were cognate splicing factors from all stages of the splicing process including components of the A-, B- (U1 snRNP associated factors), and B*-complexes, U2 snRNP associated factors (17S particle), U4/U6 core and associated components, and U4/U6 · U5 specific factors. The HA-CLIP was designed to identify E complexes and perhaps unravel some splicing later-stage factors. Surprisingly, such strategy rendered also core and associated constituents of the Prp19/CDC5L complex, and of the Transcription–Export (TREX) complex, which are major components of the catalytic core of the spliceosome [45–47]. Most of these factors were also detected in the MS2-CLIP, designed to rescue introns engaged in the splicing and post-splicing stages. Although to a limited extent, our results coincide with a number of the human, yeast and chicken spliceosomal factors probed in vivo [18,26,27,30,48,49], and resemble the pattern obtained with trypanosome tagged-SMN copurifications [29,50–52].

Except U6 snRNP, the snRNP particles contain a very stable core formed by seven Sm proteins (Sm E, F, G, D1, D2, D3, and B/B′) [49], and the LSm 2–7 proteins that belong to the U6 snRNP dissociate during spliceosome activation during the formation of the B*-complexes [53]. The active particle catalyzes the first step of splicing. In our experiments, three Sm proteins were detected with the HA-CLIP, and LSm proteins were probed with MS2-tagged introns.

Consistently, SF1 was not detected in our assays but the uridine-rich binding factor TIA-1/TIA L-1 (TIAR) was identified. TIAR reinforces U1 snRNP binding to 5′ss in front of polypyrimidine (PPy) tracts [54,55] resembling most *Entamoeba* 5′ss [8]. Intriguingly, although U1 snRNA has not been predicted in *Entamoeba* [56], we probed the U1 snRNP components U1A and U1-70K.

With our combined approach we detected components of the B*, U4/U6 and U4/U6 · U5 spliceosomal particles. Moreover, we probed components of the A- and B-complexes, proteins important for the first (Prp2), and second (Prp16, Prp22, and Abstrakt) catalytic reactions, and for spliceosome disassembly, Prp43 [57].

Sub2p/UAP56 was the sole TREX component identified. TREX was probed in yeast spliceosomal complexes assembled in vivo by using UV cross-linking combined with affinity purification of exon-tagged and U2 snRNP-tagged proteins [58,59]. TREX binds introns with functional ss concentrated in speckles where splicing occurs [47], as well as to poly(A) + and intron-less mRNAs [60], and it may be associated with the cleavage stimulation factor [61]. It is possible that the *Entamoeba* Mtr4 and Cca1p homologs found here may be also part of the TREX complex.

As expected and in agreement with intron retention, the preferred route of alternative splicing in *Entamoeba* [31], no SRSF proteins (positive splicing modifiers) [62], or hnRNP proteins (negative splicing modifiers) [63] were detected here.

Eight RNA helicases, conserved between human and yeast, are required at each of the spliceosome assembly steps, including the catalytic processes [12,57], and they can be isolated in complexes and purified by using molecular and biochemical tools [26,30]. The best characterized are Sub2p/UAP56, U5-100K (DDX3X)/Prp28, U5-200K/Brr2, and SF3b3/DDX42. From the factors here identified, UAP56 associates with U2AF65 and, together with Prp5, facilitates the interchange between SF1 and U2 snRNP during A-complex formation. In vitro, it helps U2 snRNP recruiting to the BP [57]. The U5-100K helicase is involved in the switch of U1 for U6 at the 5′ss, which also requires U5-220K/Prp8 [12,30]. During spliceosome activation (B to B* complex transition), Prp2 participates with U5-200K/Brr2 for unwinding U4/U6 [57,64], rearranging the spliceosome and repositioning the substrate [65], probably involving the Brr2 modulator U5-116K/Snu114 GTPase. As expected for splicing superfamily 2 helicases [66], the C-terminus of most *Entamoeba* DExH/D-box RNA helicases involved in splicing is conserved (Supplemental Fig. 3A). Furthermore, EhDEAD18 maintains the amino methionine terminal (NMT) domain of Sub2p/UAP56 (Supplemental Fig. 3B) involved in mRNA export and genomic stability in yeast independently of the helicase activity in vitro [67]. Surprisingly, even though *E. histolytica* expresses and recruits in spliceosomal complexes the RNA helicases/ATPases that proofread incorrectly selected 5′ss (Prp28), BS (Prp5, Prp16) or exon-ligation events (Prp16, Prp22, hPrp43) [15,16,68–71], numerous splicing products lack BS and 3′ss selection fidelity [72], suggesting either that additional splicing checkpoint factors might be absent in this parasite, or that splicing defects are somehow buffered by the amoebae.

Another twelve DNA/RNA helicases were identified (Rad3p, Ssl2p, Ssl2p-like, Rvb1p-like, Mtr4, EhDExH12, EhDEAD6, EhDEAD9, and two gene products each of Sgs1p recQ and isw2p). However, their function during spliceosome assembly or catalysis remains to be investigated. Some DNA helicases have dual functions as RNA or DNA helicases [73]. Therefore their participation in spliceosome assembly cannot be ruled out, since other DNA helicases, such as DNJC-8, have been identified previously in human and chicken spliceosome preparations [30].

Actin, myosin, Sec7-, Zn finger-, and CXXC-rich transcription factors, DNA directed polymerase II, and the 40S ribosomal proteins S13 and S19, the cleavage stimulating factors, and nucleotide binding and GTP binding proteins have been reported to belong to spliceosome particles. The remaining 39 factors listed in Table 2 include: 16 protein kinases (seven of them tyrosine-kinases), one phosphatidylinositol kinase, DNA topoisomerase III, two cysteine surface proteins, one Gal/GalNac lectin, one vacuolar targeting protein, and 17 hypothetical *Entamoeba* proteins (labeled H). Only 3 or the 17 hypothetical proteins could not be assigned by BLAST analysis.

As reported for humans and yeasts, the spliceosome has a complex and dynamic composition of proteins, which changes during assembly of each of its functional particles [29,74]. Fig. 5 summarizes the spliceosomal components previously reported along with the factors identified here. The list of spliceosomal components is still growing, and new factors are likely to be identified in the future. Whether all of these newly identified factors are components of the spliceosomal particle, or are just part of the mRNP are questions that will be addressed in the future.

Our work is the first to report the identification of the snRNP and non-snRNP splicing and mRNP factors immunoprecipitated with one *Entamoeba* Early complex splicing factor and one aptamer-tagged functional intron. All of the data suggested that the profiles of the factors identified in this work include factors involved in all steps of the splicing process as well as mRNP factors.

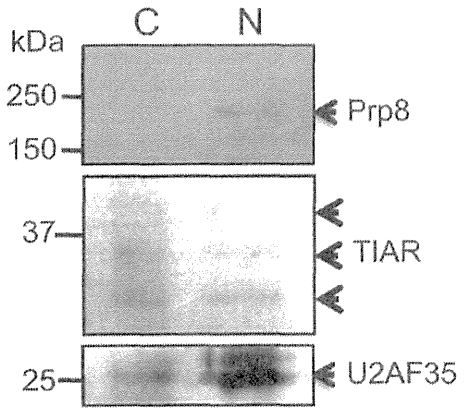


Fig. 6 – Subcellular localization of the *Entamoeba histolytica* splicing factors Prp8, TIAR and U2AF35. *Entamoeba histolytica* proteins isolated from cytoplasmic (C) and nuclear enriched (N) fractions were resolved by 10% SDS-PAGE, blotted onto nitrocellulose and probed with human antibodies to Prp8, TIAR, and U2AF35 (top, middle and bottom panels, respectively).

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2014.07.027.

Conflict of interest

I hereby state that none of the authors have any conflict of interest with any researcher in the field.

