species, some phenotypes that were originally considered merely intra-specific polymorphisms become associated with isolated genetic groups. For example, *C. posadasii* presents slower growth in high salt concentration media than *C. immitis*, however, this feature is not used for diagnosis (Fisher et al. 2002). In some cases, when the host selective pressure on the pathogen population is sufficient to be considered an allopatric speciation factor, the host can be considered a marker for species recognition. For example, the genus *Pneumocystis*, characterized by non-cultivable and highly diversified fungal pathogens that dwell in the lungs of mammals, comprises numerous host-specific species, such as *Pneumocystis jirovecii* in humans, *Pneumocystis carinii* and *Pneumocystis wakefieldiae* in rats, *Pneumocystis murina* in mice, and *Pneumocystis oryctolagi* sp. nov. in old world rabbits (Dei-Cas et al. 2006).

The search for morphological and/or molecular markers that efficiently differentiate the cryptic species in pathogenic human fungi is essential for practical medical purposes. But what constitutes a good marker? Most morphological characteristics are extremely variable because they usually result from the expression of innumerable genes. A good morphological marker should clearly distinguish between various species and also be constant and repeatable among members of the same species. The same is required for molecular markers but, in this case, polymorphisms can be objectively distinguished as it is based on nucleotide substitutions. Generally, coding nuclear genes with sufficient polymorphism(s) to separate the species or genetic groups are the most useful markers.

Inteins also seem to constitute a promising source of phylogenetic information, especially the full-length inteins, because they are expected to have more sequence variation in the endonuclease domain than in the splicing domain due to a more relaxed selection. This is particularly true if the HEG is no longer active (Gogarten & Hilario 2006). In the case of PRP8 inteins, even though their HEG domain appears to be constrained by selection (most PRP8 inteins may have an active endonuclease), they are very diverse, probably due to an ancient origin and a long evolutionary history (Butler et al. 2006) and are useful sequences for phylogenetic inferences in intein-carrying fungal species.

Since the *prp8* gene is highly conserved among eukaryotes, primer design for its amplification by PCR is possible for phylogenetic studies of fungal species and varieties. Using this approach, Butler and Poulter (2005) distinguished between strains of two varieties of *C. neoformans* (*neoformans* and *grubii*) and from strains of *C. gattii* (recently designated as a new species) by the neighbour joining method. These encapsulated yeasts are capable of causing serious infections in both immunocompromised and immunocompetent persons. The varieties *grubii* (serotype A), *neoformans* (serotype D), and *gattii* (serotypes B and C) were initially designated according to antigenic differences in the capsular polysaccharide. This was also confirmed by molecular data analyses (Franzot et al. 1999, Meyer et al. 1999, Ellis et al. 2000). The *gattii* variety has recently been accorded species status as *C. gattii* (Boekhout et al. 2001, Kwon-Chung et al. 2002). Although the use of the PRP8 intein for phylogenetic analysis at a higher taxonomic level is not viable due to its erratic distribution, it proved to be a suitable taxonomic marker at the species or sub-species level for epidemiological studies of *Cryptococcus*.

We applied the same approach to sequence and phylogenetically analyze the PRP8 intein in different isolates of the *P. brasiliensis* species complex (S1, PS2, PS3 and Pb01-like species) (Theodoro et al. 2008). This thermophilic fungus is the etiologic agent of paracoccidioidomycosis, the most important systemic mycosis in Latin America (Franco 1987). It has been frequently recovered from clinical samples and tissues of the nine-banded armadillo, *Dasypus novemcinctus* (Bagagli et al. 2003). It was also recently detected in and isolated from dogs (Ricci et al. 2004, Farias et al. 2005). However, mo-

lecular polymorphism analysis suggested that the same “ecopathogenotypes” can infect both humans and other animals (Hebeler-Barbosa et al. 2003).

