American Journal of Biochemistry and Biotechnology

Original Research Paper

The Endoribonuclease Domain of IRE1 and its Substrate HAC1 are Structurally Linked Components of the Unfolded Protein Response in Fungi

Luis David Maldonado-Bonilla

Universidad del Mar, Institute of Genetics, Puerto Escondido, Mexico

Article history
Received: 16-09-2020
Revised: 14-10-2020
Accepted: 14-10-2020

Email: maldonado@zicatela.umar.mx

Abstract: The unfolded protein response sustains the folding capacity of the endoplasmic reticulum under conditions that increase the activity of the secretory pathway. In fungi, the unfolded protein response exclusively depends on the kinase/endoribonuclease IRE1, which senses the accumulation of unfolded proteins in the endoplasmic reticulum lumen and catalyzes the committed step in the unconventional splicing of HAC1 mRNA that encodes a bZIP transcription factor. The RNase domain of the IRE1 in representative fungal species was analyzed *in silico*. This domain shows high conservation of the residues that are required to dimerize, catalyze HAC1 mRNA cleavage and undergo regulation by flavonols. The predicted structure of the RNase domain in *N. crassa* IRE1 is reminiscent of the active domain of *S. cerevisiae*. Sequence analysis of the HAC1 gene in a select group of ascomycetes revealed features that suggested that there were conserved mechanisms of transcriptional and post-transcriptional regulation, especially the recognition and cleavage that is mediated by IRE1. These bioinformatic analyses revealed the close and conserved relationship between IRE1 and HAC1 as a mechanism that secures endoplasmic reticulum function in fungi.

Keywords: ER Stress, Intron, Endoribonuclease, bZIP Transcription Factor

Introduction

Biogenesis of secretory proteins and plasma membrane proteins is critical for organism development and adaptation. Proteins that are synthetized at the Endoplasmic Reticulum (ER) surface are sorted to the ER lumen to be folded and undergo post-translational modifications. Once the proteins are processed, they are transferred to the Golgi apparatus and they are then loaded into vesicles that move towards the plasma membrane to be fused. Cargo proteins will then be released and secreted or inserted into the plasma membrane (Anelli and Sitia, 2008). In fungi, the secretory pathway plays prominent roles in hyphal growth, degradation of polymers for nutrition and establishment of pathogenesis and symbiosis (Lo Presti *et al*., 2015; Steinberg, 2007). Furthermore, knowledge of this pathway in filamentous fungi can be exploited to produce secreted recombinant proteins (Nevalainen and Peterson, 2014). Unfolded proteins are accumulated at the lumen of ER when the demands of protein secretion overcomes the processing capacity of the ER, which is known as ER stress. The Unfolded Protein Response (UPR) is a surveillance system that senses the accumulation of unfolded protein and triggers the expression of ER components to rescue the ER function and homeostasis (Hollien, 2013). In the yeast *Saccharomyces cerevisiae*, the UPR depends upon activity of the Inositol-requiring enzyme 1 (Ire1p), the bZIP transcription factor Homologous to ATF-CREB1 (Hac1p) and the tRNA ligase 1 (Rlg1p/Trl1p) (Hetz, 2012). Ire1p is an ER-resident enzyme with kinase and endoribonuclease (RNase) activities that is located at the cytoplasmic face of the ER. A domain in the luminal region interacts with the chaperone BiP. When unfolded proteins accumulate in the lumen, the interaction Ire1p-BiP is disrupted and the kinase domain is autophosphorylated and activated, which enables Ire1p dimerization (Credle *et al*., 2005). Aggregates of Ire1p are detected during UPR, but the dimer is the functional enzyme (Aragón *et al*., 2009). Because of dimerization, the two RNase domains or Kinase Extension Nuclease domains (KEN) get close to each other and catalyze the endonucleolytic cleavage of the unspliced *Hac1* (*Hac1*^{u}) mRNA in two sites that are located in loops. Each loop is

constituent of a secondary structure that is located in the mRNA coding region and both secondary structures form the unconventional intron. After removal of the unconventional intron, Rlg1p joins the other two halves of the *Hac1* mRNA. This novel mRNA is known as induced *Hac1* (*Hac1*^{i}) (Cox and Walter, 1996). During normal ER functioning, *Hac1* cannot be translated because of interaction of the unconventional intron with a segment of the 5′-Untranslated Region (5'-UTR) (Rüegsegger *et al*., 2001). This unconvetional splicing process enables the translation of the Hac1p. Hac1p dimerizes and binds to the UPR Element (UPRE) in the promoters of genes encoding luminal chaperones, peptidyl-prolyl isomerases, disulfide isomerases, enzymes of phospholipid biosynthesis and other ER components (Cox *et al*., 1997). To sustain the execution of UPR, Hac1p autoregulates its own gene expression (Ogawa and Mori, 2004). Knowledge about the IRE1 pathway in fungi is focused on yeast species (Hernández-Elvira *et al*., 2018) and some models such as *Aspergillus* spp. (Mulder *et al*., 2004; Richie *et al*., 2009) and *Neurospora crassa* (Montenegro-Montero *et al*., 2015).

The PERK and ATF6-dependent pathways also regulate the UPR. However, in fungi the UPR operates only through the IRE1 pathway. IRE1 endoribonuclease activity and the *HAC1* mRNA sequence and structure have a close relationship and represent a critical point at which to induce the UPR.

Here, the structural *in silico* analysis of the IRE1 ribonuclease domain in representative fugal species from different phyla is presented and most of them are non-model species. The structure of the *HAC1* gene and transcript was also analyzed in related fungi. These novel data provide structural features that might have biological relevance and offer information about regulation of the UPR in organisms where the function of the ER and the secretory pathway are pivotal for their development and adaptive responses.

Materials and Methods

Gene and Protein Sequences

S. cerevisiae Ire1p and the human homolog of ScHac1p, X-BOX BINDING PROTEIN X (XBP1) were used as queries to retrieve the corresponding sequences of IRE1 and HAC1 from the Ensembl Fungi genome browser (https://fungi.ensembl.org/index.html) using the BLASTP algorithm (Kersey *et al*., 2018). The ID codes for each gene and protein are enlisted in Table 1 and 2.

Sequence Alignment and Identification of Functional Domains

The RNase domain (PFAM domain PF06479) sequence of IRE1 and the bZIP domain of HAC1 proteins were aligned using the MUSCLE algorithm (Edgar, 2004). These domains were automatically identified in the Ensembl Fungi browser because it is linked to UniProt resource database (http://www.uniport.org) (Consortium, 2017). Identification of domains was confirmed by searches in HMMER (http://hmmer.org/), which also locates transmembrane helices (TMs) using the Phobius program (Käll *et al*., 2004). The TMpred server was used to confirm the presence of TM (https://embnet.vital-it.ch/software/TMPRED_form.html).

Protein Modelling

The RNase domain of *Nc*IRE1 was modelled by swiss model (Waterhouse *et al*., 2018) using the structure of *Sc*Ire1p at 2.4 A° resolution as template (3LJ1) (Wiseman *et al*., 2010) because it presents the highest identity with respect to NcIRE1. The protein 3D model was visualized and edited using UCSF Chimera 1.13 (Pettersen *et al*., 2004).

