Biochemical and Biophysical Research Communications 438 (2013) 526–532

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Characterization of bud emergence 46 (BEM46) protein: Sequence, structural, phylogenetic and subcellular localization analyses

Abhishek Kumar, Krisztina Kollath-Leiß, Frank Kempken *

Abteilung für Botanik mit Schwerpunkt Genetik und Molekularbiologie, Botanisches Institut und Botanischer Garten, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

ARTICLE INFO

Article history:
Received 17 July 2013
Available online 31 July 2013

Keywords:
Bud emergence 46 (*bem46*)
Evolution
Fungi
Indels
Endoplasmic retention signal
eGFP-fusion
α/β-hydrolase

ABSTRACT

The bud emergence 46 (BEM46) protein from *Neurospora crassa* belongs to the α/β-hydrolase superfamily. Recently, we have reported that the BEM46 protein is localized in the perinuclear ER and also forms spots close by the plasma membrane. The protein appears to be required for cell type-specific polarity formation in *N. crassa*. Furthermore, initial studies suggested that the BEM46 amino acid sequence is conserved in eukaryotes and is considered to be one of the widespread conserved "known unknown" eukaryotic genes. This warrants for a comprehensive phylogenetic analysis of this superfamily to unravel origin and molecular evolution of these genes in different eukaryotes. Herein, we observe that all eukaryotes have at least a single copy of a *bem46* ortholog. Upon scanning of these proteins in various genomes, we find that there are expansions leading into several paralogs in vertebrates. Usingcomparative genomic analyses, we identified insertion/deletions (indels) in the conserved domain of BEM46 protein, which allow to differentiate fungal classes such as ascomycetes from basidiomycetes. We also find that exonic indels are able to differentiate BEM46 homologs of different eukaryotic lineage. Furthermore, we unravel that BEM46 protein from *N. crassa* possess a novel endoplasmic-retention signal (PEKK) using GFP-fusion tagging experiments. We propose that three residues namely a serine 188S, a histidine 292H and an aspartic acid 262D are most critical residues, forming a catalytic triad in BEM46 protein from *N. crassa*. We carried out a comprehensive study on *bem46* genes from a molecular evolution perspective with combination of functional analyses. The evolutionary history of BEM46 proteins is characterized by exonic indels in lineage specific manner.

1. Introduction

Bud emergence 46 (*bem46*) genes encode proteins belonging to the BEM46 family within the α/β-hydrolase superfamily. This α/β-hydrolase superfamily possesses a typical α/β-hydrolase domain, which is characterized by a β-sheet core of five to eight strands connected by α-helices to form a α/β/α sandwich [1,2]. Several hydrolytic enzymes share the common α/β-hydrolase domain with a wide array of catalytic functions with different phylogenetic history [2,3]. In fact, there are over 30,000 manually annotated members of this domain in the ESTHER database [1]. However, the functional characterization of BEM46 family members remainslargely unknown with no information of their hydrolase activity and substrate binding. The *S. pombe bem46* gene was originally described as a temperature-sensitive suppressor of *S. cerevisiae bem1* and *bud5* double mutant [4]. Mutations in *bem1* and *bud5* show defects in cell polarization and establishment of non-random budding patterns [5,6]. BUD5 is a GTP–GDP exchange factor (GEF) for a BUD1 Ras-like small G protein and is necessary for bud site selection [7]. BEM1 is a scaffold protein binding to BUD1, the Cdc24p GEF for the Cdc42p Rho-like small G protein, GTP-CDC42p, and to actin [5,8,9]. The *bem46* homologous gene-YNL320W in *S. cerevisiae* is not an essential single-copy gene [10]. Likewise, a *bem46* homolog *WAV2* in *Arabidopsis thaliana* was identified and it has no essential function [11]. Global analysis of protein interactions in *D. melanogaster* indicate that there is a low-confidence interaction between *Drosophila* BEM46 and RAPSYNOID [12], which is a putative GEF for Gα protein and involved in the control of asymmetric cell divisions [13]. In *N. crassa*, BEM46 may play a role in a signal transduction pathway involved in determining or maintaining cell type-specific polarity [14]. Hence members of the BEM46 family may modulate the function of some morphogenic determinants, like GEF or actin, on the cell surface. However, *bem46* remains as one of the widespread conserved "known unknown" eukaryotic genes, whose function remains elusive [15,16]. Further studies are clearly required for a better understanding of the function(s) of members belonging to the BEM46 family.

In the current study, we analyzed the molecular evolution of BEM46 family and we compiled BEM46 repository from selected eukaryotic genomes. We demonstrate that BEM46 from different

* Corresponding author.
E-mail address: fkempken@bot.uni-kiel.de (F. Kempken).

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.07.103

eukaryotic lineages show lineage-specific indels that can be used as molecular marker for differentiating BEM46 in different lineages. Using GFP-fusion tagging, we provide evidence that the BEM46 protein (from *N. crassa*) has a novel and previously uncharacterized endoplasmic-retention signal (PEKK).

2. Materials and methods

We describ details of strains, DNA isolation, gel electrophoresis and PCR amplification in supplementary section S1. We also describ details of vector construction, transformation, transformant analysis and microscopic methods in supplementary section S2. We extracted sequences using full-length *N. crassa* BEM46 via NCBI, JGI and Ensembl [17] using BLAST suite [18]. We generated protein alignments of BEM46 sequences with MUSCLE [19] and visualized using GENEDOC [20]. We constructed Bayesian phylogenetic tree in the MrBayes 3.2.1 [21]. We performed divergence analysis for fungal BEM46 protein as described previously [22,23]. We created the homology model of BEM46 protein from *N. crassa* using the I-TASSER [24] and visualized it using YASARA [25]. We predicted active site residues of BEM46 from *N. crassa* using COFACTOR [26]. We provide details of data collection, sequence-structural and phylogenetic analyses in supplementary section S3.

