

UNIVERSIDADE ESTADUAL PAULISTA “Júlio de Mesquita Filho”

Instituto de Química

Programa de Pós-Graduação em Biotecnologia

Fatores de transcrição reguladores do metabolismo do glicogênio em *Neurospora crassa*. Caracterização parcial e identificação de alvos de ligação por CHIP-Seq

Rodrigo Duarte Gonçalves

RODRIGO DUARTE GONÇALVES

Fatores de transcrição reguladores do metabolismo do glicogênio em *Neurospora crassa*. Caracterização parcial e identificação de alvos de ligação por CHIP-Seq

Tese apresentada ao Instituto de Química, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Doutor em Biotecnologia.

Orientadora: **Prof^a. Dr. Maria Célia Bertolini**

ARARAQUARA

2012

Dados curriculares

Rodrigo Duarte Gonçalves

1. Dados pessoais

Nascimento: 12 de agosto de 1981 – São Paulo-SP, Brasil.

Nacionalidade: Brasileira

Filiação: Hermes Gonçalves e Maria Laura Duarte Gonçalves

Documento de Identidade: 30.596.508-6 / SSP-SP

Cadastro de Pessoa Física: 304.079.778-64

Endereço Pessoal: Rua XV de Novembro, 295 – Centro.

CEP: 19700-000 – Paraguaçu Paulista – SP

Endereço Profissional: Instituto de Química de Araraquara – UNESP

Departamento de Bioquímica e Tecnologia Química

Rua Prof. Francisco Degni 55 – 14800-900 Araraquara-SP

2. Formação acadêmica

Doutorado em Biotecnologia: Programa de Pós-Graduação em Biotecnologia, área de concentração Biotecnologia, Instituto de Química de Araraquara, UNESP, Araraquara, São Paulo (em andamento).

Mestrado em Biotecnologia: Mestrado em Biotecnologia, Programa de Pós-Graduação em Biotecnologia, área e concentração Biotecnologia, Instituto de Química de Araraquara, UNESP, Araraquara, São Paulo, concluído em 26 de junho de 2008.

Bacharel e Licenciado em Ciências Biológicas: Faculdade de Ciências e Letras de Assis, Universidade Estadual Paulista, UNESP, São Paulo, Brasil, concluído em 13 de janeiro de 2006.

3. Publicações

BARBOSA, L. C. B.; GARRIDO, S. S.; GARCIA, A.; GONÇALVES, R. D.; MARCHETTO, R. Easy Bioinformatics Analysis (EBiAn): a package for manipulating and analysis of short biological sequences. *Bioinformatics* (2010) v.5, p. 46-48.

GONÇALVES, R. D.; CUPERTINO, F. B.; FREITAS, F. Z.; LUCHESSI, A. D.; BERTOLINI, M. C. A Genome-wide screen for *Neurospora crassa* transcription factors regulating glycogen metabolism. *Molecular & Cellular Proteomics* (2011) 10 (11), M111.007963-M111.007963.

BERTOLINI, M. C.; FREITAS, F. Z.; de Paula, R. M.; CUPERTINO, F. B.; GONÇALVES, R. D. Glycogen metabolism regulation in *Neurospora crassa*. In: Biocommunication of Fungi Guenther Witzany (ed.), Springer Science+Business Media B.V., 2012, 10.1007/978-94-007-4264-2_3.

TAKEDA, A. A. S.; FREITAS, F. Z.; MAGRO, A. J.; BERNARDES, N. E.; FERNANDES, C. A. H.; GONÇALVES R. D.; BERTOLINI, M. C.; FONTES, M. R. M. Biophysical characterization of the recombinant importin- α from *Neurospora crassa*. *Protein and Peptide Letters* (2012), *in press*

“A ciência nunca resolve um problema sem criar pelo menos outros dez.”

George Bernard Shaw

Dedico este trabalho aos meus pais,
Hermes e Maria Laura

Agradecimentos

Gostaria de agradecer meus pais pelo carinho e apoio em todos os momentos.

A Tainara pela companhia, apoio e pela compreensão da minha ausência.

A Prof^a. Maria Célia Bertolini pela confiança e pela bagagem adquirida ao longo de minhas atividades em seu laboratório.

Ao Departamento de Bioquímica de Tecnologia Química e ao Programa de Pós-Graduação em Biotecnologia da UNESP, pela oportunidade de realizar meu Doutorado.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelas bolsas de Doutorado e Doutorado Sanduiche (PDEE) concedidas.

Ao Professor Dr. Michael Freitag por todo suporte que obtive no período que estive em seu laboratório.

Aos meus amigos de laboratório Thiago, Fernanda Freitas, Fernanda Cupertino, Eliane, Carol, Stela e Susi, Heloisy, Erick.

