

ANÁLISE *IN SILICO* DE CINCO ORFs DE *Saccharomyces cerevisiae*

Guedes, RF¹, Barros, MH², Nóbrega, FG¹

¹Universidade do Vale do Paraíba/ Instituto de Pesquisa e Desenvolvimento/ Laboratório de Genética Molecular e Genomas, Av. Shishima Hifumi, 2911, Urbanova, SJCampos, SP.

raquelfg@univap.br, fgdnobre@univap.br

²Universidade Estadual Paulista / Instituto de Biociências/ Departamento de Genética, Distrito de Rubião Júnior, s/nº, Botucatu, SP. mariohb@ibb.unesp.br

Resumo - Diversos estudos foram e estão sendo realizados com o intuito de se esclarecer o proteoma completo da mitocôndria, uma organela complexa presente nas células eucarióticas, responsável por diversos processos metabólicos. Um organismo considerado modelo nos estudos que envolvem a função respiratória celular é a levedura *Saccharomyces cerevisiae*. Neste trabalho foram selecionadas cinco ORFs de *S. cerevisiae*, estudadas apenas em trabalhos globais, cujas características fenotípicas estavam relacionadas à deficiência respiratória. Com o intuito de se obter informações sobre a função destas ORFs, foram realizados estudos *in silico* de localização protéica, utilizando-se quatro ferramentas computacionais distintas, todas disponíveis na internet, assim como uma análise dos domínios protéicos presentes nas seqüências de proteínas. Observou-se, de uma maneira geral, falta de congruência nos resultados obtidos em relação ao endereçamento protéico. Os resultados obtidos sobre os domínios protéicos foram bastante informativos. Apenas uma ORF não apresentou domínio protéico relacionado à atividade mitocondrial, sendo este resultado condizente com sua suposta localização protéica.

Palavras-chave: *Saccharomyces cerevisiae*, deficiência respiratória, ORF, *in silico*

Área do Conhecimento: II – Ciências Biológicas

Introdução

A mitocôndria é uma organela complexa presente nas células eucarióticas. Possui papel fundamental na montagem dos centros ferro/enxofre, na degradação oxidativa dos ácidos graxos, na apoptose [1,2] e na produção de energia da célula através da fosforilação oxidativa [3]. Possui seu próprio genoma, porém a maioria das proteínas mitocondriais é codificada por genes nucleares. Estas chegam na organela através de uma seqüência de endereçamento, localizada geralmente na porção n-terminal do peptídeo [4].

Saccharomyces cerevisiae é um fungo unicelular considerado modelo nos estudos da atividade respiratória celular, pois sobrevive à perda funcional da mitocôndria. Apesar do seu genoma ter sido completamente sequenciado em 1996, diversas ORFs (*open reading frames*) permanecem sem função definida.

Com base no trabalho de Steinmetz e colaboradores [5], foram selecionadas cinco ORFs de *S. cerevisiae* possivelmente envolvidas na respiração celular.

Neste trabalho foram feitas buscas *in silico* sobre o endereçamento e o domínio protéico das ORFs em questão com o objetivo de se analisar informações importantes que possam auxiliar na futura caracterização funcional destas proteínas.

Materiais e Métodos

As ORFs foram selecionadas com base nos resultados de Steinmetz e colaboradores [5] que se encontram disponíveis no portal *Yeast Deletion Project and Proteomics of Mitochondria Database* [6]. Todas fazem parte de um grupamento de ORFs cuja característica fenotípica é a deficiência respiratória. Essas ORFs têm sua função definida baseada apenas em estudos globais.

As seqüências das ORFs em nucleotídeos e em aminoácidos foram obtidas através do portal *Saccharomyces Genome Database* [7].

A análise das seqüências de endereçamento protéico foi realizada utilizando-se ferramentas distintas disponíveis nos seguintes portais: SGD [7], TARGETP v1.01 [8], PREDOTAR v1.03 [9], MITOPRED [10].