In our assay, the PRP8 intein from 22 isolates of *P. brasiliensis* belonging to the four previously recognized species was sequenced. Phylogenetic analysis by maximum-parsimony, maximum likelihood and Bayesian analysis clearly separated the isolates from the four species and revealed a significant difference between Pb01-like and the remaining species (Theodoro et al. 2008). This corroborated previous phylogenetic analyses of several nuclear encoding regions, such as *GP43* (the gene coding for the immunodominant antigen of *P. brasiliensis*), *CH4*, *actin*, *ODC*, *URA3*, *CHS2*, *FKS1*, *HSP70*, *Hydrophobin*, *Kex*, *Catalase A*, *Catalase P*, *Formamidase* and *Glyoxalase* (Carrero et al. 2008). The Pb01 genotype has been isolated from the central region of Brazil and, interestingly, there is no clear geographic barrier that could explain this divergence, since this species is practically sympatric with respect to the S1 and PS2 groups. Morphological assays could clarify some important features about the exploration of the host and environmental ecological niches of this pathogen. Additionally, other mycological characteristics, such as virulence and antifungal resistance, are also necessary to determine to address whether the different genetic groups present different clinical aspects.

All the evaluated *P. brasiliensis* isolates presented a full-length intein in the *prp8* gene. The HEG domain of the PRP8 intein from *P. brasiliensis* appears to be inactive because the second aspartic acid residue, which is indispensable for functionality, was replaced by a serine in the Pb01 isolate (from the Pb01-like genetic group) and a glycine in the rest of the isolates. However, the ratio between synonymous and nonsynonymous substitutions (dS/dN) produced values greater than 1, indicating that this region is under conservative selection and probably was active until recently. Although the degree of polymorphisms in the PRP8 intein from *P. brasiliensis* was not as high as in the commonly used nuclear-coding genes, this sequence contains sufficient phylogenetic information to separate the cryptic species of *P. brasiliensis* and therefore constitutes a reliable molecular marker for this pathogen.

PRP8 as a therapeutic target

The discovery of inteins has offered new biotechnological tools for protein purification on account of the protein splicing nature of these parasitic genetic elements. The traditional methods for protein purification, which require histidine tags used in affinity chromatography, have low yield and can compromise the stability, solubility and activity of the desired protein (Gogarten et al. 2002). By mutating the N or C terminus of an intein, a desired gene can be ligated and expressed in-frame with the intein and excised with it. The protein can be purified by releasing it from the mutant intein through a change in reducing conditions (Chong et al. 1997, 1998). Inteins have also been used in protein splicing, autoprolysis and biosynthesis of reconstructed polypeptides with reactive ends necessary for intein-mediated protein ligation (Evans et al. 1998).

Besides these biotechnological applications, the distribution, location and functionality of inteins make them promising therapeutic targets against important pathogens. First, intein splicing is required for the production of mature and functional protein (extein). Since inteins are commonly inserted into vital proteins, the inhibition of its splicing would lead to microorganism death. Second, since no intein has been found in animals or humans, this therapeutic target seems to be very safe for the pathogens hosts, and third, screening systems for the detection of mutations or inhibitors of the splicing domain can be tested in *Escherichia coli* (Derbyshire et al. 1997).

Some inteins occur not in vital proteins, but in some proteins required in specific conditions. For example, the VMA intein is in the vacuolar H⁺-ATPase, which is essential for growth in the presence of CaCl₂. This can be useful for the selection of splicing inhibitors that inhibit yeast growth in medium with CaCl₂ (Nogami et al. 1997). Selection of systems for the inhibition of protein splicing can be done artificially by inserting inteins into proteins that provide some selective advantage.

One of the first systems used to examine the intein splicing function in a non-native host context was based on the β-galactosidase complementation assay for the RecA intein of *Mycobacterium tuberculosis*. RecA splicing yielded an active LacZα fragment, which could be easily detected by the blue/white assay (Davis et al. 1992). Using this approach, Daugelat and Jacobs (1999) inserted a RecA intein from *M. tuberculosis* with a C-terminal cysteine residue next to the enzyme responsible for kanamycin resistance. This same intein was also used in an in vitro screening system for protein splicing inhibitors based on green fluorescent protein (GFP), as an indicator when expressed in *E. coli* (Gangopadhyay et al. 2003).

The first potential genetic selection for intein splicing in yeast was the transfer of the Sce intein (*S. cerevisiae* VMA1 intein) into the *vat2* gene, which encodes the 60kDa subunit of the vacuolar H⁺-ATPase in yeast (Cooper et al. 1993). Knockout strains for this gene grow at pH 5.0, but not at pH 7.5 (Yamashiro et al. 1990). Therefore, growth at pH 7.5 using an intein-Vat2 fusion would select for active splicing.