Acknowledgments

This work was supported by CONACyT grants 49355M and 127557M and ICYT grant 71/2012 to J.V. and a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan (23390099), a Grant-in-Aid for Scientific Research on Innovative Areas from MEXT of Japan (3308, "Matryoshka-type evolution", 23117001, 23117005), a grant for research on emerging and re-emerging infectious diseases from the Ministry of Health, Labour and Welfare (MHLW) of Japan (H23-Shinkosaiko-ippan-014) to T.N.

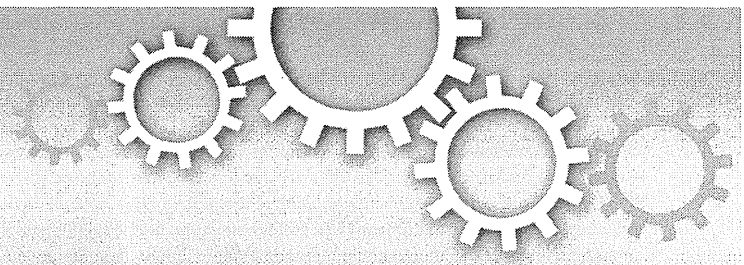
REFERENCES

[1] Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, et al. The genome of the protist parasite *Entamoeba histolytica*. Nature 2005;433:865–8.
[2] Davis CA, Brown MP, Singh U. Functional characterization of spliceosomal introns and identification of U2, U4, and U5 snRNAs in the deep-branching eukaryote *Entamoeba histolytica*. Eukaryot Cell 2007;6:940–8.
[3] Lohia A, Samuelson J. Cloning of the Eh cdc2 gene from *Entamoeba histolytica* encoding a protein kinase p34cdc2 homologue. Gene 1993;127:203–7.
[4] Marchat LA, Orozco E, Guillen N, Weber C, Lopez-Camarillo C. Putative DEAD and DExH-box RNA helicases families in *Entamoeba histolytica*. Gene 2008;424:1–10.
[5] Plaimauer B, Ortner S, Wiedermann G, Scheiner O, Duchene M. An intron-containing gene coding for a novel 39-kilodalton antigen of *Entamoeba histolytica*. Mol Biochem Parasitol 1994;66:181–5.
[6] Sanchez-Lopez R, Gama-Castro S, Ramos MA, Merino E, Lizardi PM, Alagon A. Cloning and expression of the *Entamoeba histolytica* ERD2 gene. Mol Biochem Parasitol 1998; 92:355–9.
[7] Urban B, Blasig C, Forster B, Hamelmann C, Horstmann RD. Putative serine/threonine protein kinase expressed in complement-resistant forms of *Entamoeba histolytica*. Mol Biochem Parasitol 1996;80:171–8.
[8] Wilihoeft U, Campos-Gongora E, Touzni S, Bruchhaus I, Tannich E. Introns of *Entamoeba histolytica* and *Entamoeba dispar*. Protist 2001;152:149–56.
[9] Willhoeft U, Hamann L, Tannich E. A DNA sequence corresponding to the gene encoding cysteine proteinase 5 in *Entamoeba histolytica* is present and positionally conserved but highly degenerated in *Entamoeba dispar*. Infect Immun 1999;67:5925–9.
[10] Will CL, Lührmann R. In: Gesteland RF, T. R. C., Atkins JF, editors. The RNA world. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2006. p. 369–400.
[11] Nilsen TW. In: Grunberg-Manago RWSaM, editor. RNA structure and function. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 1998. p. 279–307.
[12] Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 1998;92:315–26.
[13] Valadkhan S. snRNAs as the catalysts of pre-mRNA splicing. Curr Opin Chem Biol 2005;9:603–8.
[14] Burgess S, Couto JR, Guthrie C. A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. Cell 1990;60:705–17.
[15] Mayas RM, Maita H, Staley JP. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol 2006;13: 482–90.
[16] Xu YZ, Query CC. Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. Mol Cell 2007;28:838–49.
[17] Behzadnia N, Golas MM, Hartmuth K, Sander B, Kastner B, Deckert J, et al. Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J 2007;26:1737–48.
[18] Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, Kastner B, et al. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol Cell Biol 2006;26:5528–43.
[19] Bessonov S, Anokhina M, Will CL, Urlaub H, Luhrmann R. Isolation of an active step I spliceosome and composition of its RNP core. Nature 2008;452:846–50.
[20] Boehringer D, Makarov EM, Sander B, Makarova OV, Kastner B, Luhrmann R, et al. Three-dimensional structure of a pre-catalytic human spliceosomal complex B. Nat Struct Mol Biol 2004;11:463–8.
[21] Hartmuth K, Urlaub H, Vornlocher HP, Will CL, Gentzel M, Wilm M, et al. Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci U S A 2002;99:16719–24.
[22] Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA 2002; 8:426–39.
[23] Jurica MS, Sousa D, Moore MJ, Grigorieff N. Three-dimensional structure of C complex spliceosomes by electron microscopy. Nat Struct Mol Biol 2004;11:265–9.