Para verificar os domínios protéicos, primeiramente as seqüências de aminoácidos das ORFs em questão foram confrontadas contra o banco de dados do NCBI [11] através da ferramenta BLASTP (comparação entre proteínas). Em seguida estes foram analisados com o auxílio da ferramenta *Conserved Domain Search* também localizada no banco de dados do NCBI.

Resultados

As ORFs selecionadas para este trabalho foram YDL119C, YDR115W, YDR332W localizadas no cromossomo IV, YHR009C, no cromossomo VIII e YOR022C no cromossomo XV.

Como descrito em Materiais e Métodos, a pesquisa sobre o endereçamento das proteínas relacionadas às ORFs em questão foi realizada em quatro bancos de dados. Os resultados obtidos estão na Tabela 1.

Tabela 1 – Análise computacional das seqüências de endereçamento protéico

	SGD	TargetP	Predotar	Mitopred
YDL 119C	Membrana interna mitocondrial	15,6% MT 13,6% SP 61,7% OT	0% MT 14%ER 86% OT	99% MT
YDR 115 W	Subunidade maior ribossomo mitocondrial	91,5% MT 3,9% SP 7,8% OT	57% MT 1% ER 42% OT	100% MT
YDR 332 W	Mitocondrial	73,9% MT 6,8% SP 17,2% OT	20% MT 1% ER 80% OT	99% MT
YHR 009C	Citoplasma	4,1% MT 12,2% SP 94,6% OT	0% MT 1% ER 99% OT	Não MT
YOR 022C	Mitocondrial	89,1% MT 5,2% SP 10,8% OT	87% MT 1% ER 13% OT	99% MT

MT – mitocondrial; SP – via secretora; OT – outra localização; ER – retículo endoplasmático

Ao confrontar a seqüência de aminoácidos da ORF YDL119C contra o banco de dados do NCBI foi obtida alta similaridade com uma proteína carreadora mitocondrial de *Candida albicans* (e-value $2e^{-60}$), fato que diz com os domínios protéicos presentes neste peptídeo, como pode ser observado na Figura 1.



Figura 1: Domínios protéicos presentes na ORF YDL119C, segundo o NCBI Conserved Domain Search

A ORF YDR115W mostrou similaridade com a proteína L34 ribossomal de *E. coli* (e-value $6e^{-46}$). Analisando os domínios protéicos presentes ao longo da seqüência (Figura 2) foi possível verificar que este peptídeo faz parte da família de proteínas ribossomais L34.

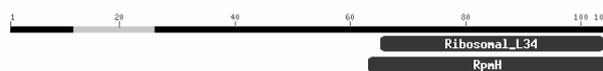


Figura 2: Domínios protéicos presentes na ORF YDR115W segundo o NCBI Conserved Domain Search

A Figura 3 refere-se aos domínios protéicos da ORF YDR332W, todos relacionados a uma suposta função de helicase.

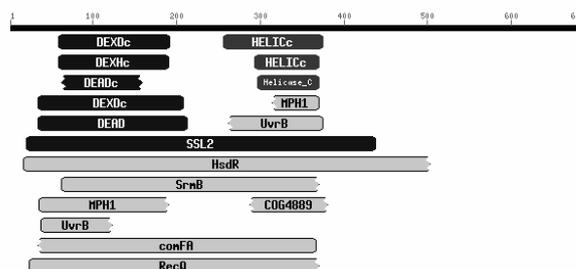


Figura 3: Domínios protéicos presentes na ORF YDR332W segundo o NCBI Conserved Domain Search

A ORF YHR009C apresentou similaridade com uma proteína citoplasmática putativa de *Cryptococcus neoformans var. neoformans JEC21* (e-value $3e^{-28}$). O domínio protéico *DadA* está relacionado ao metabolismo e transporte de aminoácidos (Figura 4).



Figura 4: Domínio protéico presente na ORF YHR009C segundo o NCBI Conserved Domain Search

Na Figura 5 é possível observar o domínio *DDHD* presente na ORF YOR022C. A ferramenta BLASTP também gerou resultado semelhante, onde houve similaridade da ORF em questão com o domínio *DDHD* de *Homo sapiens*, com um evalue de $2e^{-15}$.