Table 1: List of species and sequences of IRE1 used in this study. The ID corresponds to the identification code of Ensembl Fungi (https://fungi.ensembl.org/index.html)

Species ID	
Ascomycota	
Saccharomyces cerevisiae S288c	YHR079C
Kluyveromyces lactis NRRL Y-1140	KLLA0_D13266g
Schizosaccharomyces pombe 972	SPAC167.01
Saitoella complicata NRRL Y-17804	G7K_1322-t1
Hortaea werneckii EXF-2000	BTJ68_11496
Aureobasidium namibiae CBS 147.97	M436DRAFT_36726
Aspergillus niger An76	ABL_00408
Fusarium oxysporum f. sp. *lycopersici 4287*	FOXG_08844
Trichophyton rubrum CBS 202.88	H107_01948
Dactylellina haptotyla CBS 200.50	H072_2682
Phaeomoniella chlamydospora UCRPC4	UCRPC4_g03305
Tuber borchii Tbo3840	B9Z19DRAFT_1100132
Neurospora crassa OR74A	NCU02202
Xylona heveae TC171	L228DRAFT_243362
Basidiomycota	
Agaricus bisporus var. burnettii JB137-S8	AGABI1DRAFT_60620
Laccaria amethystina LaAM-08-1	K443DRAFT_106542
Piloderma croceum F 1598	PILCRDRAFT_316
Fibularhizoctonia sp. CBS 109695	FIBSPDRAFT_934984
Scleroderma citrinum Foug A	SCLCIDRAFT_587916
Paxillus rubicundulus Ve08.2h10	PAXRUDRAFT_822623
Puccina coronata var. *avenae*	PCANC_14392
Melampsora larici-populina 98AG31	MELLADRAFT_46493
Ustilago maydis 521	UMAG_03841
Anthracocystis flocculosa Pf-1	PFL1_06479
Mucoromycota	
Mucor circinelloides 1006PhL	HMPREF1544_01638
Choanephora cucurbitarum KUS-F28377	A0J61_02401
Blastocladiomycota	
Allomyces macrogynus ATCC38327	AMA G_00580
Catenaria anguillulae PL171	BCR44DRAFT_124952
Chytridiomycota	
Batrachochytrium dendrobatidis JAM81	BATDEDRAFT_20725
Spizellomyces punctatus DAOM BR117	SPPG_05208
Cryptomycota	
*Paramicrosporidium soccamoeba**e** KSL3*	PSACC_01985
Rozella allomycis CSF55	O9G_003659
Zoopagomycota	
Basidiobolus meristosporus CBS931.73	K493DRAFT_410752
Microsporidia	
Spraguea lophii 42_110	SLOPH_1635

Table 2: List of sequences of HAC1 of ascomycetes from different orders used in this study. The ID corresponds to the identification code of Ensembl Fungi (https://fungi.ensembl.org/index.html)

Order	Species	Gene ID
Capnodiales	*Hortaea werneckii*	BTJ68_13422
Dothideales	*Aureobasidium namibiae*	M436DRAFT_53835
Eurotriales	*Aspergillus niger*	ABL_09977
Hypocreales	*Fusarium oxysporum*	FOXG_15625
Onygenales	*Trichophyton rubrum*	H107_02842
Orbiliales	*Dactylellina haptotyla*	H072_7878
Phaeomoniellales	*Phaeomoniella chlamydospora*	UCRPC4_g04397
Pezizales	*Tuber borchii*	B9Z19DRAFT_475277
Sordariales	*Neurospora crassa*	NCU01856
Xylonales	*Xylona heveae*	L228DRAFT_249044

Prediction of RNA Secondary Structures

The sequences downstream of the bZIP coding region in selected *HAC1* mRNAs were used as queries to scan for stem–loop structures in the RNAfold server that is available in the ViennaRNA Websuite (http://rna.tbi.univie.ac.at/) (Gruber *et al*., 2008). Models with minimum free energy are presented in this report.

Results

The species that were selected for this study are heterogeneous in their lifestyles and habitats and they encompassed the current fungal phyla (Table 1). Some selected species are saprophytic such as the ascomycetes *A. niger* and *N. crassa*, the basidiomycete *A. bisporus*, the mucoromycete *M. circinelloides* and the chytridiomycete *S. punctatus*. Phytopathogens such as the ascomycete *F. oxysporum* and the basidiomycetes *U. maydis* and *P. coronata*; the halotolerant asomycete *H. werneckii*; the nematode-trapping *D. haptotyla*; the endophyte *X. heveae*; and the basidiomycete mycorrhizal fungi *L. amethystina* and *P. croceum* were included. The dermatophyte *T. rubrum* and the fish-infecting microsporidal pathogen *S. lophii* were also part of the list. The diversity of the selected organisms allows a broader understanding of IRE1 structure and function.

The RNase Domain of IRE1 is Conserved Among Fungi

IRE1 is a multidomain protein with kinase and endoribonuclease activities, its architecture is presented in (Fig. 1A). The Signal Peptide (SP) and the TM allows anchoring to the ER and membrane spanning. The wide luminal region that is located at the luminal face of the ER and the kinase and RNase domains are oriented towards the cytoplasm. The luminal region harbors a luminal domain that interacts with the chaperone BiP/GRP78 in absence of ER stress to avoid IRE1 dimerization (Zhou *et al*., 2006). Upon UPR, dimerization of IRE1 forms the active endoribonuclease that cleaves the mRNA substrates. To determine the conservation of the RNase domain and its potential functionality in fungi, the IRE1 homolog sequences that are representative of different fungal species were retrieved from the Ensembl Fungi genome browser and aligned using the MUSCLE algorithm. The result of the alignment is presented in Fig. 1B. The RNase domain, which is oriented towards the cytoplasm, is a trilobal structure composed of eight α-helices (Lee *et al*., 2008). Residues that are involved in catalysis are located at the first and fourth helices (Lee *et al*., 2008) and most of the residues are conserved in fungi, regardless of the phylum to which they belong (Fig. 1B). In yeast, this loop, which is formed by a leucine, basic residues and tyrosines, is required to confer specific recognition of the *HAC1* mRNA stem–loop structures. Basic residues are mostly conserved in fungi and in the second tyrosine that correspond to Tyr1043 in *S. cerevisiae*. The role of Tyr1043 in the endoribonuclease activity has been experimentally confirmed (Lee *et al*., 2008). Tyr1041 is conserved in yeasts such as *S. cerevisiae*, *K. lactis*, *S. pombe* and *S. complicata*, but this residue is variable in other selected ascomycetes (Fig. 1B). Species of Basidiomycota, Mucoromycota and Blastocladiomycota presented an aromatic residue in this position. Conservation of the residues that are involved in the catalysis and in the loop suggests that the IRE1 homologs that are analyzed in this study may be functional endoribonucleases that operate in a similar manner as ScIre1p. Flavonols such as quercetin increase the endoribonuclease activity of ScIre1p by electrostatic interactions with residues of the first helix at the dimerization interface in the RNase domain (Wiseman *et al*., 2010). Residues that are involved in recognition of quercetin are conserved in fungi (Fig. 1B). Lys992 is necessary to cleave the mRNA substrate independent of the presence of flavonols (Lee *et al*., 2008). An aspartate in the RNAse domain is most conserved compared to lysine and in some cases, such as for *S. complicata* or *A. bisporus*, the aspartate is replaced by glutamate.