[24] Makarov EM, Makarova OV, Urlaub H, Gentzel M, Will CL, Wilm M, et al. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science 2002; 298:2205–8.
[25] Makarova OV, Makarov EM, Urlaub H, Will CL, Gentzel M, Wilm M, et al. A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. EMBO J 2004;23:2381–91.
[26] Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. Genome Res 2002;12:1231–45.
[27] Zhou Z, Licklider LJ, Gygi SP, Reed R. Comprehensive proteomic analysis of the human spliceosome. Nature 2002; 419:182–5.
[28] Herold N, Will CL, Wolf E, Kastner B, Urlaub H, Luhrmann R. Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes. Mol Cell Biol 2009;29: 281–301.
[29] Chen W, Shulha HP, Ashar-Patel A, Yan J, Green KM, Query CC, et al. Endogenous U2 · U5 · U6 snRNA complexes in *S. pombe* are intron lariat spliceosomes. RNA 2014;20:308–20.
[30] Chen YI, Moore RE, Ge HY, Young MK, Lee TD, Stevens SW. Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. Nucleic Acids Res 2007;35:3928–44.
[31] McGuire AM, Pearson MD, Neafsey DE, Galagan JE. Cross-kingdom patterns of alternative splicing and splice recognition. Genome Biol 2008;9:R50.
[32] Miranda R, Salgado LM, Sanchez-Lopez R, Alagon A, Lizardi PM. Identification and analysis of the u6 small nuclear RNA gene from *Entamoeba histolytica*. Gene 1996;180:37–42.
[33] Collins L, Penny D. Complex spliceosomal organization ancestral to extant eukaryotes. Mol Biol Evol 2005;22:1053–66.
[34] Azubel M, Wolf SG, Sperling J, Sperling R. Three-dimensional structure of the native spliceosome by cryo-electron microscopy. Mol Cell 2004;15:833–9.
[35] Diamond LS, Harlow DR, Cunnick CC. A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. Trans R Soc Trop Med Hyg 1978;72:431–2.
[36] Diamond LS, Mattern CF, Bartgis IL. Viruses of *Entamoeba histolytica*. I. Identification of transmissible virus-like agents. J Virol 1972;9:326–41.
[37] Saito-Nakano Y, Mitra BN, Nakada-Tsukui K, Sato D, Nozaki T. Two Rab7 isotypes, EhRab7A and EhRab7B, play distinct roles in biogenesis of lysosomes and phagosomes in the enteric protozoan parasite *Entamoeba histolytica*. Cell Microbiol 2007;9:1796–808.
[38] Biller L, Davis PH, Tillack M, Matthiesen J, Lotter H, Stanley Jr SL, et al. Differences in the transcriptome signatures of two genetically related *Entamoeba histolytica* cell lines derived from the same isolate with different pathogenic properties. BMC Genomics 2010;11:63.
[39] Nozaki T, Asai T, Sanchez LB, Kobayashi S, Nakazawa M, Takeuchi T. Characterization of the gene encoding serine acetyltransferase, a regulated enzyme of cysteine biosynthesis from the protist parasites *Entamoeba histolytica* and *Entamoeba dispar*. Regulation and possible function of the cysteine biosynthetic pathway in *Entamoeba*. J Biol Chem 1999;274:32445–52.
[40] Furukawa A, Nakada-Tsukui K, Nozaki T. Novel transmembrane receptor involved in phagosome transport of lysozymes and beta-hexosaminidase in the enteric protozoan *Entamoeba histolytica*. PLoS Pathog 2012;8:e1002539.
[41] Yoshimoto R, Kataoka N, Okawa K, Ohno M. Isolation and characterization of post-splicing lariat-intron complexes. Nucleic Acids Res 2009;37:891–902.
[42] Lutz CS, Cooke C, O'Connor JP, Kobayashi R, Alwine JC. The snRNP-free U1A (SF-A) complex(es): identification of the largest subunit as PSF, the polypyrimidine-tract binding protein-associated splicing factor. RNA 1998;4:1493–9.
[43] McGugan Jr GC, Joshi MB, Dwyer DM. Identification and biochemical characterization of unique secretory nucleases of the human enteric pathogen, *Entamoeba histolytica*. J Biol Chem 2007;282:31789–802.
[44] Lu Y, Qian XY, Krug RM. The influenza virus NS1 protein: a novel inhibitor of pre-mRNA splicing. Genes Dev 1994;8: 1817–28.
[45] Grote M, Wolf E, Will CL, Lemm I, Agafonov DE, Schomburg A, et al. Molecular architecture of the human Prp19/CDC5L complex. Mol Cell Biol 2010;30:2105–19.
[46] Bono F, Gehring NH. Assembly, disassembly and recycling: the dynamics of exon junction complexes. RNA Biol 2011;8: 24–9.
[47] Dias AP, Dufu K, Lei H, Reed R. A role for TREX components in the release of spliced mRNA from nuclear speckle domains. Nat Commun 2010;1:97.
[48] Agafonov DE, Deckert J, Wolf E, Odenwalder P, Bessonov S, Will CL, et al. Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. Mol Cell Biol 2011;31:2667–82.
[49] Wolf E, Kastner B, Deckert J, Merz C, Stark H, Luhrmann R. Exon, intron and splice site locations in the spliceosomal B complex. EMBO J 2009;28:2283–92.
[50] Luz Ambrosio D, Lee JH, Panigrahi AK, Nguyen TN, Cicarelli RM, Gunzl A. Spliceosomal proteomics in *Trypanosoma brucei* reveal new RNA splicing factors. Eukaryot Cell 2009;8: 990–1000.
[51] Palfi Z, Jae N, Preusser C, Kaminska KH, Bujnicki JM, Lee JH, et al. SMN-assisted assembly of snRNP-specific Sm cores in trypanosomes. Genes Dev 2009;23:1650–64.
[52] Tkacz ID, Gupta SK, Volkov V, Romano M, Haham T, Tulinski P, et al. Analysis of spliceosomal proteins in Trypanosomatids reveals novel functions in mRNA processing. J Biol Chem 2010;285:27982–99.
[53] Chan SP, Kao DI, Tsai WY, Cheng SC. The Prp19p-associated complex in spliceosome activation. Science 2003;302:279–82.
[54] Forch P, Puig O, Martinez C, Seraphin B, Valcarcel J. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites. EMBO J 2002;21:6882–92.
[55] Zhu H, Hasman RA, Young KM, Kedersha NL, Lou H. U1 snRNP-dependent function of TIAR in the regulation of alternative RNA processing of the human calcitonin/CGRP pre-mRNA. Mol Cell Biol 2003;23:5959–71.
[56] Davila Lopez M, Rosenblad MA, Samuelsson T. Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. Nucleic Acids Res 2008;36: 3001–10.
[57] Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell 2009;136:701–18.
[58] Gornemann J, Kotovic KM, Hujer K, Neugebauer KM. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell 2005; 19:53–63.
[59] Lacadie SA, Tardiff DF, Kadener S, Rosbash M. In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. Genes Dev 2006;20: 2055–66.
[60] Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, et al. TREX is a conserved complex coupling transcription with messenger RNA export. Nature 2002;417:304–8.
[61] Katahira J, Okuzaki D, Inoue H, Yoneda Y, Maehara K, Ohkawa Y. Human TREX component Thoc5 affects alternative polyadenylation site choice by recruiting mammalian cleavage factor I. Nucleic Acids Res 2013;41: 7060–72.

[62] Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. Biochem J 2009;417: 15–27.
[63] Han SP, Tang YH, Smith R. Functional diversity of the hnRNPs: past, present and perspectives. Biochem J 2010;430: 379–92.
[64] Wlodaver AM, Staley JP. The DExD/H-box ATPase Prp2p destabilizes and proofreads the catalytic RNA core of the spliceosome. RNA 2014;20:282–94.
[65] Hahn D, Kudla G, Tollervey D, Beggs JD. Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. Genes Dev 2012;26: 2408–21.
[66] Fairman-Williams ME, Guenther UP, Jankowsky E. SF1 and SF2 helicases: family matters. Curr Opin Struct Biol 2010;20: 313–24.
[67] Saguez C, Gonzales FA, Schmid M, Boggild A, Latrick CM, Malagon F, et al. Mutational analysis of the yeast RNA helicase Sub2p reveals conserved domains required for growth, mRNA export, and genomic stability. RNA 2013;19: 1363–71.
[68] Yang F, Wang XY, Zhang ZM, Pu J, Fan YJ, Zhou J, et al. Splicing proofreading at 5′ splice sites by ATPase Prp28p. Nucleic Acids Res 2013;41:4660–70.
[69] Koodathingal P, Novak T, Piccirilli JA, Staley JP. The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5′ splice site cleavage during pre-mRNA splicing. Mol Cell 2010;39:385–95.
[70] Egecioglu DE, Chanfreau G. Proofreading and spellchecking: a two-tier strategy for pre-mRNA splicing quality control. RNA 2011;17:383–9.
[71] Query CC, Konarska MM. Splicing fidelity revisited. Nat Struct Mol Biol 2006;13:472–4.
[72] Hon CC, Weber C, Sismeiro O, Proux C, Koutero M, Deloger M, et al. Quantification of stochastic noise of splicing and polyadenylation in *Entamoeba histolytica*. Nucleic Acids Res 2013;41:1936–52.
[73] Jankowsky E. RNA helicases at work: binding and rearranging. Trends Biochem Sci 2011;36:19–29.
[74] Will CL, Luhrmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol 2011;3.

SCIENTIFIC REPORTS



OPEN

SUBJECT AREAS:
COMPUTATIONAL BIOLOGY AND BIOINFORMATICS
EVOLUTION
CELL BIOLOGY
PARASITOLOGY

Received
18 November 2014

Accepted
23 January 2015

Published
25 February 2015

Correspondence and requests for materials should be addressed to T.N. (nozaki@niid.go.jp)

*These authors contributed equally to this work.