Similar studies on PRP8 fungal inteins were carried out to evaluate the splicing activity of this intein in a non-native host protein in *E. coli*. The PRP8 intein from *C. neoformans* has been studied by mutagenesis to identify the minimal sequence requirements, as well as the importance of a number of conserved residues for its splicing in a foreign protein context (Pearl et al. 2007). The same mutagenic approach was used for the PRP8 mini-intein of *Penicillium chrysogenum*, the smallest known nuclear-encoded active splicing protein, by introducing deletions at different sites. The authors removed a total of 14 amino acid residues (8 and 6 at 2 different sites) and the resulting 143aa intein was functional when expressed in *E. coli* (Elleuche et al. 2008). Elleuche et al. (2006) also expressed the PRP8 mini-intein of *P. chrysogenum*, *Penicillium expansum* and *Penicillium*

vulpinum in *E. coli* in a model host protein and a GFP model system. This same intein, from *P. chrysogenum*, was also tested in a *trans*-splicing assay by splitting it artificially. Protein *trans*-splicing involves the joining of two halves of an intein sequence through their high affinity for each other followed by ligation of the two extein sequences. Therefore, these split inteins self-associate and catalyze protein-splicing activity in *trans* (Tavassoli et al. 2008). Elleuche & Pöggeler (2007) tested three split sites in the *P. chrysogenum* PRP8 intein. Two of these sites resulted in protein *trans*-splicing: one that corresponded to the insertion site of the endonuclease domain in allelic large PRP8 inteins and the other was detected at a new position, located N-terminal of the endonuclease insertion site.

The splicing of PRP8 inteins from *A. fumigatus*, *A. nidulans*, *H. capsulatum* and *C. neoformans* was also evaluated by inserting them into the expression plasmid pMST between the *Xho*I and *Age*I sites, replacing the *Ssp* DnaB intein coding sequence of the plasmid. This insertion site was flanked by a maltose binding protein at its N-terminus and a thioredoxin at its C-terminus. Western blots were carried out using anti-thioredoxin antibody. The precursor protein, spliced protein and excised intein were identified by their predicted sizes. Intein splicing was evaluated at 25°C and 36°C for the four fungal species and in general splicing efficiency was higher at 25°C. This relationship was more significant for *H. capsulatum*, whose splicing efficiency at 37°C was only 25%, while at 25°C it was 100% (no band for the precursor protein was observed by western blot) (Liu & Yang 2004).

It is interesting to note that, as a dimorphic fungus, *H. capsulatum* must convert from the mycelia (M) to yeast (Y) form during infection or under high temperature (37°C). If the splicing efficiency of the intein in the *prp8* gene at 37°C is as low as in the non-native gene host, the M-Y transition could be disturbed because many different genes must be transcribed and correctly translated for the fungus to assume the yeast morphology and physiology. In this case, the PRP8 intein could decrease the fitness of the pathogen and, as the PRP8 intein from *H. capsulatum* is a full-length intein that codes for a functional HEG (Poulter et al. 2007), one may ask why this HEG appears to be maintained by selection. According Gogarten and Hilario (2006), the balance between decreased fitness of the host and active spread of the parasitic element could lead to long persistence time of the HEG, especially when the selective disadvantage to the carrier and the mobility of the intein were not the same in different subpopulations. Since *H. capsulatum* is a cosmopolite pathogen, with many subpopulations, including cryptic species, this hypothesis for the maintenance of the PRP8 intein in this pathogen seems very reasonable.

The ability of the PRP8 intein to self-splice from a non-native protein from *E. coli* was merely the first step for future studies and the identification of splicing inhibitors may constitute potential antifungal drugs. Since PRP8 inteins from different pathogens have the same corresponding insertion site in the PRP8 protein and, differently from the HEG domain, their splicing domain

is highly conserved (Liu & Yang 2004), splicing inhibitor drugs may have the advantage of simultaneous action against most fungal diseases.