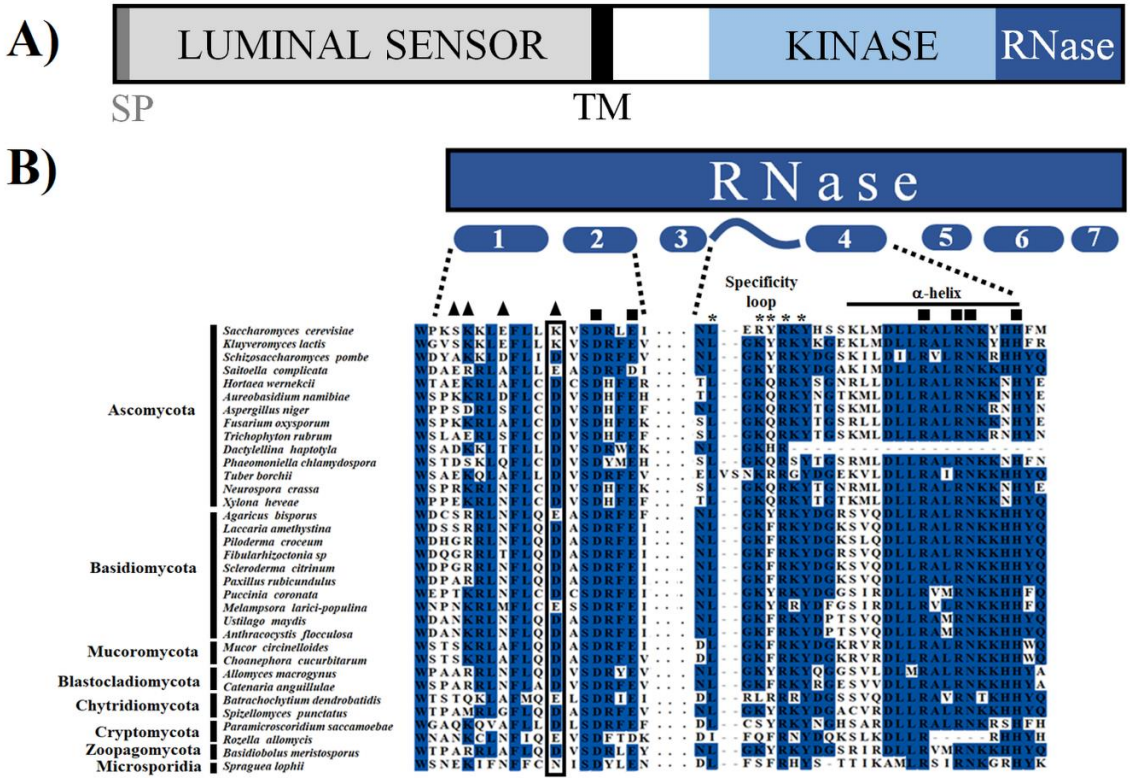

Fig. 1: Structure of the IRE1 protein. (A) Signal peptide (SP) at the N terminus precedes the luminal region of the ER and the transmembrane helix (TM), the KINASE and RNase domain are oriented towards the cytoplasm. (B) Alignment of the RNase domain of fungal species. The 8 α-helices are represented by Blue bars with white numbers. The phylum is indicated in the left of the species list. Conserved residues are shaded in blue. Black squares, asterisks and black triangles indicate residues of *Sp*Ire1p that are involved in catalysis, specificity and activation by flavonols, respectively. The K992 of SpIre1p and the other residues of this position are framed by a rectangle

Predicted Structural Features of the NcIRE1 RNAse Domain

The *N. crassa* RNase domain was modelled using the *Sc*Ire1p (3LJ1) model to define whether conservation in the RNase domain implies a similar tertiary structure (Wiseman *et al*., 2010). The *N. crassa* domain was selected because there has been experimental evidence of unconventional *HAC1* splicing (Montenegro-Montero *et al*., 2015). Analysis of the *Nc*IRE1 RNAse domain sequence in the secondary structure prediction servers such as Psipred or Jpred revealed that the variations in the residues do not affect the formation of α-helices (data not shown). The predicted model as a potentially functional dimer, this model and residues that are potentially required for catalysis are shown in (Fig. 2A). In particular, the residues Asn1077 and Asp1081 constitute the first α-helix of the domain at the interface of dimerization and they are involved in flavonol activity and regulation. The specificity loop is located at the top of the structure. Overlap of the reported and predicted *Sc*Ire1p RNase domain and predicted *Nc*IRE1 RNase domain reveals quite similar structures (Fig. 2B), although the slight differences in residues in the first α-helix do not compromise the structure, but they may suggest differences in the regulation of enzymatic activity among IRE1 homologs.

The Structure of HAC1 Genes in Pezizomycotina Suggests Common Post-Transcriptional Regulation

Unconventional splicing of *HAC1* mRNA is reported in fungal yeast models (Hernández-Elvira *et al*., 2018) and filamentous models such as *A. niger*, *A. fumigatus* and *N. crassa* (Mulder *et al*., 2004; Richie *et al*., 2009) (Montenegro-Montero *et al*., 2015). Conservation of the RNase domain that is presented above suggests that IRE1 is functional and cleavage of *HAC1* mRNA at specific sites is a conserved mechanism in fungi. The *HAC1* gene structure in species from different orders of the subphylum Pezizomycotina was analyzed to obtain insight about potential mRNA processing. To avoid a sequence bias, one species from ten different orders of Pezizomycotina were selected for this analysis. Proteins were found using the BLASTP algorithm in Ensembl Fungi, with *Nc*HAC1 as a query. Identification of a HAC1 homolog often failed in other phyla because the

best hits showed remarkably high e values or a lack of conservation in the bZIP domain. The structural *HAC1* gene and 1000 bp upstream of the ATG start codon are shown in Fig. 3. In *A. niger*, HACA/HAC1 recognizes the consensus UPR elements 5′-CAN(G/A)NTG(T/G)CCT-3′ and the 10 *mer* variant 5′-CACTGTCCT-3′ (Mulder *et al*., 2006). These elements and similar sequences were identified in the potential *HAC1* gene promoters, suggesting a positive feedback loop (Table 3). In *A. niger*, upstream Open Reading Frames (uORF) have been identified in the 5′-UTR of *HAC1*/*HACA* mRNA and it inhibits translation of the major open reading frame. When UPR is triggered, an alternative transcription start site produces a mRNA with a shorter 5′-UTR that exempts the uORF, which facilitates translation (Mulder *et al*., 2004). Putative uORF were searched along with 5′-UTRs. Each analyzed gene presented at least one ORF. The average length of the uORFs was 56 amino acids (aa), the minimum length was a 19-aa uORF in *T. rubrum HAC1* and the maximum length of an uORF was 137 aa, which was found in the *D. haptotyla HAC1* gene (Fig. 3). None of the amino acid sequences are homologous with protein sequences or annotated as an independent ORFs in the corresponding genomes. Potential stem-loop structures were identified in the *HAC1* mRNAs that are downstream of the region encoding for the bZIP domain (Fig. 3). A sequence encoding at least one TM was also predicted downstream of the stem–loop structures (Fig. 3).