A Novel Mitosomal β-Barrel Outer Membrane Protein in *Entamoeba*

Herbert J. Santos[1,2,3]*, Kenichiro Imai[4]*, Takashi Makiuchi[5], Kentaro Tomii[4], Paul Horton[4], Akira Nozawa[6], Mohamed Ibrahim[7], Yuzuru Tozawa[8] & Tomoyoshi Nozaki[1,2]

[1]Department of Parasitology, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan, [2]Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan, [3]Institute of Biology, College of Science, University of the Philippines Diliman, Quezon City, 1101 Philippines, [4]Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan, [5]Department of Infectious Diseases, Tokai University School of Medicine, Isehara, Kanagawa 259-1193, Japan, [6]Proteo-Science Center, Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan, [7]Botany Department, Faculty of Science, Ain Shams University, Khalifa El-Maamon St, Abbasiya Sq., Cairo, 11566, Egypt, [8]Graduate School of Science and Engineering, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama, Saitama 338-8570, Japan.

Entamoeba **possesses a highly divergent mitochondrion-related organelle known as the mitosome. Here, we report the discovery of a novel protein in *Entamoeba*, which we name Mitosomal β-barrel Outer Membrane Protein of 30 kDa (MBOMP30). Initially identified through *in silico* analysis, we experimentally confirmed that MBOMP30 is indeed a β-barrel protein. Circular dichroism analysis showed MBOMP30 has a predominant β-sheet structure. Localization to *Entamoeba histolytica* mitosomes was observed through Percoll-gradient fractionation and immunofluorescence assay. Mitosomal membrane integration was demonstrated by carbonate fractionation, proteinase K digestion, and immunoelectron microscopy. Interestingly, the deletion of the putative β-signal, a sequence believed to guide β-barrel outer membrane protein (BOMP) assembly, did not affect membrane integration, but abolished the formation of a ~240 kDa complex. MBOMP30 represents only the seventh subclass of eukaryotic BOMPs discovered to date and lacks detectable homologs outside *Entamoeba*, suggesting that it may be unique to *Entamoeba* mitosomes.**

Mitochondria can possess highly divergent, and often degenerate, morphology, function, and components in eukaryotes that are adapted to anoxic or hypoxic environments. In cases in which morphology is drastically changed and some hallmark mitochondrial processes such as oxidative phosphorylation, the TCA cycle, and β-oxidation are lost, these organelles are called Mitochondrion-Related Organelles (MROs), or specifically, hydrogenosome and mitosome. Mitosomes are particularly degenerate organelles, lacking cristae structure, and even the ability to synthesize ATP. It is believed that mitosomes, as well as hydrogenosomes, have occurred multiple times during eukaryotic evolution because organisms that possess mitosomes do not cluster together in eukaryote phylogenies, and the size, function, and content of mitosomes differ between organisms[1–3]. Like Gram-negative bacteria and chloroplasts in primary land plants, mitochondria and MROs possess a double membrane. Transport of proteins and metabolites across the outer membrane is mediated by pore-forming β-barrel Outer Membrane Proteins. (Hereafter we use "BOMP" to denote any β-barrel outer membrane protein and "MBOMP" to denote BOMPs from mitochondria and MROs).

In mitochondria, six subclasses of MBOMPs have been previously identified: Tom40, Sam50, VDAC, Mdm10, ATOM and Tac40. Tom40 is the core pore component of the Translocase of the Outer Membrane (TOM) complex required for the import of mitochondrial precursor proteins into mitochondria[4,5]. Sam50 is the central component of the Sorting and Assembly Machinery (SAM) complex and promotes the integration of MBOMPs[6–8] to the outer membrane. Both Tom40 and Sam50 are essential for yeast viability[4,5,8]. VDAC (Voltage-Dependent Anion Channel) primarily serves as a non-specific diffusion pore for small molecules entering or leaving the mitochondria[9]. Mdm10 (Mitochondrial Dynamics and Morphology 10) has only been clearly identified in fungi, and is involved in mitochondrial morphogenesis and dynamics[10], as well as in the biogenesis of mitochondrial BOMPs as it was reported to be a part of the SAM complex[11]. It is also a member of the ER-mitochondria tethering complex known as the ER-Mitochondria Encounter Structure, ERMES[12]. Trypanosomatids lack Tom40, and instead have a unique translocase called ATOM[13] (Archaic Translocase of the Outer Mitochondrial membrane). Recently another trypanosome-specific MBOMP Tac40, a member of the

Tripartite Attachment Complex, was identified. Tac40 also belongs to the mitochondrial porin family, and is essential to mitochondrial DNA inheritance, as it physically links the mitochondrial genome to cytoskeletal components of both the mitochondrion and flagellum of *Trypanosoma brucei*[14]. Among all MBOMPs, only Sam50 and ATOM have recognizable homologs outside of the Eukaryota. Sam50 has a bacterial homolog, Omp85/BamA, a 16-stranded bacterial BOMP[15], which also integrates and assembles BOMPs[16]; while the homolog of ATOM belongs to the Omp85 subgroup called YtfM[13], which is a BOMP required for normal growth in *Escherichia coli*[17].

Entamoeba histolytica is an anaerobic unicellular parasite that causes dysentery and extra-intestinal abscesses that are responsible for an estimated 100,000 deaths annually. This organism possesses highly divergent mitosomes, as predicted by a recent proteomic study[18]. It appears that the mitosome proteome in *E. histolytica* is far less complex than mitochondria (e.g., yeast mitochondria are believed to harbor around 1000 proteins[19,20]), and remarkably different from other MROs. Indeed, even Fe-S cluster biogenesis, which is the only known common function of mitochondria and MROs, is uncertain in *E. histolytica* mitosomes[21,22]. It was reported that the iron-sulfur cluster assembly genes *iscS* and *iscU* of *E. histolytica* were acquired by horizontal gene transfer[22,23] and unlike other organisms, *E. histolytica*[21,22] and the distantly related *Mastigamoeba balamuthi*[24] use a NIF (nitrogen fixation)-like system for Fe-S cluster biogenesis; which is primarily, if not exclusively, cytosolic and appears to have been obtained by horizontal transfer from ε-proteobacteria. Thus far, the only established role of the *Entamoeba* mitosome is sulfate activation[18,25].

Among the many "missing links" in understanding mitosomal biology is the conspicuous absence of detectable homologs to VDAC, an MBOMP which is the usual channel for metabolites in the mitochondrial outer membrane. Intrigued by this and encouraged by the fact that our previously developed method[26,27] for predicting MBOMPs from amino acid sequence had already been able to predict a candidate novel plastid BOMP TGD4 (At3g06960), subsequently confirmed to localize to the plastid outer membrane[28]; we endeavored to combine our informatics and experimental techniques to search for novel *Entamoeba* BOMPs. These efforts led us to discover MBOMP30, a novel lineage-specific BOMP which localizes to the outer membrane of the *E. histolytica* mitosome.

Results

***In silico* screening of novel MBOMP candidates in *E. histolytica*.** To identify novel MBOMP candidates in *E. histolytica*, we customized our MBOMP predictor[27] for MRO's, and also refined the screening method in general (see Materials and methods) (Fig. 1, Supplementary Table S1). We screened 8,306 proteins in the *E. histolytica* genome database using an updated version of our MBOMP prediction pipeline, yielding six MBOMP candidates: EHI_178630, EHI_007460, EHI_163510, EHI_050690, EHI_068370, and EHI_104420 (Table 1). EHI_104420 is a homolog of Tom40, which has been previously characterized[29]. Four of the other five proteins are annotated as non-mitochondrial proteins and have sequence similarity to known functional domains. Thus, in this study, we chose to focus on the remaining protein EHI_178630 (Uniprot accession C4LUS8), which attained the second highest predictor score, second only to the known MBOMP Tom40, and we hereafter denote as *E. histolytica* MBOMP30 (Mitosomal β-barrel Outer Membrane Protein of 30kDa).