REFERENCES

- Bagagli E, Franco M, Bosco SMG, Hebel-Barbosa F, Trinca LA, Montenegro MR 2003. High frequency of *Paracoccidioides brasiliensis* infection in armadillo (*Dasypus novemcinctus*): an ecological study. *Med Mycol* 41: 217-223.
- Bagagli E, Theodoro RC, Bosco SMG, McEwen J 2008. *Paracoccidioides brasiliensis*: phylogenetic and ecological aspects. *Mycopathologia* 165: 197-207.
- Belfort M, Reaban ME, Coetzee T, Dalgaard JZ 1995. Prokaryotic introns and inteins: a panoply of form and function. *J Bacteriol* 177: 3897-3903.
- Boekhout T, Theelen B, Diaz M, Fell JW, Hop WC, Abeln EC, Dromer F, Meyer W 2001. Hybrid genotypes in the pathogenic yeast *Cryptococcus neoformans*. *Microbiol* 147: 891-907.
- Burt A, Koufopanou V 2004. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr Opin Genet Dev* 14: 609-615.
- Butler ML, Goodwin TJD, Poulter RTM 2001. A nuclear-encoded intein in the fungal pathogen *Cryptococcus neoformans*. *Yeast* 18: 1365-1370.
- Butler ML, Gray J, Goodwin TJD, Poulter RTM 2006. The distribution and evolutionary history of the PRP8 intein. *BMC Evol Biol* 6: 1-26.
- Butler ML, Poulter RTM 2005. The PRP8 inteins in *Cryptococcus* are a source of phylogenetic and epidemiological information. *Fungal Genet Biol* 42: 452-463.
- Carrero LL, Niño-Vega G, Teixeira MM, Carvalho MJA, Soares CMA, Pereira M, Jesuino RSA, McEwen JG, Mendoza L, Taylor J W, Felipe MS, San-Blas G 2008. New *Paracoccidioides brasiliensis* isolate reveals unexpected genomic variability in this human pathogen. *Fung Genet Biol* 45: 605-612.
- Casadevall A, Steenbergen JN, Nosanchuk JD 2003. "Ready-made" virulence and "dual use" virulence factors in pathogenic environmental fungi: the *Cryptococcus neoformans* paradigm. *Curr Opin Microbiol* 6: 332-337.
- Chong S, Mersha FB, Comb DG, Scott ME, Landry D 1998. Single-column purification of free recombinant proteins using a self-cleavable affinity target derived from a protein splicing element. *Gene* 192: 271-281.
- Chong S, Montello GE, Zhang A, Cantor EJ, Liao W 1997. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nuc Acids Res* 26: 5109-5115.
- Chute IC, Hu Z, Liu XQ 1998. A topA intein in *Pyrococcus furiosus* and its relatedness to the r-gyr intein of *Methanococcus jannaschii*. *Gene* 210: 85-92.
- Cooper AA, Chen YJ, Lindorfer MA, Stevens TH 1993. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J* 12: 2575-2583.
- Cooper AA, Stevens TH 1995. Protein splicing: self-splicing of genetically mobile elements at the protein level. *Trends Biochem Sci* 20: 351-356.
- Dassa B, Pietrokovski S 2008. Origin and evolution of inteins and other Hint domains. In M Belfort, V Derbyshire, BL Stoddard, DW Wood, *Homing endonucleases and inteins*, Springer-Verlag Berlin, Heidelberg, p. 211-229.