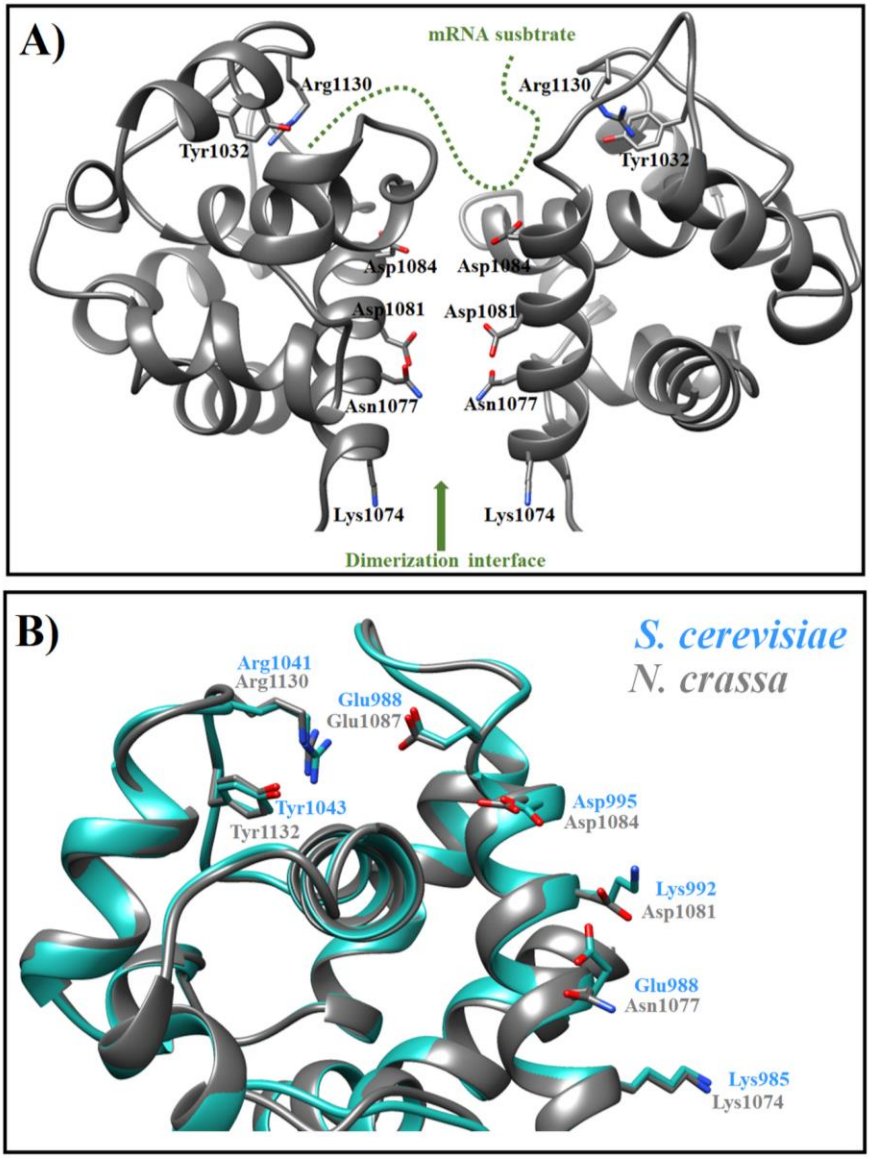

Fig. 2: Predicted tertiary structure of the dimerized RNase domain of *Nc*IRE1. (A) The first helix required of dimerization interface and residues involved in recognition of substrate are pointed out. (B) Superimposed structures reveal conservation of the RNase domain in *Sc*Ire1p and *Nc*IRE1

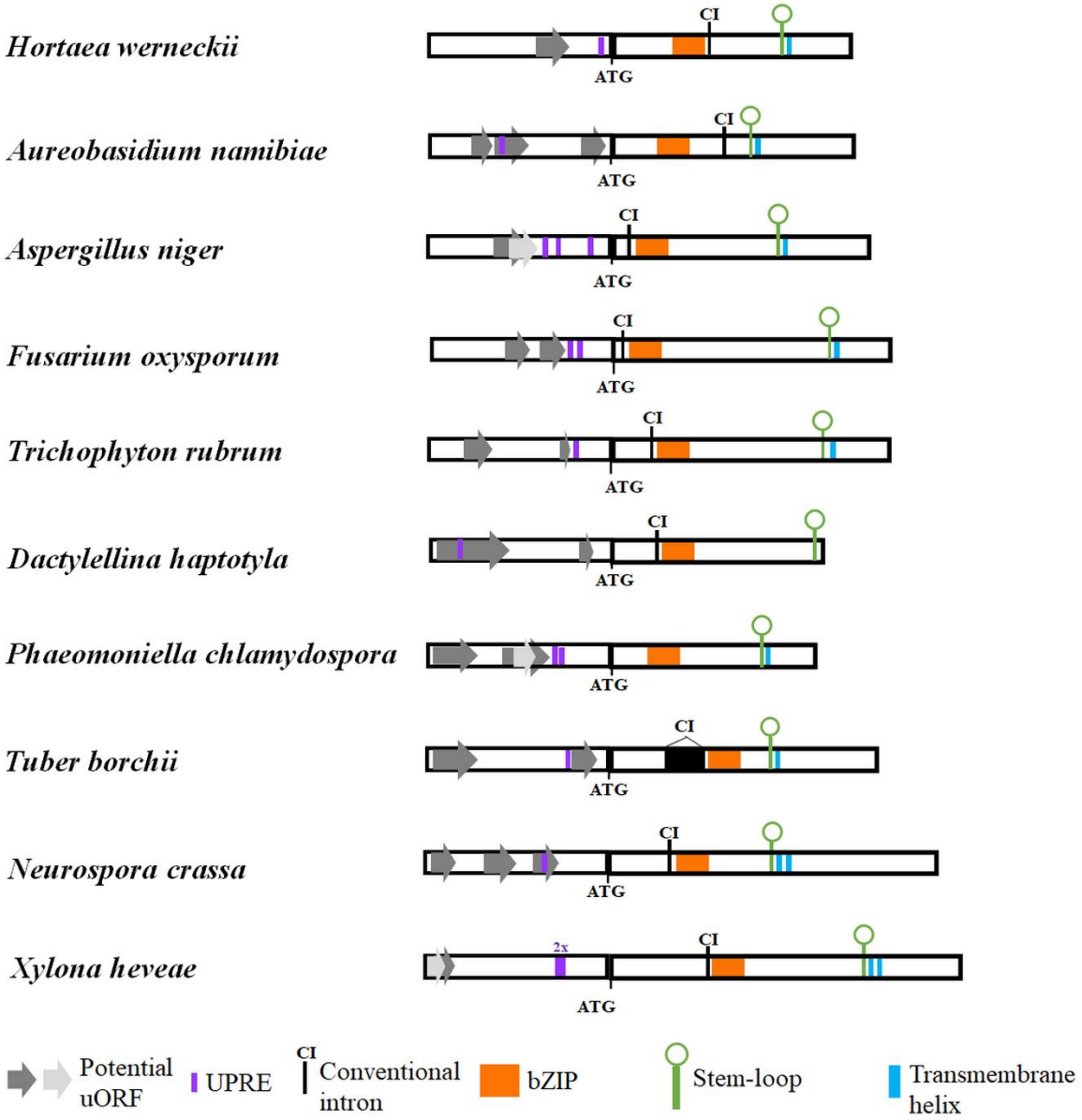

Fig. 3: Structure of the *HAC1* gene in selected ascomycetes. Potential uORF and UPRE sequences (purple boxes) were identified in 1000 bp upstream ATG start codon. Two-toned grey arrows were used because some of the uORF are overlapped. The coding sequence comprises the bZIP domain (orange box), conventional intron and a stem-loop sequence that is potentially recognized as an unconventional intron and a transmembrane helix. Note that the transmembrane helix coding sequence is downstream of the stem–loop

Comparison of the predicted RNA stem–loop structures reveals conservation in two motifs that are potentially recognized by IRE1 (Fig. 4). *A. niger* and *N. crassa* sequences were experimentally confirmed, which allows the identification of the potential IRE1 cleavage sites. The length and constitution of the sequence between the potential cleavage sites is variable (Fig. 4). The stem–loop structures are presented in Fig. 5. The structures share no similarities among them, but the nucleotides G↓C at the cleavage sites are mostly localized into loops. These RNA secondary structures might function as unconventional introns that are cleaved by IRE1 during the UPR.