MBOMP30 has no detectable homologs outside of *Entamoeba*. To examine the phylogenetic distribution of MBOMP30 among the three domains of life, we used the highly sensitive iterative profile HMM comparison search methods JackHMMR and HHblits[30,31] to search for homologs of *E. histolytica* MBOMP30 in other organisms. We found no significant hits (E-value < 1e-5) in either bacteria or archaea. Among eukaryotes, we found homologs only in genus *Entamoeba*, namely, *E. nuttalli*, *E. dispar* and *E. invadens*. Notably, we found no hits in the relatively well-characterized free-living aerobic amoebozoan *Dictyostelium discoideum*. Both highly related to *E. histolytica, E. nuttalli*, a pathogenic parasite of non-human primates, and *E. dispar*, a non-pathogenic species, possess very similar sequences ENU1_140620 and EDI_035580, alignable to MBOMP30 at 97.2% and 86.5% identity (although EDI_035580 is truncated), respectively. On the other hand, the more distant reptilian parasite *E. invadens* yielded a more diverged, but still highly significant hit EIN_041060 (with an identical sequence assigned as EIN_066350 in NCBI), having a 32.8% identity (E-value 1.8e-27) (Supplementary Table S2).

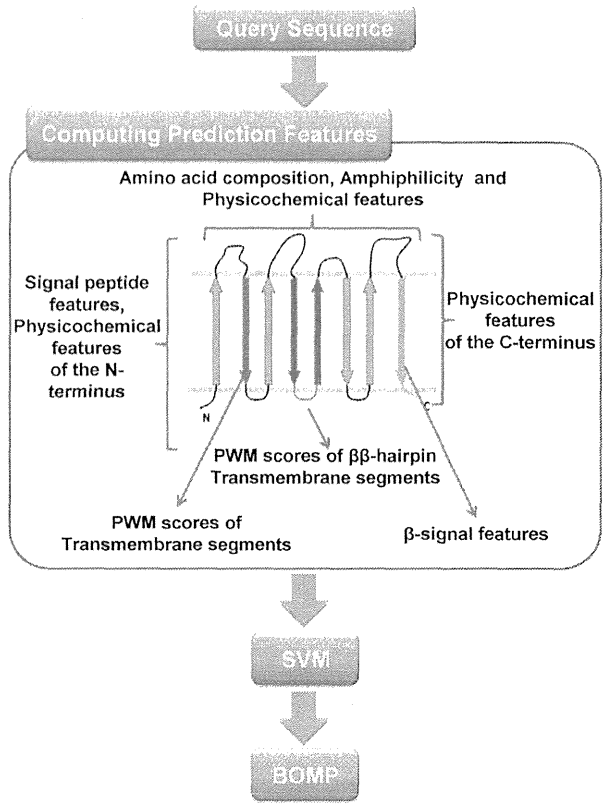


Figure 1 | The architecture of our MBOMP prediction system. Sequence features used by our MBOMP predictor are listed in relation to a schematic depiction of MBOMP structure.

MBOMP30 is predicted to contain transmembrane β-strands. Our MBOMP predictor gives a high probability score for all four *Entamoeba* MBOMP30 homologs. To assure that this result was not a quirk of our predictor, we analyzed the *E. histolytica* MBOMP30 sequences using two tools designed for topology prediction of bacterial BOMPs, which may be expected to share some structural properties with MBOMPs. The topology prediction tools BOCTOPUS and TMBETAPRED-RBF[32,33] predicted the *Entamoeba* MBOMP30 to contain multiple transmembrane β-strand regions (Fig. 2, Supplementary Fig. S1a), consistent with the premise that MBOMP30 is a mitosomal BOMP. The outside surface of MBOMPs faces a lipid environment and is expected to display a relatively high hydrophobicity. The MBOMP30 sequences indeed share this property, and their predicted β-strand regions and hydrophobic stretches align well (Supplementary Fig. S1b).

Table 1 | List of predicted MBOMP candidates in *E. histolytica*. JackHMMER searches were performed against NCBI non-redundant protein database with the best E-value amongst annotated proteins shown. The β-signal motif ($P_xGxxH_yxH_y$) was matched against the β-strand closest to the C-terminus, as predicted by BOCTOPUS

Protein ID	Length	MBOMP prediction probability	Predicted β-strand ratio	Predicted α-helix ratio	Annotation inferred from homology	JackHMMER Hit (E-value)	Pfam domain (E-value)	β-signal motif hit
EHI_104420	284	0.961	0.29	0.21	Tom40	*C. parvum*, Tom40p like translocase (0.026)	Eukaryotic porin (2.9e-09)	Yes
EHI_178630	282	0.954	0.34	0.23	Unknown	No significant hit	No significant hit	Yes
EHI_007460	304	0.871	0.27	0.20	Phospholipase C	*T. thermophile*, Endonuclease/ Exonuclease/ phosphatase family protein (1.2e-76)	Endonuclease/ Exonuclease/ phosphatase family (3.5e-15)	Yes
EHI_163510	263	0.853	0.31	0.18	Exosome complex exonuclease RRP4	*D. rerio*, Exosome complex exonuclease RRP4 (8.9e-100)	S1 RNA binding domain (0.19)	No
EHI_050690	321	0.749	0.43	0.09	Apyrase	*P. sergenti*, salivary apyrase (3.1e-79)	Apyrase (5.4e-09)	Yes
EHI_068370	466	0.734	0.56	0.00	Cysteine protease binding protein family like protein	*E. histolytica*, cysteine protease binding protein family 4 (8.5e-178)	No significant hit	No

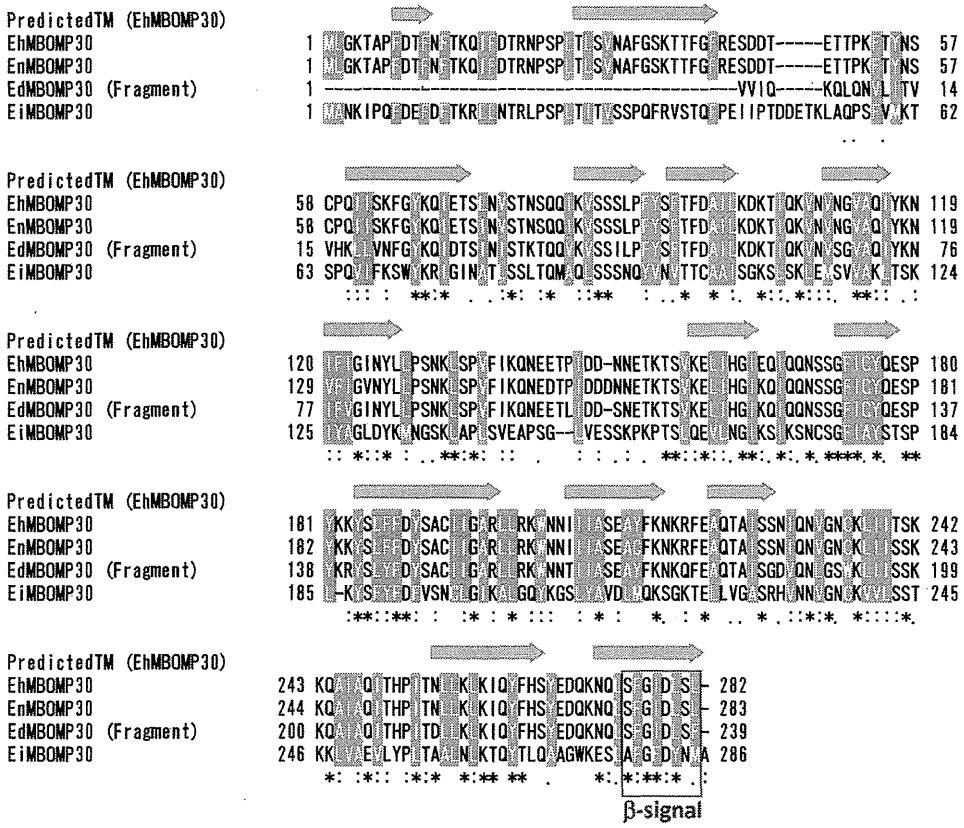


Figure 2 | Prediction of transmembrane β-strands of EhMBOMP30 and multiple alignment with its homologs. Multiple alignment of EhMBOMP30, EnMBOMP30, EdMBOMP30, and EiMBOMP30, built with Clustal omega[59] is shown in CLUSTAL format. Predicted transmembrane β-strands of EhMBOMP30 are indicated with yellow arrows. The arrows cover all positions predicted as part of a β-strand by either of two BOMP topology predictors: BOCTOPUS and TMBETAPRED-RBF[32,33] [See Fig. S1]. Conserved hydrophobic residues are highlighted in green. A red dashed box indicates the region corresponding to the β-signal motif in the last predicted transmembrane β-strand.