- Daugelat S, Jacobs WR 1999. The *Mycobacterium tuberculosis recA* intein can be used in an ORFTRAP to select for open reading frames. *Prot Sci* 8: 644-653.
- Davis EO, Jenner PJ, Brooks PC, Colston MJ, Sedwick SG 1992. Protein splicing in the maturation of *M. tuberculosis recA* protein: a mechanism for tolerating a novel class of intervening sequence. *Cell* 71: 201-210.
- Dawkins R 1989. *The selfish gene*, Oxford Univ. Press, Oxford, 540 pp.
- Dei-Cas E, Chabe M, Moukhliis R, Durand-Jolly I, Aliouat el M, Stringer JR, Cushion M, Noel C, de Hoog GS, Guillot J, Viscogliosi E 2006. *Pneumocystis oryctolagi* sp. nov., an uncultured fungus causing pneumonia in rabbits at weaning: review of current knowledge and description of a new taxon on genotypic, phylogenetic and phenotypic bases. *FEMS Microbiol Rev* 30: 853-871.
- Derbyshire V, Wood DW, Wu W, Dansereau JT, Dagaard JZ, Belfort M 1997. Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. *Proc Natl Acad Sci USA* 94: 11466-11471.
- Elleuche S, Döring K, Pöggeler S 2008. Minimization of a eukaryotic mini-intein. *Biochem Biophys Res Commun* 366: 239-243.
- Elleuche S, Nolting N, Pöggeler S 2006. Protein splicing of PRP8 mini-inteins from species of the genus *Penicillium*. *Appl Microbiol Biotechnol* 72: 959-967.
- Elleuche S, Pöggeler S 2007. Trans-splicing of an artificially split fungal mini-intein. *Biochem Biophys Res Commun* 355: 830-834.
- Ellis D, Marriott D, Hajjeh RA, Warnock D, Meyer W, Barton R 2000. Epidemiology, surveillance of fungal infections. *Med Mycol* 38: 173-182.
- Evans TC, Benner J, Xu MQ 1998. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* 7: 2256-2264.
- Farias MR, Werner J, Muro MD, Marques SA, Marques MEA, Franco MF, Ribeiro MG, Custodio CC, Condas LAZ, Bosco SMG, Bagagli E 2005. Canine paracoccidioidomycosis: case report of generalized lymphadenitis. *Rev Inst Med Trop São Paulo* 47: 64.
- Fisher MC, Koenig GL, White TJ, Taylor JW 2002. Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognized as the non-California population of *Coccidioides immitis*. *Mycologia* 94: 73-84.
- Franco M 1987. Host-parasite relationships in paracoccidioidomycosis. *J Med Vet Mycol* 25: 5-18.
- Franzot SP, Salkin IF, Casadevall A 1999. *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. *J Clin Microbiol* 37: 833-840.
- Gangopadhyay JP, Jiang S, Paulus H 2003. An *in vitro* screening system for protein splicing inhibitors based on green fluorescent protein as an indicator. *Anal Chem* 75: 2456-2462.
- Gimble FS, Thorner J 1992. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature* 357: 301-306.
- Gogarten JP, Hilario E 2006. Inteins, introns and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol Biol* 6: 1-5.
- Gogarten JP, Senejani AG, Zhaxybayeva O, Olenzinski L, Hilario E 2002. Inteins: structure, function and evolution. *Ann Rev Microbiol* 56: 263-287.
- Grainger RJ, Beggs JD 2005. Prp8 protein: at the heart of the splicosome. *RNA* 11: 533-557.
- Haber JE, Wolfe KH 2005. Function and evolution of HO and VDE endonucleases in fungi. *Nucleic Acids Mol Biol* 16: 161-175.
- Hebeler-Barbosa F, Morais FV, Montenegro MR, Kuramae EE, Montes B, McEwen JG, Puccia R, Bagagli E 2003. Sequence comparison of the internal transcribed spacer regions and gp 43 in *Paracoccidioides brasiliensis* for patients and armadillos *Dasyurus novemcinctus*. *J Clin Microbiol* 41: 5735-5737.
- Hirata R, Ohsumi Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y 1990. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H (+)- translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem* 265: 6726-6733.
- Hutchinson GE, MacArthur RH 1959. A theoretical ecological model of size distribution among species of animals. *Am Nat* 93: 117-125.
- Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebel M, Stevens TH 1990. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 250: 651-657.
- Kasuga T, White TJ, Koenig G, McEwen J, Restrepo A, Castañeda E, Lacaz CS, Heins-Vaccari, EM, Freitas RS, Zancopé-Oliveira RM, Qin Z, Negroni R, Carter DA, Mikami Y, Tamura M, Taylor ML, Miller GF, Poonwan N, Taylor JW 2003. Phylogeography of the fungal pathogen *Histoplasma capsulatum*. *Mol Ecol* 12: 3383-3401.
- Koufopanou V, Burt A, Szaro T, Taylor JW 2001. Gene genealogies, cryptic species and molecular evolution in the human pathogen *Coccidioides immitis* and relatives (Ascomycota, Onygenales). *Mol Biol Evol* 18: 1246-1258.
- Kwon-Chung KJ, Boekhout T, Fell JW, Diaz M 2002. Proposal to conserve the name *Cryptococcus gattii* against *C. Hondurianus* and *C. Bacillisporus* (Basidiomycota, Hymenozetes, Tremellomycetidae). *Taxon* 51: 804-806.
- Liu XQ 2000. Protein-splicing intein: genetic mobility, origin and evolution. *Ann Rev Genet* 34: 61-76.
- Liu XQ, Hu Z 1997. A DnaB intein in *Rhodothermus marinus*: indication of recent intein homing across remotely related organisms. *Proc Natl Acad Sci USA* 94: 7851-7856.
- Liu XQ, Yang J 2004. Prp8 intein in fungal pathogens: target for potential antifungal drugs. *FEBS Letters* 572: 46-50.
- Matute DR, McEwen JG, Montes BA, San-Blas G, Bagagli E, Rauscher JT, Restrepo A, Morais F, Nino-Veja G, Taylor JW 2006. Cryptic speciation and recombination in the fungus *Paracoccidioides brasiliensis* as revealed by gene genealogies. *Mol Biol Evol* 23: 65-73.
- Meyer W, Marszewska K, Amirmostofian M, Igreja RP, Hardtke C, Methling K, Viviani MA, Chindamporn A, Sukroongreung S, John MA, Ellis DH, Sorrell TC 1999. Molecular typing of global isolates of *Cryptococcus neoformans* var. *neoformans* by polymerase chain reaction fingerprinting and randomly amplified polymorphic DNA—a pilot study to standardize techniques on which to base a detailed epidemiological survey. *Electrophoresis* 20: 1790-1799.
- Nogami SY, Satow Y, Anraku Y 1997. Probing novel elements for protein splicing in yeast Vma1 protozyme: a study of replacement mutagenesis and intragenic suppression. *Genetics* 147: 73-85.
- Pearl EJ, Tyndall JDA, Poulter RTM, Wilbanks SM 2007. Sequence requirements for splicing by the Cne PRP8 intein. *FEBS letters* 581: 3000-3004.
- Perler FB 2000. InBase, the intein database. *Nucleic Acids Res* 22: 344-345.
- Poulter RTM, Goodwin TJD, Butler ML 2007. The nuclear encoded inteins of fungi. *Fungal Genet Biol* 44: 153-179.