Aos amigos que passaram pelo laboratório Messias, Luiz, Flávia, Ana Paula, Amanda, Monica, Eduardo, Juliana, Margareth e Vivia.

Ao Kyle, Jonathan, Pallavi, Lanelle e especialmente a Erin pela amizade e auxílio prestado nas análises de ChIP-Seq durante meu Doutorado Sanduíche.

Aos amigos, Tarcísio, Waldenir, Fátima, Zilda e Fernando pelo suporte e colaboração e amizade.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo suporte financeiro concedido ao laboratório.

Lista de abreviaturas

5'- e 3'-UTR:	5'- e 3'- "Untranslated Region"
ADP:	Adenosina difosfato
AMP:	Adenosina monofosfato
ATP:	Adenosina trifosfato
bHLH	<i>Basic helix-loop-helix</i>
BSA:	Soroalbumina bovina
cDNA:	Ácido desoxirribonucleico complementar
com	Cintilações por minuto
C-terminal:	Carboxi-terminal
D.O.:	Densidade ótica
dATP:	Desoxiadenosina trifosfato
DEPC:	Dietil pirocarbonato
DNA:	Ácido desoxirribonucleico
dNTPs:	Deoxiribonucleosídeos 5'-trifosfatados (dATP, dCTP, dTTP, dGTP)
DTT:	Ditiotreitol
EDTA:	Ácido etilenodaminotetracético
FGSC.	<i>Fungal Genetics Stock Center</i>
GSN:	Glicogênio sintase de <i>Neurospora crassa</i>
G1P:	Glicose-1-fosfato
G6P:	Glicose-6-fosfato
His:	Histidina
HSE	<i>Heat shock elements</i>
Kb:	kilobase
kDa:	kilodalton
MOPS:	Ácido 3-(N-morfolino) propanosulfônico
mRNA:	RNA mensageiro
N-terminal:	Amino-terminal
ORF:	<i>Open reading frame</i>
Pb:	pares de bases
PCR:	<i>Polymerase Chain Reaction</i>

PKA:	Proteína quinase dependente de AMPc
PSMF:	Fluoreto de metilfenilsulfonil
RNA:	Ácido ribonucléico
RNAse A:	Ribonuclease pancreática A
RPM:	Rotações por minuto
SDS:	Dodecil sulfato de sódio
Ser:	Serina
SSPE	Solução salina fosfato-EDTA
TAE:	Tampão Tris-acetato-EDTA
TCA:	Ácido tricloroaético
TLCK:	N-p-tosil-L-lisina clorometilcetona
Tris:	Hidroximetilaminometano
Tyr:	Tirosina
VM	Meio mínimo de Vogel
UDP:	Uridina difosfato
UDPG:	Uridina difosfato glicose
UTR:	<i>Untranslated region</i>

Resumo

Resultados prévios, em nosso laboratório, permitiram identificar fatores de transcrição provavelmente atuando como reguladores do metabolismo de glicogênio no fungo *Neurospora crassa* utilizando uma coleção de linhagens mutantes do fungo contendo ORFS codificadoras de fatores de transcrição individualmente nocauteadas. O presente trabalho teve como objetivo realizar análises funcionais de alguns dos fatores de transcrição anteriormente identificados, tais como o regulador transcricional XlnR, descrito em outros fungos filamentosos como regulador de genes codificadores de enzimas xilanolíticas e/ou celulolíticas e o produto da ORF NCU04390, uma proteína sem função conhecida. Inicialmente foram realizadas análises de acúmulo de glicogênio e da expressão dos genes *gsn* (codificador da enzima glicogênio sintase) e *gpn* (codificador da enzima glicogênio fosforilase) durante o crescimento vegetativo (30°C) e sob condição de estresse térmico (45°C) nas linhagens mutantes e comparadas à linhagem selvagem. Ambas as linhagens mutantes apresentaram alterações tanto no conteúdo de glicogênio como na expressão dos genes citados anteriormente, o que indicou o possível envolvimento dos fatores de transcrição na regulação do metabolismo do carboidrato. Foram realizados experimentos com o objetivo de investigar a função do fator de transcrição XlnR na regulação do metabolismo de glicogênio. Foi verificado que esta proteína regula os níveis do carboidrato, bem como a expressão do gene *gsn* quando o fungo foi crescido em fontes alternativas de carbono. Entretanto, o fator de transcrição XLR-1 de *N. crassa* produzido na forma recombinante em *E. coli* (inteiro e truncado) não foi capaz de ligar ao *motif* presente no promotor *gsn*. Genes codificadores de duas endoxilanasases foram identificados no genoma de *N. crassa* e as análises de expressão gênica nas linhagens selvagem e XLR-1^{KO} mostraram que os níveis dos transcritos codificadores de ambas as endoxilanasases foram expressos somente na presença de xilana, e não na presença de xilose, e sugeriram que o fator de transcrição XLR-1 pode não atuar sozinho na ativação de genes codificadores de endoxilanasases. O fator de transcrição codificado pela ORF NCU04390 foi caracterizado por ChIP-Seq com o intuito de mapear no genoma do fungo genes alvos possivelmente regulados por esta proteína, portanto determinar o provável *motif* de DNA no qual esta proteína interage. Estas análises mostraram que genes envolvidos em diversos processos celulares são possivelmente regulados por esta proteína, sendo a maioria deles relacionada com o metabolismo de carbono. Para a confirmação desta análise foram realizados ensaios de interação DNA-proteína *in vitro*, mostrando que o produto gênico produzido na forma recombinante em *E. coli* foi capaz de interagir com o *motif* de DNA 5'-TACATA-3' presente na região 5'-flanqueadora do gene codificador da enzima desramificadora do glicogênio. Este resultado mostrou um papel regulador do fator de transcrição na regulação do metabolismo de glicogênio através da regulação do gene codificador da enzima desramificadora.