Table 2 | Estimation of EhMBOMP30 secondary structure organization based on CD spectra analysis using CONTIN software

Trial	Predicted β-strand %	Predicted α-helix %	Predicted β-turns %	Predicted random coil %	NMRSD	λ_{min}, nm	λ_{max}, nm
1	33.5	12.2	19.2	35.1	0.059	221.5	196.0
2	27.8	14.8	9.5	47.9	0.088	221.9	192.5
3	30.3	15.6	15.9	38.5	0.094	220.9	192.8
Avg ± SD	30.5 ± 2.9	14.1 ± 1.7	14.9 ± 4.9	40.5 ± 6.6			

NMRSD is the normalized root mean square deviation from the CONTIN analysis and should be >0.1.

Far-UV spectroscopy and circular dichroism (CD) suggests high β-strand content of *E. histolytica* MBOMP30. We synthesized *E. histolytica* MBOMP30 by wheat germ cell-free expression system in the presence of 1,2-diphytaynyl-sn-glycero-3-phosphocholine (DPhPC) liposomes (Supplementary Fig. S2), and estimated the secondary structure composition of the protein by CD spectroscopy. The *E. histolytica* MBOMP30 far-UV spectra indicated a pattern representative of β-strand-rich proteins, having minimum and maximum ellipticities near 220 and 195 nm, respectively (Table 2, Fig. 3). This was similar to that of a β-barrel control, green fluorescent protein[34]. In addition, using the CONTIN algorithm, the deconvoluted CD spectra predicted that *E. histolytica* MBOMP30 has a high β-strand-content estimated at 30.5 ± 2.9%, while having only 14.1 ± 1.7% α-helices. It was also estimated that the protein has 14.9 ± 4.9% β-turns, and 40.5 ± 6.6% random/unordered secondary structure (Table 2). Likewise, a previous report on the CD spectra of human VDAC (hVDAC1) in DPhPC liposome estimated the β-strand, α-helix, β-turn, and random coil contents of hVDAC1 as 37.3%, 7.7%, 22.6%, and 32.4%, respectively[35]. The association of MBOMP30 in DPhPc liposomes and the far-UV and CD data further support our prediction that MBOMP30 is a β-barrel protein.

Localization of MBOMP30 to the *Entamoeba* mitosome. We experimentally verified the localization of MBOMP30 in *E. histolytica* trophozoites using an *E. histolytica* cell line expressing MBOMP30 tagged with the hemagglutinin epitope (HA) at the amino terminus (HA-MBOMP30). We avoided potential interference with the putative β-signal, which is located at the carboxyl terminus, by using amino-terminal HA tagging. We confirmed the molecular mass of the expressed product was as expected using whole amoebic lysates and anti-HA antibody (Fig. 4a). We further fractionated lysates from HA-MBOMP30-expressing trophozoites by two rounds of Percoll gradient ultracentrifugation, followed by immunodecoration with anti-HA antibody and antiserum raised against Chaperonin 60 (Cpn60), a canonical mitosomal protein. The distribution of the band corresponding to HA-MBOMP30 throughout the fractions was similar to that of Cpn60 (Fig. 4b), suggesting that MBOMP30 localizes to mitosomes. In addition, immunofluorescence assay (IFA) using anti-HA antibody, and anti-adenosine-5′-phosphokinase (APSK; XP_656278; also mitosomal[18]) antisera showed that HA-MBOMP30 colocalizes with APSK (Fig. 4c). This result confirms the mitosomal localization of MBOMP30 reported previously[18].

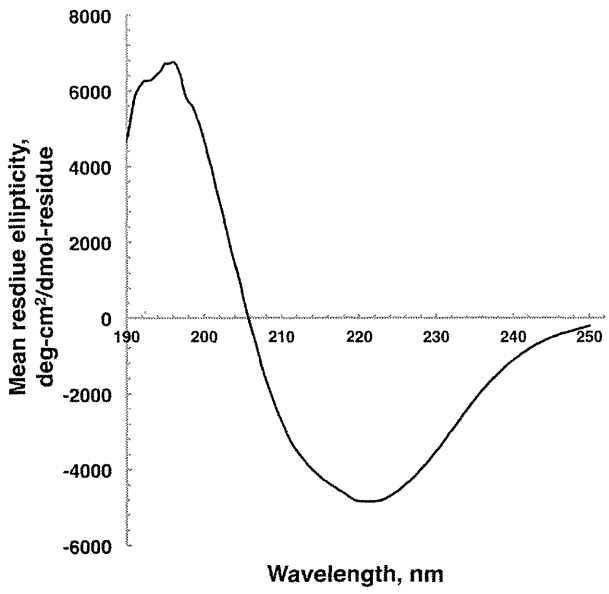


Figure 3 | Circular dichroism spectroscopy. The far-ultraviolet CD spectra of MBOMP30 in DPhPc liposomes, expressed in a cell-free system, were taken from the accumulation of 9 consecutive scans.

MBOMP30 is integrated to the mitosomal membrane. To verify our prediction that MBOMP30 is an integral outer membrane protein, we subjected the particulate (membrane) fraction to 100,000 *g* ultracentrifugation with Na_2CO_3, which is known to liberate soluble matrix and peripheral membrane proteins from organelles[36]. Immunoblot analysis showed that MBOMP30-HA was detected in the pellet fraction after Na_2CO_3 treatment, similar to the positive control, mitosomal BOMP EhTom40-HA, suggesting membrane integration of HA-MBOMP30 (Fig. 5a). In contrast, the soluble mitosomal matrix protein Cpn60 was observed only in the supernatant fraction after Na_2CO_3 treatment (Fig. 5a). We also performed proteinase K protection assay to ascertain the membrane topology of MBOMP30 in amoeba trophozoites. Immunoblots revealed that the sensitivity pattern of MBOMP30 to proteinase K degradation is intermediate to that of an outer membrane control, Tom40-HA, and an inner membrane control, AAC (ATP/ADP Carrier[25,37])-HA, while expectedly, a matrix marker APSK-HA, showed the least susceptibility to proteinase K digestion (Fig. 5b, Supplementary Fig. S3). Interestingly, when we observed mitosomes of HA-MBOMP30 trophozoites by immunoelectron microscopy, it appears that the protein is localized on the membrane of the mitosomes (Fig. 5c). The micrographs reveal mitosomes having an electron-dense region, with a diameter ranging from 100–600 nm, enclosed by a double membrane, with the matrix marked by anti-Cpn60 staining (small gold particles). Notably, abundant anti-HA staining (large gold particles) of mitosomal membranes was observed in HA-MBOMP30-expressing transformants. In addition, the micrographs of randomly selected mitosomes (n = 11) in HA-MBOMP30 trophozoites, revealed that the gold anti-HA particle distribution of 25.2 ± 15.7 gold/μm² in the mitosomal membranes was significantly higher compared to just 0.18 ± 0.22 gold/μm² in the cytoplasm ($p < 0.001$, using Student's *t*-test).