Structure of the HAC1 Proteins in Pezizomycotina

A TM is predicted in the HAC1^{u} proteins, except in *P. chlamydospora* (Fig. 6A). There was no similarity in the TM sequence, but its identification suggests that,

similar to the model plant Arabidopsis, the unspliced mRNA produces a version of the protein that is anchored to the ER membrane (Iwata and Koizumi, 2005).

The bZIP domain is a long helix that is divided by a basic motif and it interacts with DNA and a leucine zipper that mediates dimerization (Yin *et al*., 2017). The HAC1 protein bZIP domain in the species Pezizomycotina was aligned with the human homolog XBP1(Yoshida *et al*., 2001) and the human bZIP protein FosB. The bZIP domain of FosB is crystalized and it functions as a reference to locate amino acids from a basic motif and the leucine zipper (Yin *et al*., 2017). The basic motif is highly conserved, but the leucine zipper is more divergent (Fig. 6B). A motif of charged amino acids, mostly basic, which were adjacent to the basic motif was highly conserved in the species Pezizomycotina in this study and it might influence the HAC1[i] transcription factor activity.

Table 3: Putative *An*HACA/HAC1-binding consensus sites found in 1000 bp upstream of the start codon of *HAC1* genes. The asterisk denotes sites similar to the consensus of *An*HACA/HAC1 previously reported[27]. The position of the sequence regarding the start codon is shown in the far right column

Species	Element	Position respect to start codon
Hortaea werneckii	5’-CACCTGTCCT-3’	70-61
Aureobasidium namibiae	5’-CAAGAGGGCCT-3’*	569-559
Aspergillus niger	5’-CACGTTGGCCT-3’	359-349
	5’-CACCTGTCCT-3’	258-249
	5’-CACCGTGGCCT-3’	159-149
Fusarium oxysporum	5’-CAACGTGTCCT-3’	385-375
	5’-CACCTGTCCT-3’	278-268
Trichophyton rubrum	5’-CACGGTGGCCT-3’	269-259
Dactylellina haptotyla	5’CAGCGTGTCCT-3’*	756-746
Phaeomoniella chlamydospora	5’-CATCCATTCCT-3’*	247-237
	5’-CACTCTGTCCT-3’*	231-221
Tuber borchii	5’-TAGTATGGCCT-3’*	528-518
Neurospora crassa	5’-CAGCCTGGCCT-3’*	406-396
Xylona heveae	5’-CAATCTGTCCT-3’*	288-278
	5’-CTCTCTGTCCT-3’*	277-267

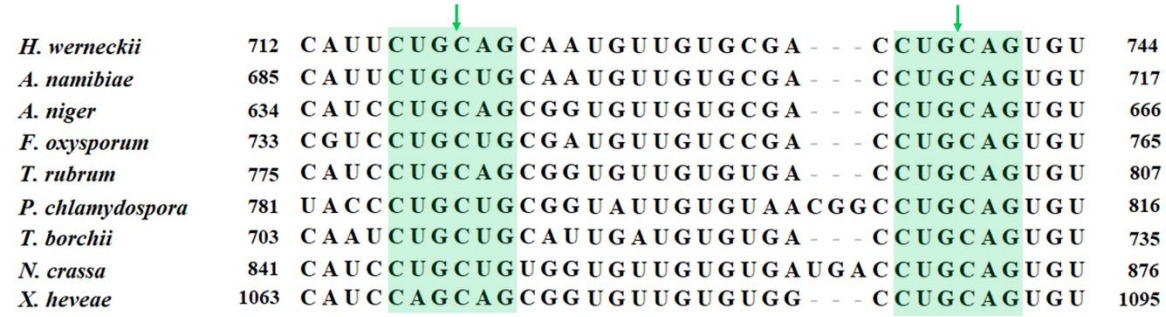

Fig. 4: Alignment of the RNA sequences that are potentially recognized by the RNase domain of IRE1 of ascomycetes. Two highly conserved motifs are shaded in green. The arrow indicates the potential G↓C cleavage site

Discussion

The endoribonuclease activity of IRE1 depends on the RNase/KEN domain that is adjacent to the kinase domain, which is located at the C terminus of the protein (Fig. 1A). This RNase domain is composed of eight α-helices and the residues of the first and fourth helices mediate the cleavage of the RNA substrates when they dimerize (Lee *et al*., 2008). The loop is located between the third and fourth α-helix in *Sc*Ire1p and it recognizes the two loop structures in *Hac1* mRNA; both of these loops will be subsequently cleaved at a G↓C site. RNase domain residues that are involved in catalysis and the specificity loop residues are conserved in fungi (Fig. 1B). The Lys992 in *Sc*Ire1p is crucial to cleave the *Hac1* mRNA and this residue is indicated by the black rectangle in Fig. 1B. Among the RNase domain sequences that were analyzed in this research, K992 is only conserved in the yeast *K. lactis*. This residue is substituted by acid amino acids in the rest of the sequences, including *S. pombe*, which lacks a HAC1 homolog (Guydosh *et al*., 2017). The exception is the *S. lophii* RNase domain, which has an asparagine in this position. The substitution K992D in *Sc*Ire1p decreases the endoribonuclease activity (Lee *et al*., 2008), suggesting that the basic nature of lysine is required for

catalysis. Paradoxically, the acid residue is more conserved in fungi. K992 in *Sc*Ire1p also forms hydrogen bonds with E988 (Wiseman *et al*., 2010). This residue is highly divergent in the analyzed sequences and such divergence might be necessary to set electrostatic interactions with other residues. It will be worth comparing the *Sc*Ire1p RNase domain activity against a RNase domain that harbors an acid residue in this position because it will define the function of this substitution. The mutations K985A and K992L in the *Sc*Ire1p RNase domain attenuate endoribonuclease activity and activation by flavonols such as quercetin. Quercetin is placed at the dimerization interface of the *Sc*Ire1p RNase domain and may boost its activity by clinching the dimerization of the RNase domains. Interaction with flavonols might have biological implications. Favonols are secondary metabolites that are broadly produced by plants and they have roles in the interaction between pathogens and symbionts (Buer *et al*., 2010). A group of RNase domains from phytopathogens such as *F. oxysporum* or mycorrhizae such as *L. amethystina* were analyzed and showed conservation in the residues in the first α-helix that is involved in flavonol binding. The RNase domain of IRE1 in these organisms might be altered by flavonols that are produced by plants during pathogenic or beneficial interactions. This potential regulation of the endoribonuclease activity mediated by flavonols might affect the fungal protein secretion that is necessary to set the interaction with plants.