3. Results and discussion

3.1. All eukaryotic organisms possess at least one copy of bem46 gene

We have compiled a comprehensive repository of BEM46 from representative eukaryotic organisms, as summarized by the Bayesian phylogenetic tree (Fig. 1). The majority of eukaryotic organisms have a single copy of BEM46 homolog, however, vertebrates have several paralogs of BEM46 (originated by duplication events), which separated in three different branches in this Bayesian phylogenetic tree (Fig. 1). Fungal BEM46 proteins are grouped into four sub-branches namely ascomycetes, basidiomycetes, *saccharomycetes* and methylotrophic yeasts. Interestingly, the parasitic fungus *Batrachochytrium dendrobatidis* (infecting frogs) has two copies of BEM46 and one copy is a close homolog of vertebrate BEM46 homologs (marked by arrow in Fig. 1). Possibly, this bem46 gene originated from horizontal gene transfer.

We have not detected *bem46* genes in bacterial genomes, however they possess other α/β-hydrolase superfamily members such as lysophospholipase in *Thioalkalivibrio* sp. ALJ9 (Genbank accession id: WP_018173930.1). The members of α/β-hydrolase superfamily are one of the largest and most diverse protein families known and these are comprised of proteases, lipases, esterases, dehalogenases, peroxidases, and epoxide hydrolases with different phylogenetic origin and catalytic function [2,3]. These members have been extensively studied in last two decades after the first report on α/β-hydrolases in 1992 [3]. Hence, we limited our study to the BEM46 protein family. To evaluate the sequence features of BEM46, we examined protein sequences in different eukaryotic lineages and were able to illustrate several insertions/deletions (indels) that might have played important roles in enhancing local diversities within overall conserved BEM46 core domain as described below. The existence of lineage-specific indels may also explain the differences between Bem46 homologs within different eukaryotic lineages.

3.2. Fungal BEM46 proteins have indels specific for fungal phyla

Using protein sequence analyses, we illustrate that there are distinct differences in terms of indels for ascomycetes and basidiomycetes (Fig. 2). A large indel is found in the region between 87P-128I amino acids (according to BEM46 protein numbering from *N. crassa*) as an insertion in basidomycetes but a deletion in ascomycetes. Another two amino acid indels, located between positions 205G-206D, are present in ascomycetes except yeast, but absent in basidiomycetes except rusts. These indels were mapped to BEM46 protein model structure to highlight their respective locations. The large indel is localized in between the β-sheet β2 to the helix α3, just before the α/β-hydrolase domain, while the two amino acid indel between positions 205G-206D is localized in the loop L10 in the α/β-hydrolase domain.

3.3. BEM46 protein from N. crassa possesses novel endoplasmic reticulum (ER)-retention signal

We considered BEM46 protein from *N. crassa* as a standard fungal BEM46, which is characterized by the presence of a signal peptide at the N-terminus with hydrophobic residues and an endoplasmic reticulum (ER)-retention signal PEKK at the C-terminal end (Fig. 3A). The ER-retention signal is designated by a prosite pattern as [KRHQSA]-[DENQ]-E-L. The ER-retention signal for *N. crassa* BEM46 does not fit into this typical pattern, however, it fits into recently reported 35 newly defined ER-retention signals [27] by using various permutation and combinations at these four positions. To examine if this putative motif serves as an authentic ER retention signal, we performed a localization study with (Fig. 3B, C) and without PEKK (Fig. 3D) at the C-terminal end using eGFP as a tag. The native protein carrying the putative ER retention signal coupled with eGFP showed the previously described localization to the perinuclear ER and in additional small spots near to the plasma membrane [14]. In contrast, the eGFP tagged protein without the PEKK sequence was not located to the ER any more, but forms several small cytoplasmic spots (Fig. 3C). Our experiment indicates that the PEKK motif is indeed an ER retention signal.

3.4. Structural analysis of BEM46 protein reveals catalytic triad

We modeled the BEM46 protein from *N. crassa* using I-TASSER [24]. It illustrates that BEM46 has a core α/β-hydrolase domain with eight β-sheets (Fig. 4A). In addition, BEM46 has an extra region with two β-sheets and two α-helices in the N-terminal region. To evaluate diversity within secondary structural elements, we carried out divergence analysis of individual secondary structural elements of different BEM46 proteins during fungal evolution. Fig. 4B illustrates the divergence pattern of these elements among different BEM46 proteins. We found that four loops L5, L11, L13 and L15 are particularly conserved. Normally, loops are conserved in a typical α/β-hydrolase and they harbor active site for ligand bindings. We computed putative binding site residues for three different ligands (Table 1) with C-score higher than −1.5, which reflects correct fold prediction whereas TM-score illustrates structural similarity between the template and the predicted model, [24]. We found that loop L2 possess an important residue, a glycine 117G, for two ligands namely acetic acid and glycerol. Furthermore, amino acid positions between 186G-189L spanning over β-sheet β6 and the loop L9 appear to be critical positions for putative binding sites of fungal BEM46 protein as summarized in Table 1. Among β-sheets, the β-sheet β7 is highly conserved and it harbors a putative binding site residue (isoleucine 211I). The loop L11 harbors an important residue, a phenyl alanine (216F). Among α-helices, the helix α6 appears to be a critical one, as it possesses an important residue, leucine 222L, while the next loop L12 possesses another critical position (proline 228P). Additionally, a potential binding site residue, valine 265 V, is also located near the loop L15, and also a histidine (292H) is located in loop L17. By combining these data, we propose that positions 117G, 188S,