Abstract

Previous results from our laboratory using a collection of *Neurospora crassa* strains mutated in genes encoding transcription factors allowed us to identify transcription factors likely involved in glycogen metabolism regulation. The present work aimed to perform a functional analysis of some transcription factors previously identified, such as the transcriptional regulator XInR, described in many filamentous fungi as a transcriptional regulator of xylanolytic and/or cellulolytic genes and a protein with unknown function, the ORF NCU04390 product. Initially, analysis of glycogen accumulation and *gsn* (coding glycogen synthase) and *gpn* (coding glycogen phosphorylase) gene expression were performed under vegetative growth (30 °C) and under heat shock condition (45 °C) in the two mutant strains and compared to the wild-type strain. Both mutant strains showed changes in the glycogen levels as well as in the *gsn* and *gpn* expression, suggesting an involvement of these transcription factors in the regulation of glycogen metabolism. Additional analysis was performed in order to investigate the function of the XInR transcription factor. It was observed that this transcription factor regulates the glycogen levels as well as the *gsn* expression under alternative carbon sources. However, the recombinant XLR-1 transcription factor produced in *E. coli* (entire and truncated proteins) was not able to bind to the XLR-1 motif present in the *gsn* promoter. Two endoxylanases coding genes were identified in the *N. crassa* genome database and analysis of their gene expression showed the upregulation of both genes during growth in xylan instead of xylose as carbon sources. This result suggested that the XLR-1 transcription factor is not the only protein that regulates the endoxylanases genes. The transcription factor encoded by NCU04390 ORF was functionally characterized using the ChIP-Seq analysis in order to identify in the *N. crassa* genome genes likely regulated by this transcription factor. This analysis showed that genes involved in different cellular process could be regulated by this protein, most of them related to carbon metabolism. Analysis of DNA-protein binding by EMSA was performed in order to confirm the ChIP-Seq data. The results showed that the NCU04390 gene product, produced as a recombinant protein in *E. coli*, bound to the 5'-TACATA-3' DNA motif existent in the glycogen debranching enzyme gene promoter. This result indicated a regulatory role of this transcription factor in the glycogen metabolism by regulating the gene coding glycogen debranching enzyme.

Sumário

Introdução.....	17
1. O fungo <i>Neurospora crassa</i>	18
2. Glicogênio. Estrutura e metabolismo	22
3. Controle do metabolismo do glicogênio	24
4. Estrutura da cromatina e fatores de transcrição	27
5. O fator de transcrição XlnR em fungos filamentosos	34
Objetivos.....	37
Materiais e Métodos.....	39
1. Linhagens.....	40
1.1 <i>Neurospora crassa</i>	40
1.2. <i>Escherichia coli</i>	40
2. Meios de cultivo	40
2.1. <i>Neurospora crassa</i>	40
2.1.1. Meio mínimo de Vogel (VM)	40
2.1.2. Meio utilizado pra cruzamento (SC)	41
2.1.3. <i>Top ágar</i>	41
2.1.4. <i>Base ágar</i>	42
2.2. <i>Escherichia coli</i>	42
2.2.1. 2YT + 0,2% de glicose	42
2.2.2. <i>Terrific Broth</i> (TB)	42
3. Propagação das linhagens de <i>N. crassa</i> e obtenção dos conídios	42
4. Choque térmico	42
5. Crescimento do fungo utilizando pentoses como fontes de carbono.....	43
6. Preparo do extrato celular e determinação do conteúdo de glicogênio	43
7. Análise da expressão gênica	44
7.1. Extração de RNA	44
7.2. Análise da expressão gênica por <i>Northern blot</i>	44
8. Análise das linhagens mutantes por citometria de fluxo	46
9. Clonagem molecular das ORFs NC06971 e NCU04390 por RT-PCR	46
10. Transformação de células competentes de <i>E. coli</i>	48
11. Mini extração de DNA plasmidial	48
12. Produção de proteínas recombinantes em <i>E. coli</i>	49
12.1. Subclonagem das ORFs em vetores de expressão	49
12.2. Ensaio de indução e análise de solubilidade	49