MBOMP30 has a putative β-signal. A short, carboxyl-terminal sequence known as the β-signal[38] plays an important role in MBOMP integration and/or assembly in the mitochondrial outer membrane of *Saccharomyces cerevisiae* and by comparative sequence analysis is inferred to do so in a wide range of eukaryotes. Based on a comparison of numerous putative MBOMP sequences, we have proposed a slightly refined β-signal consensus sequence: $P_oxGh_yxH_yxH_y$ (P_o: non-negatively charged polar, G: glycine, H_y: large hydrophobic, h_y: loosely defined hydrophobic

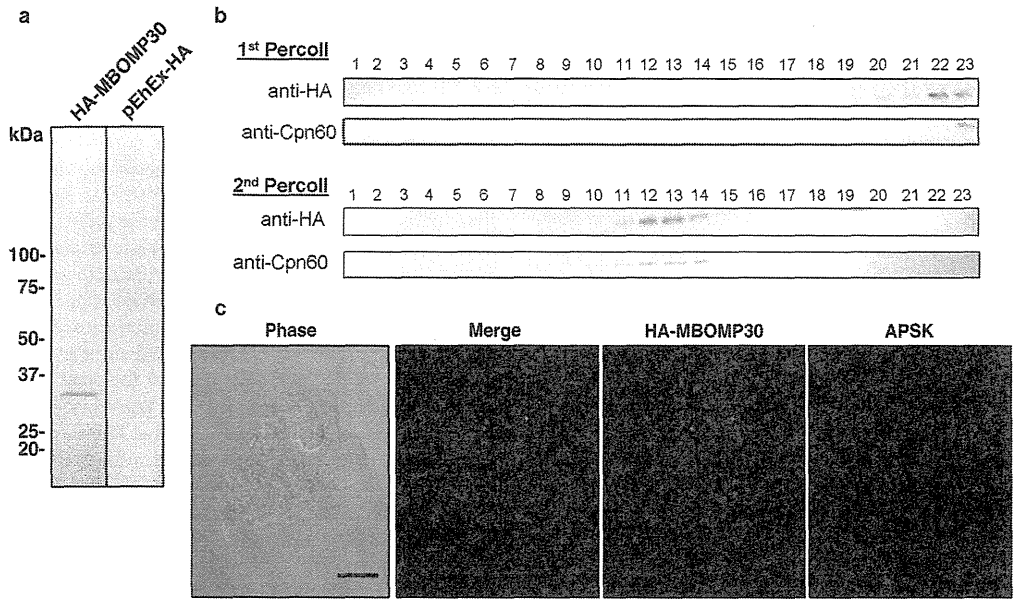


Figure 4 | Localization analyses of MBOMP30. (a) Ectopic expression of HA-MBOMP30 in *E. histolytica* trophozoites. Approximately 20 μg of cell lysates from HA-MBOMP30 and control strain was fractionated on SDS-PAGE and subjected to immunoblot analysis using anti-HA antibody. (b) Immunoblot analysis of the fractions by two series of Percoll gradient ultracentrifugation. Approximately 20 μL of each fraction of the first and second ultracentrifugation was separated by SDS-PAGE and blotted to nitrocellulose membranes. The blots were cut into strips containing the region of the target proteins, before reacting with anti-HA and anti-Cpn60 antibodies. (c) Immunofluorescence analysis of HA-MBOMP30 trophozoites. Colocalization of punctate anti-HA (green) and anti-APSK (red; mitosomal marker) signals is shown [scale bar, 10 μm].

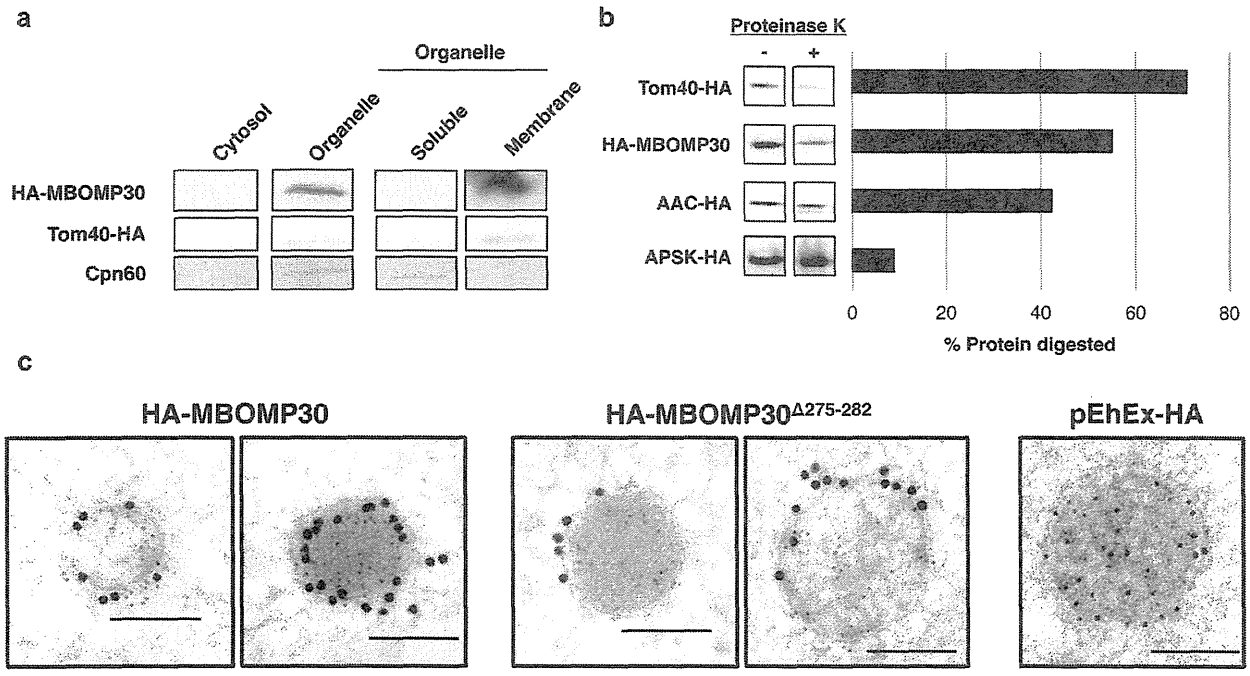


Figure 5 | Mitosomal membrane localization of MBOMP30. (a) Na_2CO_3 fractionation. Homogenates from amoebae expressing HA-MBOMP30 and Tom40-HA were fractionated. The organelle fraction was treated with Na_2CO_3 and NaCl to lyse the organelle and liberate loosely bound proteins. The resulting fractions were separated on SDS-PAGE followed by immunoblotting. Parts of the full-length immunoblot for the anti-HA (first two rows) and anti-Cpn60 (last row) antibody reactions are shown respectively. Tom40-HA and Cpn60 serve as a control for mitosomal BOMP and soluble mitosomal proteins, respectively. The original blots are shown in Supplementary Fig. S4a–b. (b) Proteinase K protection assay. Cropped immunoblots of proteinase K-treated (+) and untreated (−) organelle fractions are shown on the left panel, with the corresponding ratio of protein digestion on the right panel (See full-length immunoblots on Supplementary Fig. S4c) (c) Immunoelectron microscopy. Immunodecoration of mitosomes of HA-MBOMP30, HA-MBOMP30$^{\Delta 275-282}$, and control trophozoites with anti-HA (large gold particles) and anti-Cpn60 antibodies (small gold particles) showed electron-dense organelles surrounded by double membranes [scale bar, 200 nm].

including alanine and cysteine, x: any residue, see Materials and Methods)[27]. We also note that the β-signal almost never contains the secondary structure breaker proline in any position. As shown in Fig. 2, *E. histolytica* MBOMP30 has an appropriately placed match to the β-signal, but in *E. invadens* MBOMP30 the match is not perfect, because an alanine is aligned where a polar residue should occur. However, a few MBOMPs with imperfect matches to the β-signal are known, for example the *Saccharomyces pombe* Mdm10 β-signal (FFGVHFEY) has a phenylalanine residue in place of a polar residue in the first position[26].