To determine whether the α-helices are distinctive to the RNase domain of IRE1, the corresponding *N. crassa* domain was modelled (Fig. 2A). This protein was selected because there is experimental evidence of unconventional splicing of *NcHAC1* mRNA; thus, *N. crassa* has a functional IRE1 (Montenegro-Montero *et al*., 2015). The predicted structure conserves the α-helices that are involved in catalysis, dimerization and interaction with flavonols as well as and the specificity loop that recognizes the substrate. Figure 2B illustrates the overall structure of the RNase domain of *Nc*IRE1 compared to the respective domain of *Sc*IRE1p. Regardless of the differences in the residues of the first α-helix, this secondary structure may be also formed in *N. crassa* and constitutes part of a potentially active enzyme. The *Nc*IRE1 specificity loop harbors the residue Y1132, which is equivalent to the Y1043 of *Sc*Ire1p (Lee *et al*., 2008). The mutation Y1043A in the *Sc*Ire1p specificity loop dramatically decreases the endoribonuclease activity (Lee *et al*., 2008) and therefore, the conserved Y1132 in *Nc*IRE1 may be critical to catalysis.

The ascomycetes *HAC1* gene structure from different orders was analyzed *in silico* to determine whether their mRNAs are substrates of IRE1. The ascomycetes that were selected belonged to the Pezizomycotina subphylum and included *A. niger* and *N. crassa*. Schemes of the *HAC1* genes illustrates the conservation of the sequence encoding bZIP domain (Fig. 3), a TM at the C terminus and a stem–loop sequence in the nearby of the TM-coding sequences. Analysis of sequences upstream of the start codon detected a UPRE consensus in *A. niger* and suggests hypothetical positive feedback that potentiates *HAC1* expression upon UPR (Ogawa and Mori, 2004). The uORFs encoded in the 5′-UTR attracts ribosomes as a mechanism to attenuate the translation of the major ORF (Hood *et al*., 2009). In *A. niger*, an uORF encoding 44 amino acids is excluded from the *HACA/HAC1* mRNA upon ER stress by truncation of the 5′-UTR (Mulder *et al*., 2004). A similar condition is reported in *A. fumigatus* (Richie *et al*., 2009). Inspection of the 1000 bps that are upstream from the start codon identified a short ORF with no homology to the fungal polypeptides that might function as uORFs, but they have to be experimentally validated, especially the ORFs near the start codon.

Sequences that are reminiscent of the unconventional intron were identified in the group of predicted *HAC1* mRNAs (Fig. 4). Two 6-nt motifs are highly conserved in the transcripts, except for *D. haptotyla*, which has stem–loop structures without the motifs. It was excluded from Fig. 4. Both motifs harbor the G↓C site that is cleaved by IRE1. The composition and sequence length between motifs are variable. The presence of the G↓C sites inside the loops is the factor that enables cleavage (Lee *et al*., 2008). Regulated IRE1-Dependent Decay of mRNA (RIDD) is a pathway of mRNA degradation that is conserved in eukaryotes and it prevents the protein load for maintaining the protein folding demand of the ER (Maurel *et al*., 2014). Substrates of RIDD are cleaved by IRE1, which recognizes a sequence that is similar to the motifs of the *HAC1* unconventional intron. However, IRE1 activity might not be specific to loops because the human IRE1α cleaves miRNA precursors at G↓C sites that are localized in stems or double-stranded RNA (dsRNA) segments (Upton *et al*., 2012).

Figure 5 shows the secondary structures that are predicted in the potential unconventional introns. Stem–loop structures are formed in those mRNAs and the potential cleavage sites are mostly located inside loops. They are likely recognized and cleaved by IRE1 enzymes and its RNase domain conservation and functionality were previously discussed. The first G↓C site in *H. werneckii* and the two sites in *P. chlamydospora* are examples of potential cleavage sites that are situated in dsRNA (Fig. 5). The affinity of fungal IRE1 for the G↓C sites in *HAC1* dsRNA segments is unknown, but the activity of human IRE1α towards the miRNA precursors might indicate that loops harboring the G↓C sites are not the only conditional factors for *HAC1* mRNA cleavage.

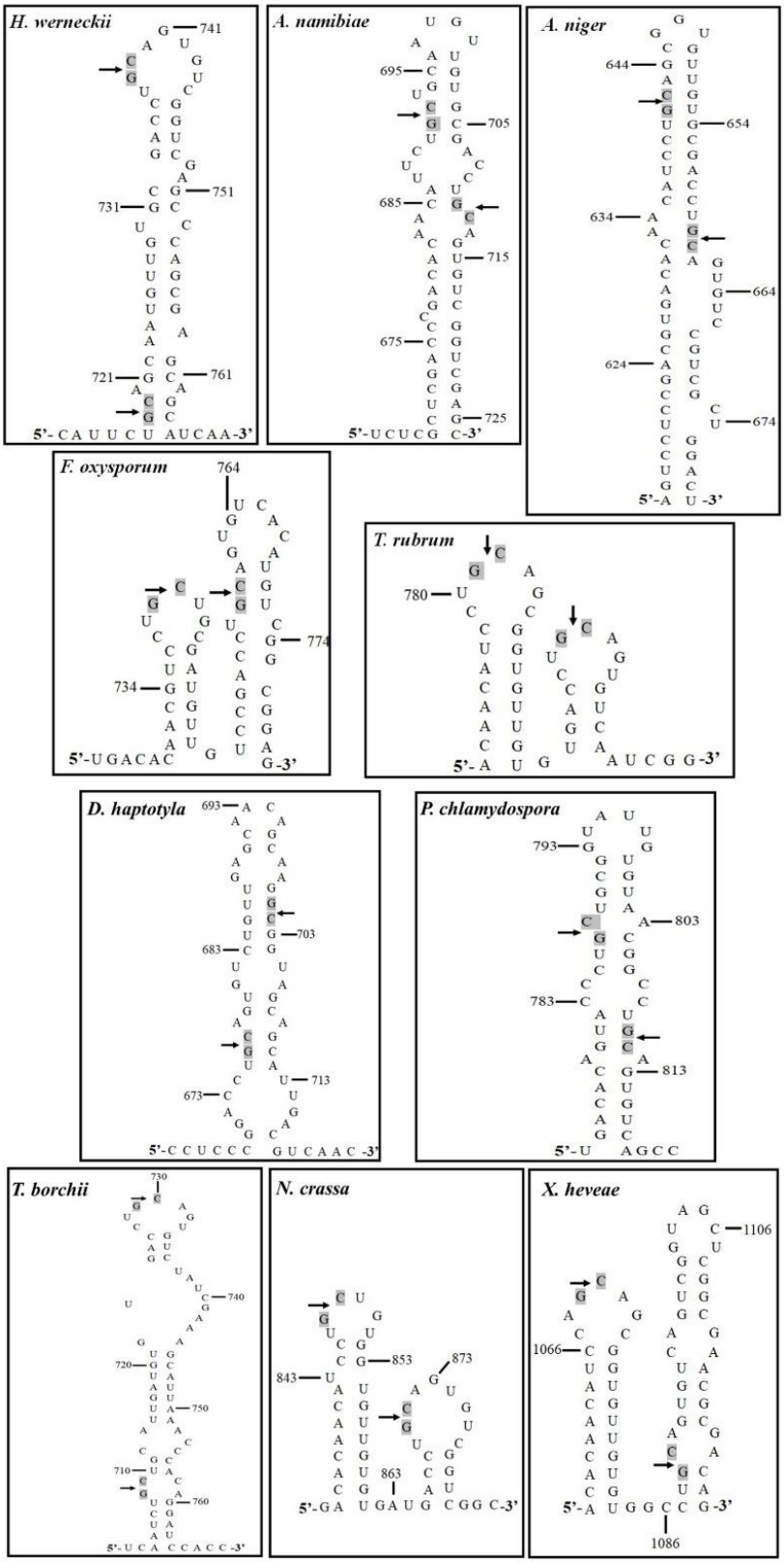

Fig. 5: Predicted stem–loop structures of *HAC1* mRNAs in ascomycetes. The arrows show the potential G↓C cleavage site, which are shaded in grey. Numbers indicate the position of the nucleotide in the mature mRNA