12.3. Análise das proteínas recombinantes por <i>Western blot</i>	50
12.4. Purificação das proteínas recombinantes por cromatografia de afinidade.....	50
13. Determinação dos sítios de ligação ao DNA por CHIP-Seq.....	51
13.1. Preparo dos <i>split markers</i> para a produção de proteínas fusionadas a GFP.....	51
13.2. Transformação de conídios do fungo <i>N. crassa</i>	53
13.3. Extração do DNA genômico.....	54
13.4. Confirmação por <i>Southern blot</i>	55
13.5. Condições de hibridização e detecção.....	55
13.6. Cruzamento das linhagens de <i>N. crassa</i> e seleção dos transformantes	56
13.7. Imunoprecipitação da cromatina (ChIP).....	56
13.8. Preparo da biblioteca de CHIP-Seq.....	58
13.9. Sequenciamento de DNA	59
13.10. Análise computacional dos resultados de CHIP-Seq.....	59
14. <i>Electrophoretic mobility shift assay</i> (EMSA).....	60
14.1. Obtenção e marcação das sondas.....	60
14.2. Ensaio de retardamento em gel	62
Resultados.....	63
1. Fatores de transcrição selecionados	64
2. Análise da expressão dos genes <i>gsn</i> e <i>gpn</i> por <i>Northern blot</i> nas linhagens mutantes	64
3. Análise das linhagens mutantes por citometria de fluxo	67
4. Produção e purificação de proteína recombinante	68
4.1. Ensaio de indução e análise de solubilidade das proteínas recombinantes	68
4.2. Produção e purificação de proteínas recombinantes	71
5. Caracterização funcional da proteína XLR-1 no fungo <i>N. crassa</i>	79
5.1. Conteúdo de glicogênio e expressão do gene <i>gsn</i> e das ORFs codificadoras de endoxilanasas em diferentes fontes de carbono	81
5.2. Análise de ligação DNA-proteína por EMSA	84
6. Caracterização funcional da proteína hipotética codificada pela ORF NCU04390	85
6.1. Introdução do <i>tag</i> GFP na região C-terminal da proteína 4390	85
6.2. Imunoprecipitação da cromatina.....	88
6.3. Sequenciamento de alto desempenho.....	90
6.4. Determinação do <i>motif</i> de DNA para a proteína 4390.....	91
6.5. Análise das regiões 5' e 3'-flanqueadoras dos genes envolvidos no metabolismo de glicogênio para <i>motifs</i> da proteína 4390	94
6.6. Ensaio de ligação com a proteína His4390 recombinante por EMSA.....	95
Discussão.....	100
1. O fator de transcrição XLR-1 e regulação do metabolismo do glicogênio	102

2. Caracterização funcional da proteína hipotética codificada pela ORF NCU04390	104
Referências Bibliográficas	107
Anexo I	117
Anexo II	131
Anexo III	150

Introdução

1. O fungo *Neurospora crassa*

Os organismos pertencentes ao reino dos fungos têm exercido um papel fundamental no avanço da compreensão dos mecanismos moleculares de organismos eucariotos devido às suas facilidades de manipulação bem como o conhecimento das características genéticas e bioquímicas envolvidas em seus respectivos ciclos de vida. O fungo *Neurospora crassa*, com um legado de mais de 70 anos de intensa pesquisa, é provavelmente um dos fungos filamentosos mais bem conhecidos. Pesquisas utilizando *N. crassa* como modelo nas primeiras décadas do século XX determinaram o destino da genética moderna e da biologia molecular. Desde os anos 1940, o fungo *N. crassa* vem sendo intensamente utilizado como organismo modelo em estudos de expressão gênica, desenvolvimento e diferenciação celular, ritmo circadiano, defesa do genoma, bem como outros aspectos da biologia de eucariotos (Perkins & Davis 2000).