Figura 5: Domínio proteico presente na ORF YOR022C segundo o *NCBI Conserved Domain Search*

Discussão

Através do portal *Yeast Deletion Project and Proteomics of Mitochondria Database* [6] que contém informações adicionais ao trabalho de Steinmetz e colaboradores [5], foi verificado que as ORFs selecionadas pertenciam a um grupo de 466 ORFs cuja característica fenotípica era a deficiência de crescimento em meios não fermentáveis. Destas, 201 (43%) codificam para proteínas sabidamente mitocondriais de acordo com a função ou localização das mesmas, sendo que a ORF YDL119C está contida neste grupo, fato bastante interessante, pois se mostra um atraente objeto de pesquisa. 35% das ORFs (161) mostram-se como possuindo localização subcelular desconhecida e na maioria dos casos, função não conhecida. Nesta situação encontram-se as demais ORFs selecionadas para este estudo (YDR115W, YDR332W, YHR009C e YOR022C), o que também é bastante animador, já que estas não estão incluídas entre as 104 ORFs (22%) que possuem funções não relacionadas à mitocôndria, como transporte de íons, transcrição, sinalização proteica, entre outras.

Apesar da ORF YDL119C ser uma proteína mitocondrial, segundo o portal *Yeast Deletion Project and Proteomics of Mitochondria Database* [6], os resultados do programa TARGETP e PREDOTAR não condizem com este fato. Já as informações fornecidas pelo SGD e MITOPRED indicam ser uma proteína com endereçamento mitocondrial, assim como o domínio proteico (*Mito_carr*) e o resultado do BLASTP.

Os resultados obtidos em relação ao endereçamento proteico da ORF YDR115W concordam entre si, indicando uma suposta localização proteica mitocondrial. O domínio proteico *Ribossoma_L34* assim como o resultado do BLASTP são condizentes entre si, indicando a provável função desta ORF.

Os resultados de endereçamento proteico da ORF YDR332W, com exceção do fornecido pelo programa PREDOTAR indicam ser um peptídeo mitocondrial. Todos os domínios proteicos encontrados para esta ORF estão relacionados a uma suposta função de helicase. O domínio HELICc (e-value do domínio: $3e^{-12}$) compreende uma superfamília de helicases de domínio c-terminal, associada com proteínas DEXDc, DEAD e DEAH. O domínio de maior e-value ($1e^{-69}$) é o SSL2, correspondente a uma DNA/RNA helicase da superfamília II. O BLASTP também resultou em

similaridade com o DEAD ($3e^{-66}$). Shiratori e colaboradores [12] realizaram um estudo de identificação, classificação e caracterização sistemática de ORFs relacionadas à atividade de helicase, em *Saccharomyces cerevisiae*, onde a ORF YDR332W foi citada.

Com a ORF YHR009C ocorreu um fato curioso, já que houve um consenso entre todas as ferramentas utilizadas de que a proteína em questão não deve ser direcionada para a mitocôndria, assim como o resultado fornecido pelo BLASTP, porém no trabalho de Steinmetz e colaboradores [5] ela está contida no grupo de 161 ORFs que não possuem classificação quanto à localização, e não no grupo de 104 ORFs sabidamente não mitocondriais. O domínio proteico *DadA* está relacionado ao metabolismo e transporte de aminoácidos (NCBI).

Para a ORF YOR022C houve um consenso entre os resultados da análise da seqüência de direcionamento da proteína, segundo os quais esta deve ser direcionada à mitocôndria. O domínio *DDHD* contém quatro resíduos conservados que formam um sítio de ligação com metal, o que pode ter aplicação na mitocôndria.