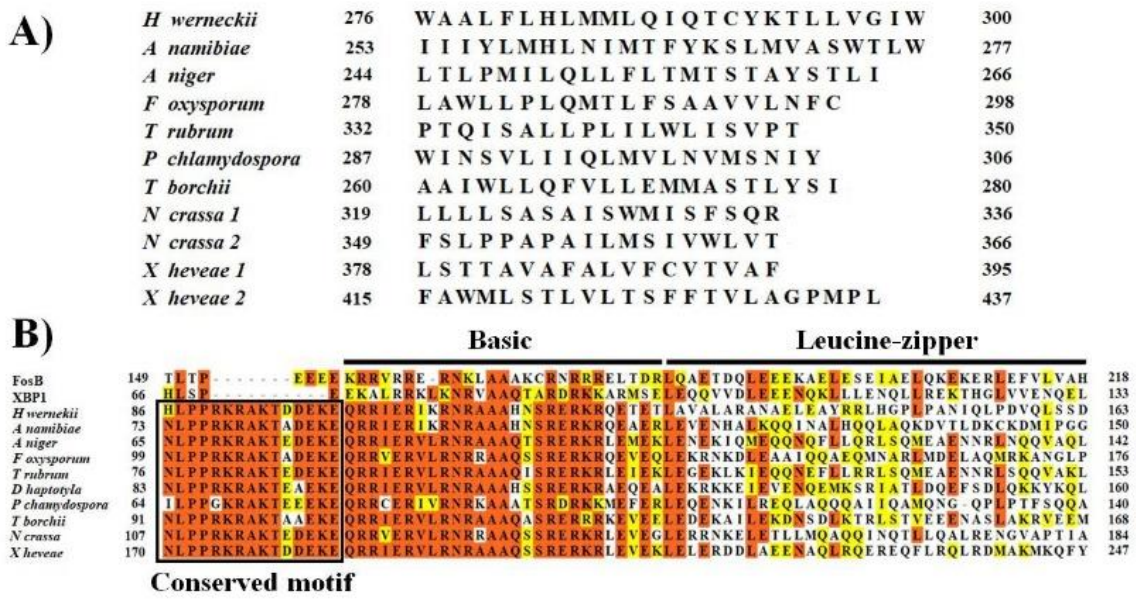

Fig. 6: Features of HAC1 proteins of ascomycetes. (A) Predicted TM with no similarity among them. Two TM in tandem were identified in *Nc*HAC1 and *Xh*HAC1. (B) Alignment of the bZIP domains of the human FOSB and Xbp1 proteins and the HAC1 of ascomycetes. Basic and leucine zipper motifs of the bZIP are marked. The black square harbors the conserved motif of ascomycetes. Identical residues are shaded in orange and similar residues are shaded in yellow. Numbers indicate the position of the residue

The putative TM that was identified in the selected group of HAC1 proteins suggests that upon ER homeostasis, inactive transcription factors accumulate in the ER membrane to avoid UPR-related gene expression, which is similar to the transcription factor ATF6 in mammals (Hollien, 2013). The ER-resident ATF6 is translocated to undergo proteolysis in the Golgi apparatus upon UPR. After lysis of its TM, ATF6 becomes active and translocates to the nucleus to activate transcription of UPR-responsive genes. A mechanism that is similar to the factors bZIP17/28 of the plant Arabidopsis is reported (Kim *et al*., 2018). Fungi lack an ATF6 homolog and HAC1 proteolysis has not been reported. Thus, unconventional splicing of the transcript HAC1^{u} that causes a frameshift might be the mechanism that generates the active HAC1^{i} and conservation of the components of this mechanism have been identified in the present report.

No similarities in the sequence of the putative TMs were identified in the selected group of HAC1 proteins and two TMs were detected in the HAC1 proteins of *N. crassa* and *X. heveae* (Fig. 6A). This finding might be molecularly relevant because the TMs are located at the C terminus and the corresponding TM-coding sequence is always downstream of the IRE1-cleavege sites. The cleavage and ligation change the open reading frame and then, the reading of the TM-coding sequence is avoided. The unrelated TMs that are found in HAC1 proteins suggest that mutations that accumulated during the course of *HAC1* gene evolution in independent species enables the diversification of TMs, but a selection pressure drove the secondary structure and position of the IRE1-cleavege sites.

HAC1 proteins of Pezizomycotina have a unique motif that is adjacent to the bZIP domain (Fig. 6B). The *Sc*Hac1^{i} binds to the UPRE-1 (5′-GGACAGCGTGTC-3′) and the UPRE-2 (5′TACGTG-3′) and such binding is influenced by the residues that are adjacent to the basic motif of the bZIP domain (Fordyce *et al*., 2012). However, heterodimerization with other bZIP factors such as Gcn4p also affect the *in vivo* recognition of targets. The homolog of HAC1 in humans, XBP1, binds to diverse sequences such as the UPRE (5′-GACGTG[G]-3′), the CCACG-box and the ACTG-core (Acosta-Alvear *et al*., 2007). Heterodimerization, as in the case of XBP1 and the neighboring residues to the bZIP domain might impact DNA binding. Therefore, variations in the composition of UPRE in the UPR-responsive genes might coevolved with the DNA binding activity and heterodimerization of HAC1.

Conclusion

The bioinformatic analyses of selected fungal proteins and genes showed the catalytic domain of IRE1 and the potential mRNA substrate are the evolutionary conserved components of the UPR that operate similarly as *Sc*Ire1p.

The *cis*-acting elements in promoters and potential uORFs identified in *HAC1* genes of selected ascomycetes are hints of conserved feedback of transcriptional regulation and translational control. The conservation of the potential cleavage motifs found within secondary structures of *HAC1* mRNAs might imply the presence of unconventional introns and the adjacent downstream sequence encoding TM suggests that the cleavage of *HAC1* mRNA mediated by IRE1 is a regulatory mechanism that operate to ameliorate the ER function in the UPR in fungi, in which the secretory pathway plays prominent roles in development and communication with other organisms.

Acknowledgment

This work was supported by the Program *Cátedras CONACYT* from the National Council of Science, Mexico (*CONACYT*), Grant No. 538.

Ethics

The author declares no ethical issues may arise by this publication.

References

Acosta-Alvear, D., Zhou, Y., Blais, A., Tsikitis, M., Lents, N. H., Arias, C., … & Dynlacht, B. D. (2007). XBP1 Controls Diverse Cell Type- and Condition-Specific Transcriptional Regulatory Networks. Molecular Cell, 27(1), 53–66.