Estudos iniciais com o fungo *N. crassa* datam da primeira metade do século XIX, onde o fungo foi descrito e utilizado em experimentos por microbiologistas franceses após sua infestação em padarias de Paris (PAYEN, 1843; MONTAGNE, 1843). Inicialmente denominada de *Monilia sitophila*, o gênero foi renomeado logo após a identificação de corpos de frutificação, demonstrando a ocorrência de um estágio sexual no fungo até então desconhecido (SHEAR & DODGE, 1927). Nos anos 30, uma série de análises realizadas por Lindegren confirmaram os estudos iniciais de Dodge, fazendo com que o fungo *N. crassa* começasse a ser utilizado em estudos de genética clássica. Tais estudos determinaram as bases genética do “fator de acasalamento” e estabeleceu os primeiros de mapas de *linkage* do fungo *N. crassa* (DAVIS, 2000). Na década de 40, partindo do pressuposto de que o desenvolvimento e funcionamento de um organismo consistia essencialmente de um sistema integrado de reações químicas controladas de alguma maneira pelos genes, uma série de experimentos utilizando o fungo *N. crassa* foram realizados por Beadle e Tatum (BEADLE & TATUM, 1941), o que resultou na hipótese “um gene, uma enzima”. Como reconhecimento, os autores foram laureados, em 1958, com o Prêmio Nobel de Medicina e Fisiologia.

N. crassa é um fungo filamentoso pertencente à classe dos *Eumicetos*, subclasse *Ascomiceta*, família *Sordariacea*, subfamília *Sphaeriales* (ESSER & KUENEN, 1967). É um organismo haploide, constituído por hifas cenocíticas multinucleadas que se anastomosam formando o micélio. Essas hifas possuem poros que permite a passagem de núcleos e organelas entre elas, fazendo com que múltiplos núcleos contribuam para o metabolismo do micélio. Ao todo, o fungo apresenta pelo menos 28 tipos distintos de células, muitos dos quais derivam das hifas, que contribuem para seu complexo ciclo de vida (BISTIS et al., 2003, BORKOVICH et al., 2004). Além disso, *N. crassa* é um organismo heterotático,

possuindo dois tipos de fator de acasalamento designados “a” e “A”, saprófito e não patogênico, embora seu genoma esteja intimamente relacionado com genomas de fungos fitopatogênicos (GALAGAN *et al.*, 2003). Apresenta tanto o desenvolvimento sexual quanto o assexual em seu ciclo de vida, o qual é altamente influenciado por fatores do meio ambiente, como disponibilidade de nutrientes, luz e temperatura. Uma representação esquemática do ciclo do fungo pode ser visualizada na figura 1.

O ciclo assexual do fungo consiste na formação de estruturas denominadas de macro e microconídios. Macroconídios são células grandes, multinucleadas (3 a 6 núcleos) que são originados ou por brotamentos repetidos a partir da célula apical de conidióforo (blastoconídio) ou pela fragmentação do conidióforo (artroconídio) (SPRINGER & YANOFSKY, 1989). Estas estruturas são revestidas por proteínas fazendo com que os macroconídios assumam uma natureza hidrofóbica, o que os mantém secos e leves facilitando sua dispersão pelo ar (DAVIS, 2000). Além dos macroconídios, existe também os microconídios, os quais apresentam uma estrutura diferenciada em relação aos macroconídios. Eles são menores, possuem natureza hidrofílica, possuem apenas um núcleo, são liberados diretamente de estruturas constituídas por pequenas células denominadas de microconidióforos. Por serem liberados mais tarde que os macroconídios, os microconídios assumem a função de agente fertilizador masculino (MAHESHWARI, 1999). Em geral, os conídios germinam em condições adequadas para o crescimento pela formação de um tubo germinativo inicial, que se estende até formar uma hifa (Davis, 2000).

Para que ocorra o ciclo sexual, além da presença de estruturas de ambos os fatores de acasalamento, é necessário que o fungo esteja em condição de limitações de nitrogênio, o que induz a formação de uma estrutura sexual especializada denominada de protoperitécio (órgão sexual feminino). Após a formação do protoperitécio ocorre a formação de hifas especializadas denominadas de tricógino (órgão sexual masculino), o qual é atraído por feromonios em direção ao protoperitécio. Após o contato entre ambas as estruturas ocorre a fertilização do protoperitécio pelo tricógino por meio da fusão das hifas e transferência do núcleo proveniente do tricógino formando uma estrutura binucleada denominada de ascogônia (RAJU, 1980). Após este evento ocorre uma mitose na qual resulta na formação de uma estrutura denominada de *crozier* dividida em 3 células, uma lateral e uma célula basal com um núcleo cada, e uma célula apical com dois núcleos. É a partir desta célula apical que ocorrerá a formação do ascósporo. Este processo se inicia após a cariogamia, a qual é seguida de uma meiose, resultando na formação de 4 células filhas haploides no interior do asco jovem. Em seguida, ocorre uma mitose para a formação de 8 células filhas mononucleadas, dispostas linearmente (tétrades) no interior dos ascos (RAJU, 1980).