- Ricci G, Mota FT, Wakamatsu A, Serafim RC, Borra RC, Franco M 2004. Canine paracoccidioidomycosis. *Med Mycol* 42: 379-383.
- Ridley M 2006. Conceitos de espécie e variação intra-específica. In M Ridley, *Evolução*, Artmed, Porto Alegre, p. 481-484.
- Steenbergen JN, Casadevall A 2003. The origin and maintenance of virulence for the pathogenic fungus *Cryptococcus neoformans*. *Microb and Infect* 5: 667-675.
- Steenbergen JN, Nosanchuk JD, Malliaris SD, Casadevall A 2004. Interaction of *Blastomyces dermatitidis*, *Sporothrix schenckii* and *Histoplasma capsulatum* with *Acanthamoeba castellanii*. *Infect Immun* 72: 3478-3488.
- Tavassoli A, Naumann TA, Benkovic SJ 2008. Production of cyclic proteins and peptides. In M Belfort, V Derbyshire, BL Stoddard, DW Wood, *Homing endonucleases and inteins*, Springer-Verlag Berlin, Heidelberg, p. 293-302.
- Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet Biol* 31: 21-32.
- Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D 2006. Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philos Trans R Soc Lond B Biol Sci* 361: 1947-1963.
- Teixeira MM 2008. *Tipagem molecular e evolução do gênero Paracoccidioides*. Master Thesis, Faculdade de Medicina da Universidade de Brasília, Brasília, 154 pp.
- Theodoro RC, Bagagli E, Oliveira C 2008. Phylogenetic analysis of PRP8 intein in *Paracoccidioides brasiliensis* species complex. *Fung Genet Biol* 45: 1284-1291.
- Yamashiro CT, Kane PM, Wolczyk DF, Preston RA, Stevens TH 1990. Role of vacuolar acidification in protein sorting and zymogen activation: a genetic analysis of the yeast vacuolar proton-translocating ATPase. *Mol Cell Biol* 10: 3737-3749.