É importante ressaltar que o programa TARGETP detecta apenas seqüências N-terminais, com uma fidelidade de cerca de 85%. Com o PREDOTAR é possível verificar o putativo endereçamento proteico para a mitocôndria, plastídeo ou retículo endoplasmático; assim como o TARGETP detecta apenas endereçamento N-terminal. Segundo os autores, os resultados refletem uma baixa taxa de falsos positivos. O MITOPRED prediz o endereçamento proteico apenas mitocondrial, para todos os organismos, incluindo plantas. A performance desta ferramenta é superior a 0,73, enquanto do TARGETP é 0,51 [10]. Porém mesmo com estas informações sobre a confiabilidade dos programas foi verificado falta de congruência em diversos resultados, fato que pode ser explicado tanto por falha computacional, quanto pela falta de consenso entre as seqüências de endereçamento n-terminal assim como a sinalização pode estar difundida ao longo da seqüência, sendo assim estes programas não capazes e detectá-las.

Conclusão

Este trabalho vem ratificar a importância das pesquisas *in silico* no intuito de direcionar e embasar os experimentos práticos, no entanto, é preciso ressaltar que a identificação experimental das características fenotípicas individuais mostram-se necessárias para a real confirmação das indicações do estudo global de Steinmetz e colaboradores e deste presente trabalho, onde foram utilizadas ferramentas de bioinformática.

Referências

- [1] LILL, R.; KIISPAL, G. Maturation of cellular Fe-S proteins an essential function of mitochondria. **Trends Biochem Sci.** V. 25, n.2, p. 352-356, 2000
- [2] NEWMAYER, D. D.; FERGUSON-MILLER, S. Mitochondria: Releasing Power for Life and Unleashing the Machineries of Death. **Cell.** V. 112, p.481-490, 2003.
- [3] SARASTE, M. Oxidative Phosphorylation at the fin de siècle. **Science.** V.283, p.1488-1493, 1999.
- [4] TRUSCOTT, K.N.; PFANNER, N.; BRANDER, K. Mechanisms of protein import into mitochondria. **Current Biology.** V.13, p.326-337, 2003.
- [5] STEINMETZ, L.M.; SCHARFE, C.; DEUTSCHBAUER, A.M.; MOKRANJAC, D.; HERMAN, Z.S.; JONES, T.; CHU, A.M.; GIAEVER, G.; PROKISCH, H.; OEFNER, P.J.; DAVIS, R.W. Systematic screen for human disease genes in yeast. **Nature Genetics.** V.31, n.4, p.400-404, 2002.
- [6] YDPM (Yeast Deletion Project and Proteomics of Mitochondria Database). Internet site address http://www-deletion.stanford.edu/YDPM/YDPM_index.html. Acessado em 05/02/2005
- [7] SGD (Saccharomyces Genome Database). Internet site address <http://www.yeastgenome.org/>. Acessado em 01/03/2005.
- [8] EMANUELSSON, O.; NIELSEN, H.; BRUNAK, S.; HEIJNE, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. **Journal of Molecular Biology.** V.300, p.1005-1016, 2000. TARGETP Internet site address <http://www.cbs.dtu.dk/services/TargetP/>. Acessado em 01/03/2005.
- [9] PREDOTAR (A prediction service for identifying putative N-terminal targeting sequences). Internet site address <http://genoplante-info.infobiogen.fr/predotar/predotar.html>. Acessado em 01/03/2005.
- [10] GUDA C.; FAHY E.; SUBRAMANIAM S. MITOPRED: A genome-scale method for prediction of nuclear-encoded mitochondrial proteins. **Bioinformatics,** V. 20, p.1785-1794, 2004. MITOPRED (A genome-scale method for predicting mitochondrial proteins). Internet site address <http://mitopred.sdsc.edu/>. Acessado em 01/03/2005.
- [11] NCBI (National Center for Biotechnology Information). banco de dados. Internet site address <http://www.ncbi.nlm.nih.gov/>. Acessado em 05/02/2005.
- [12] SHIRATORI, A.; SHIBATA, T.; ARISAWA, M.; HANAOKA, F.; MURAKAMI, Y.; EKI, T. Systematic identification, classification, and characterization of the open reading frames which encode novel helicase-related proteins in *Saccharomyces cerevisiae* by gene disruption and Northern analysis. **Yeast.** V.15, n.3, p.219-253, 1999.