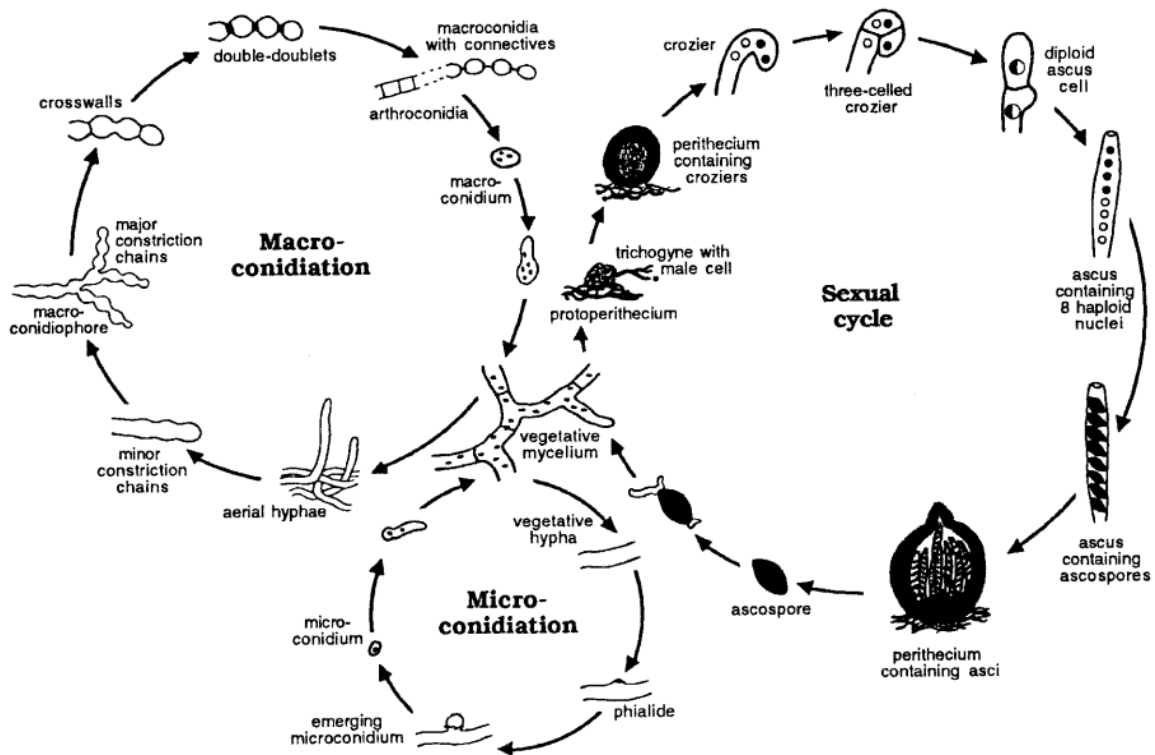


Figura 1 - Ciclo de vida do fungo *Neurospora crassa* (BORKOVICH et al., 2004).

O genoma de fungo está organizado em sete cromossomos que variam entre 4 a 10 Mb de tamanho (SCHUTLE *et al.*, 2002). O sequenciamento de seu genoma (GALAGAN et al. 2003) revelou ser constituído por cerca de 40 Mb, muito maior se comparado com outros fungos com genomas já conhecidos, como *Sacharomyces cerevisiae* e *Schizosacharomyces pombe*, os quais possuem genomas com aproximadamente 12 Mb. *N. crassa* possui um total de 10.082 genes codificadores de proteínas, que cobrem 44% do genoma. Do total de genes anotados, 9.200 codificam proteínas com mais de 100 resíduos, cerca de duas vezes o total de genes de *S. pombe* (aproximadamente 4.800 genes) e *S. cerevisiae* (aproximadamente 6.300 genes) (GALAGAN et al., 2003). Um fato que chama a atenção é que 41% dos genes que codificam proteínas não possuem ortólogos conhecidos depositados nos bancos de dados, refletindo o baixo conhecimento dos genomas dos fungos. Além disso, 57% das proteínas de *N. crassa* não possuem relação a proteínas conhecidas de leveduras. Quando comparado com genomas conhecidos de eucariotos, 1.421 genes de *N. crassa* apresentaram as melhores correlações em relação a proteínas de plantas ou animais (GALAGAN et al., 2003). Isso sugere uma relação mais estreita entre fungos filamentosos e eucariotos superiores, se comparado com leveduras e outros eucariotos inferiores. O fato de *N. crassa* ser um organismo multicelular faz com que uma

grande quantidade de seus genes não possui ortólogos em *S. cerevisiae*, sugerindo que *N. crassa* pode ser um organismo mais adequado para a compreensão dos aspectos de regulação celular dos eucariotos superiores (DUNLAP et al., 2007). A presença de um grande número de genes no genoma de *N. crassa* ainda com funções desconhecidas aponta este organismo como um promissor modelo para o estudo de novos mecanismos genéticos e bioquímicos não identificados até o momento, os quais podem possivelmente representar genes com novas funções ou novos genes desempenhando maneiras diferentes de realizar funções conhecidas (DUNLAP et al., 2007).