Anelli, T., & Sitia, R. (2008). Protein quality control in the early secretory pathway. EMBO Journal, 27(2), 315–327.

Aragón, T., van Anken, E., Pincus, D., Serafimova, I. M., Korennykh, A. V, Rubio, C. A, & Walter, P. (2009). Messenger RNA targeting to endoplasmic reticulum stress signalling sites. Nature, 457(7230), 736–740.

Buer, C. S., Imin, N., & Djordjevic, M. A. (2010). Flavonoids: New roles for old molecules. Journal of Integrative Plant Biology, 52(1), 98–111.

Consortium, T. U. (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Research, 45(D1), D158–D169.

Cox, J S, & Walter, P. (1996). A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. Cell, 87(3), 391–404.

Cox, Jeffery S., Chapman, R. E., & Walter, P. (1997). The unfolded protein response coordinates the production of endoplasmic reticulum protein and endoplasmic reticulum membrane. Molecular Biology of the Cell, 8(9), 1805–1814.

Credle, J. J., Finer-Moore, J. S., Papa, F. R., Stroud, R. M., & Walter, P. (2005). On the mechanism of sensing unfolded protein in the endoplasmic reticulum. Proceedings of the National Academy of Sciences of the United States of America, 102(52), 18773–18784.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797.

Fordyce, P. M., Pincus, D., Kimmig, P., Nelson, C. S., El-Samad, H., Walter, P., & DeRisi, J. L. (2012). Basic leucine zipper transcription factor Hac1 binds DNA in two distinct modes as revealed by microfluidic analyses. Proceedings of the National Academy of Sciences of the United States of America, 109(45).

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The Vienna RNA websuite. Nucleic Acids Research, 36(Web Server issue), 70–74.

Guydosh, N. R., Kimmig, P., Walter, P., & Green, R. (2017). Regulated Ire1-dependent mRNA decay requires no-go mRNA degradation to maintain endoplasmic reticulum homeostasis in S. pombe. ELife, 6, e29216.

Hernández-Elvira, M., Torres-Quiroz, F., Escamilla-Ayala, A., Domínguez-Martin, E., Escalante, R., Kawasaki, L., … & Coria, R. (2018). The Unfolded Protein Response Pathway in the Yeast Kluyveromyces lactis. A Comparative View among Yeast Species. Cells, 7(8), 106.

Hetz, C. (2012). The unfolded protein response: controlling cell fate decisions under ER stress and beyond. Nature Reviews Molecular Cell Biology, 13(2), 89–102.

Hollien, J. (2013). Evolution of the unfolded protein response. Biochimica et Biophysica Acta - Molecular Cell Research, 1833(11), 2458–2463.

Hood, H. M., Neafsey, D. E., Galagan, J., & Sachs, M. S. (2009). Evolutionary Roles of Upstream Open Reading Frames in Mediating Gene Regulation in Fungi. Annual Review of Microbiology, 63(1), 385–409.

Iwata, Y., & Koizumi, N. (2005). An Arabidopsis transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. Proceedings of the National Academy of Sciences of the United States of America, 102(14), 5280–5285.

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology, 338(5), 1027–1036.

Kersey, P. J., Allen, J. E., Allot, A., Barba, M., Boddu, S., Bolt, B. J., … Yates, A. (2018). Ensembl Genomes 2018: An integrated omics infrastructure for non-vertebrate species. Nucleic Acids Research, 46(D1), D802–D808.

Kim, J. S., Yamaguchi-Shinozaki, K., & Shinozaki, K. (2018). ER-Anchored transcription factors bZIP17 and bZIP28 regulate root elongation. Plant Physiology, 176(3), 2221–2230.

Lee, K. P. K., Dey, M., Neculai, D., Cao, C., Dever, T. E., & Sicheri, F. (2008). Structure of the Dual Enzyme Ire1 Reveals the Basis for Catalysis and Regulation in Nonconventional RNA Splicing. Cell, 132(1), 89–100.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., … & Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. Annual Review of Plant Biology, 66(1), 513–545.

Maurel, M., Chevet, E., Tavernier, J., & Gerlo, S. (2014). Getting RIDD of RNA: IRE1 in cell fate regulation. Trends in Biochemical Sciences, 39(5), 245–254.

Montenegro-Montero, A., Goity, A., & Larrondo, L. F. (2015). The bZIP transcription factor HAC-1 is involved in the unfolded protein response and is necessary for growth on cellulose in Neurospora crassa. PLoS ONE, 10(7), e0131415.

Mulder, H. J., Saloheimo, M., Penttilä, M., & Madrid, S. M. (2004). The transcription factor HACA mediates the unfolded protein response in Aspergillus niger and up-regulates its own transcription. Molecular Genetics and Genomics, 271(2), 130–140.

Mulder, Harm J., Nikolaev, I., & Madrid, S. M. (2006). HACA, the transcriptional activator of the unfolded protein response (UPR) in Aspergillus niger, binds to partly palindromic UPR elements of the consensus sequence 5′-CAN(G/A)NTGT/GCCT-3′. Fungal Genetics and Biology, 43(8), 560–572.

Nevalainen, H., & Peterson, R. (2014). Making recombinant proteins in filamentous fungi- Are we expecting too much? Frontiers in Microbiology, 5(FEB), 75.

Ogawa, N., & Mori, K. (2004). Autoregulation of the HAC1 gene is required for sustained activation of the yeast unfolded protein response. Genes to Cells, 9(2), 95–104.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera — A Visualization System for Exploratory Research and Analysis. Journal of Computational Chemistry, 25(13), 1605–1612.

Richie, D. L., Hartl, L., Aimanianda, V., Winters, M. S., Fuller, K. K., Miley, M. D., … & Askew, D. S. (2009). A role for the unfolded protein response (UPR) in virulence and antifungal susceptibility in Aspergillus fumigatus. PLoS Pathogens, 5(1), e1000258.

Rüegsegger, U., Leber, J. H., & Walter, P. (2001). Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell, 107(1), 103–114.

Steinberg, G. (2007). Hyphal growth: A tale of motors, lipids and the spitzenkörper. Eukaryotic Cell, 6(3), 351–360.

Upton, J.-P., Wang, L., Han, D., Wang, E. S., Huskey, N. E., Lim, F. R., … & Oakes, S. A. (2012). IRE1alpha cleaves select microRNAs During ER Stress to derepress translation of proapoptotic Caspase-2. Science, 338(6108), 818–822.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research, 46.

Wiseman, R. L., Zhang, Y., Lee, K. P. K., Harding, H. P., Haynes, C. M., Price, J., … & Ron, D. (2010). Flavonol Activation Defines an Unanticipated Ligand-Binding Site in the Kinase-RNase Domain of IRE1. Molecular Cell, 38(2), 291–304.

Yin, Z., Machius, M., Nestler, E. J., & Rudenko, G. (2017). Activator protein-1: Redox switch controlling structure and DNA-binding. Nucleic Acids Research, 45(19), 11425–11436.

Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., & Mori, K. (2001). XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell, 107(7), 881–891.

Zhou, J., Liu, C. Y., Back, S. H., Clark, R. L., Peisach, D., Xu, Z., & Kaufman, R. J. (2006). The crystal structure of human IRE1 luminal domain reveals a conserved dimerization interface required for activation of the unfolded protein response. Proceedings of the National Academy of Sciences of the United States of America, 103(39), 14343–14348.