Our immunoelectron microscopy results for HA-MBOMP30$^{\Delta 275-282}$ (HA-MBOMP30 lacking the β-signal) show its localization on the mitosomal membranes (Fig. 5c). Moreover, the gold anti-HA particle distribution in 13 randomly selected micrographs of HA-MBOMP30$^{\Delta 275-282}$ mitosomes is higher in the mitosomal membranes 13.4 ± 11.9 gold/μm^2 compared to 0.03 ± 0.07 gold/μm^2 in the cytoplasm ($p = 0.0016$, using Student's *t*-test). This data clearly suggests that membrane integration was unaffected even with the truncation of the putative β-signal, and therefore it is not required for outer membrane targeting per se. This observation is supported by the immunoblot of HA-MBOMP30$^{\Delta 275-282}$ cells fractions, showing that the protein was detected in the organelle fraction and was retained in the organellar membrane fraction even after treatment with Na_2CO_3 (Supplementary Fig. S4a).

MBOMP30 forms a ~240 kDa protein complex. We investigated whether MBOMP30 forms a complex, by immunoprecipitation of digitonin-solubilized organelle fraction of HA-MBOMP30. By performing Blue Native-PAGE of immunoprecipitated samples, followed by immunoblot analysis using anti-HA antibody, we detected a band suggestive of an MBOMP30-containing complex of approximately 240 kDa (Fig. 6a). We also tested if the complex is formed without the putative β-signal sequence. Interestingly, the ~240 kDa complex was not observed in either amino- or carboxy-terminal HA-tagged MBOMP30$^{\Delta 275-282}$ trophozoites. Furthermore, we also observed a similar phenomenon with immunoprecipitated HA-Tom40 and HA-Tom40$^{\Delta 275-284}$ from solubilized organelle fractions (Fig. 6b–c). This suggests that the β-signal may have a role in establishing stable MBOMP complex formation in *E. histolytica*.

Discussion

By a combination of *in silico* and experimental work we have identified a novel eukaryotic subclass of mitochondrion-related-organelle β-barrel proteins, found exclusively in the genus *Entamoeba*. MBOMP30 is the seventh MBOMP subclass, lacking any recognizable sequence homology to any of the six previously identified MBOMP subclasses. We have not determined the structure of MBOMP30 (e.g. via protein NMR or X-ray crystallography) and therefore our conclusion may not be considered as proven definitively, we have presented a considerable amount of experimental evidence to support our case. First, the *in silico* prediction of the secondary structure ratios by PSI-PRED (Table 1), was strongly corroborated by the deconvolution of the CD and far-UV spectroscopic data of MBOMP30 in DPhPc (Table 2), revealing a high ratio of β-sheet, consistent with the protein being a β-barrel. In a previous report, the CD spectral deconvolution analysis of human MBOMP hVDAC1 in DPhPC was also found to have high β-strand content[35], similar to what we observed in MBOMP30 and GFP. Second, Percoll gradient ultracentrifugation and Na_2CO_3 fractionation experiments showed MBOMP30 behaved like the representative *Entamoeba* MBOMP Tom40[39]. Third, imaging data, provided by immunofluorescence and immunoelectron microscopy, strongly indicate mitosomal, as well as mitosomal membrane localization of MBOMP30. Fourth, the membrane integration of MBOMP30 does not appear to be mediated by lipid modifications. It lacks a canonical amino-terminal secretory signal peptide, thus it is unlikely to go through the ER. Potential lipid modification sites including GPI attachment or isoprenylation are also absent. Highly reliable and well-benchmarked predictors of α-helical transmembrane regions such as Phobius[40] also predicted no such region in MBOMP30. Moreover, the synthesis of MBOMP30 by an *in vitro* wheat germ translation system demonstrated that it is spontaneously integrated into lipid bilayers (Supplementary Fig. S2), which was similarly observed in other integral mitochondrial membrane proteins such as the VDAC1 in *Homo sapiens*[35] and the dicarboxylate–tricarboxylate carrier in *Arabidopsis thaliana* and *Plasmodium falciparum*[41].

The mitosome of *Entamoeba* lacks most of the canonical processes and components existing in the mitochondrion and even the hydrogenosome[3]. In particular, the mitosomal outer and inner membranes appear almost bare, lacking homologs for most of the components associated with protein import. The assembly of BOMPs typically requires the outer membrane complexes TOM and SAM, containing the β-barrel proteins Tom40 and Sam50 respectively[6]. Recently, Tom60[39], a component unique to the *Entamoeba* TOM complex, was discovered. This essential tetratricopeptide repeat-containing protein acts as both a cytoplasmic carrier of soluble and membrane premitosomal proteins, as well as a lone structural component of the *Entamoeba* TOM complex. Aside from Tom60, the *Entamoeba* genome contains no detectable homologs to the non-β-barrel proteins which usually work in concert with Tom40 and Sam50[3,29], including the small Translocase of the Inner Membrane (small TIM) complexes Tim9-Tim10 and Tim8-Tim13 required for translocation of precursor BOMPs from the TOM to the SAM complex[6,7]. Also, there appears to be no homolog of Sam35 in *E. histolytica*, the non-BOMP subunit of the SAM complex which recognizes the β-signal in yeast prior to BOMP assembly via Sam50[38].

Interestingly, the *E. histolytica* MBOMP30 and Tom40 (Supplementary Fig. S5) match our slightly refined β-signal consensus sequence. However, the *E. histolytica* Sam50 homolog does not have anything close to a β-signal match, but instead has a terminal phenylalanine, which (although possibly coincidental) matches the membrane integration/assembly signal for bacterial BOMPs, simply consisting of a large aromatic residue [FYW] as the final residue of the protein[42] (Supplementary Fig. S5). This observation is intriguing in its own right, as Sam50 is an MBOMP with a recognizable bacterial BOMP homolog (Omp85/BamA), which itself typically ends in [FYW]. Given that the *Entamoeba* proteome seems to exhibit an unusual mixture of systems of diverse phylogenetic origin[18], it is conceivable that their MBOMPs utilize both eukaryotic and bacterial mechanisms for membrane insertion. The existence of MBOMPs that do not possess the β-signal motif like *E. histolytica* Sam50, suggests that there is a possibility of finding further candidates that also do not match the β-signal motif. Thus, we also performed an *in silico* screening using our MBOMP predictor without the β-signal related features. Unfortunately, this screening yielded only the known MBOMP EhSam50 and EHI_062770, a protein with a C2 domain forming a beta-sandwich fold.

To investigate the role of the putative *Entamoeba* β-signal in BOMP assembly we transfected amoebic trophozoites to overexpress MBOMP30 lacking the β-signal. Interestingly, our data from sodium carbonate fractionation and immunoelectron microscopy of HA-MBOMP30$^{\Delta 275-282}$ both indicate integration of the protein to the mitosomal membrane. From a survey of the immunoelectron micrographs of several randomly selected mitosomes, HA-MBOMP30 was observed to have double the number of anti-HA gold particles compared to the β-signal-truncated overexpressor. The decrease in the number of detected HA-particles may be due to a generally lower expression level of the protein lacking the β-signal, or the cytosolic degradation of the translated mutant protein after unsuccessful integration to the mitosomal membrane. Nevertheless, with regards to localization, the protein is highly concentrated on the mitosomal