Em eucariotos existem dois mecanismos de recombinação gênica, a homóloga que envolve a quebra da dupla fita de DNA seguida da substituição entre sequências de DNA homólogas e a recombinação não homóloga que envolve a ruptura da dupla fita de DNA, seguida da inserção do fragmento exógeno em local aleatório, independente da homologia. A recombinação homóloga é o mecanismo predominante em *S. cerevisiae*, representando o método mais eficiente pelo qual se pode inativar um gene de interesse. Apesar de ocorrer com muita frequência em leveduras, o mecanismo de recombinação homóloga ocorre em raríssimas ocasiões tanto em *N. crassa* como nos demais eucariotos, dependendo do gene e do tamanho da região nucleotídica homóloga que flanqueia o gene de interesse (DUNLAP et al., 2007). Já o mecanismo de recombinação gênica não homóloga é conservado evolutivamente em todos os reinos, sendo o mecanismo predominante em eucariotos superiores, como humanos, plantas e insetos (NINOMIYA et al., 2004). O conjunto de proteínas envolvidas neste mecanismo são as proteínas Ku70 e Ku80, conservadas tanto em leveduras quanto em humanos, que se ligam nas extremidades do DNA e o complexo DNA ligase IV-XRCC4.

Uma coleção de mutantes em todos os genes do genoma é uma fonte extremamente poderosa no estudo da genômica funcional de um organismo. Sendo assim, a interrupção sistemática de todas as ORFs utilizando mecanismos de inativação gênica é uma abordagem fundamental para a compreensão do papel de cada uma das ORFs contidas no genoma do fungo *N. crassa*, onde muitas das quais ainda não têm uma função atribuída. Com a identificação de genes homólogos aos genes codificadores das proteínas Ku70 e Ku80 no genoma de *N. crassa*, *mus-51* e *mus-52*, respectivamente, Ninomiya e colaboradores (2004) desenvolveram um procedimento, baseado na técnica de *Fusion-PCR*, para a construção dos cassetes de DNA. Tais cassetes continham o gene codificador da proteína higromicina B fosfotransferase (*hph*), conferindo ao fungo resistência ao antibiótico higromicina B. Sabe-se que a linhagem selvagem do fungo *N. crassa*, assim como outros fungos filamentosos apresenta uma taxa muito pequena de recombinação homóloga (<10%) (PAIETTA & MARZLUF, 1985). Entretanto, Ninomyia et al. (2004) mostraram que deleções únicas nos genes *mus-51* ou *mus-52*, os quais estão envolvidos em processos de

recombinação não homóloga, promovem um aumento expressivo nos níveis de recombinação homóloga (>90%).

O nocaute de todos os genes de *N. crassa* tem como objetivo determinar a função de cada deles, pois apenas 15% dos genes são associados com o fenótipo e pouco mais de um terço destes genes não possuem ortólogos em outros organismos, não existindo evidências de suas funções (DUNLAP et al., 2007). Recentemente o procedimento descrito por Ninomiya et al. (2004) foi utilizado por um consórcio de genômica funcional do fungo para a construção de um banco de linhagens mutantes contendo os 10.000 genes codificadores de proteínas individualmente nocauteados pela inserção de uma marca de seleção (COLOT et al., 2006). Linhagens contendo genes individualmente inativados estão disponíveis para a comunidade científica pelo *Fungal Genetics Stock Center* (FGSC, Kansas City, Missouri, USA), as quais constituem uma importante ferramenta para a realização de estudos relacionados à genômica funcional do fungo, ou seja, a caracterização de um gene específico.

2. Glicogênio. Estrutura e metabolismo

O microrganismos possuem a capacidade de utilizar uma grande variedade de nutrientes com o propósito de adaptar-se continuamente às mudanças das condições ambientais. Muitos deles, incluindo leveduras e bactérias, acumulam glicogênio em períodos de abundância de nutrientes e o degrada quando as células são submetidas a certos tipos de estresse, como por exemplo, escassez de nutrientes no ambiente (HARRIS, 1997). O glicogênio é uma molécula ramificada constituída por resíduos de glicose unidos covalentemente por ligações glicosídicas α -1,4 e ligações α -1,6, esta última responsável pela ramificação da molécula. Devido a sua estrutura, este carboidrato de reserva é capaz de armazenar uma grande quantidade de resíduos de glicose no interior da célula sem elevar a osmolaridade celular, além de aumentar a quantidade de extremidades não redutoras da molécula, permitindo uma rápida mobilização de seus resíduos de glicose pela ação de enzimas envolvidas em sua degradação (MELÉNDEZ et al., 1997).

De uma forma geral, os processos de síntese e degradação do glicogênio em organismos eucariotos evolutivamente distantes são muito bem conservados, envolvendo as mesmas enzimas. A síntese do glicogênio se inicia pela formação do nucleotídeo doador de glicose UDP-glicose (UDPG) a partir de glicose-1-fosfato e do nucleotídeo UTP, uma reação catalisada pela enzima UDP-glicose pirofosforilase. A seguir, a síntese continua pela ação sequencial das enzimas glicogenina, glicogênio sintase e enzima ramificadora. A enzima glicogenina foi primeiramente caracterizada na levedura *S. cerevisiae* e sua ação catalítica consiste na formação de um oligossacarídeo iniciador constituído por 8 a 12 resíduos de

glicose pela adição de resíduos glicosil provenientes da UDPG a três resíduos de tirosina da própria enzima em uma reação de auto-glicosilação (MU et al., 1996; ROACH & SKURAT, 1997). É importante ressaltar que mesmo tendo a sua estrutura resolvida (GIBBONS et al., 2002), pouco se sabe sobre como os processos de auto-glicosilação e alongamento do oligossacarídeo iniciador do glicogênio são catalisados pela enzima glicogenina. O fungo *N. crassa* possui uma isoforma de glicogenina (GNN), a qual é constituída por 687 resíduos de aminoácidos (de PAULA et al., 2005a). Esta proteína apresenta em sua região N-terminal seu domínio responsável por sua autoglicosilação, o qual é constituído por dois sítios de glicosilação (Tyr 196 e Tyr 198) (de PAULA et al., 2005b). Em sua região C-terminal, esta proteína apresenta o domínio responsável por sua interação com a enzima glicogênio sintase (de PAULA et al., 2005b). Atualmente pouco se sabe sobre os mecanismos regulatórios da expressão do gene codificador da enzima glicogenina. Entretanto, no fungo *N. crassa*, sua expressão é possivelmente controlada pela disponibilidade de nutrientes e estresse (de PAULA et al. 2005a).

Após a etapa de iniciação da síntese, a enzima glicogênio sintase é a responsável pela polimerização do oligossacarídeo iniciador ligado à molécula de glicogenina. Para isso, resíduos glicosil provenientes do UDPG são adicionados por meio de ligações glicosídicas α -1,4 à extremidade não redutora da cadeia polissacarídica em crescimento. Recentemente, a estrutura tridimensional da enzima glicogênio sintase eucariótica foi resolvida, revelando que esta proteína atua na forma de um tetrâmero (BASKARAN et al., 2011). Finalmente, a enzima glicosil-(4–6)-transferase, enzima ramificadora, catalisa a transferência de uma cadeia com aproximadamente 6 ou 7 resíduos de glicose a partir de uma extremidade para um resíduo de glicose na mesma cadeia ou em outra distinta formando as ligações ramificadas α -1,6 da molécula de glicogênio. No processo de degradação do glicogênio estão envolvidas as enzimas glicogênio fosforilase e enzima desramificadora. A enzima glicogênio fosforilase é responsável pela clivagem de resíduos de glicose unidos por meio de ligações α -1,4, nas extremidades não redutoras do glicogênio, liberando glicose-1-fosfato que é convertido em glicose-6-fosfato pela ação da enzima fosfoglicomutase. É importante ressaltar que a clivagem das ligações C1-O fornecem energia suficiente para produzir glicose-1-fosfato sem que seja necessário o uso de ATP. Isto faz com que o armazenamento do glicogênio seja mais eficiente em termos energéticos. A enzima desramificadora é responsável pelas funções de transferase e α -(1,6)-glicosidase na degradação do glicogênio. Inicialmente esta enzima é responsável pela transferência de 4 a 5 resíduos de uma ligação α -1,6 para uma extremidade não redutora de outra ramificação, assumindo uma função de transferase. Além disso, esta enzima hidroliza o resíduo de glicose no ponto de ramificação α -1,6, assumindo uma função de glicosidase. É importante ressaltar que esta enzima é energeticamente menos eficiente em relação à enzima glicogênio

fosforilase pelo fato de requerer o consumo de ATP para a liberação de resíduos livres de glicose. Um modelo esquemático do metabolismo do glicogênio, bem como as enzimas envolvidas tanto no processo de síntese quanto na degradação do metabolismo do glicogênio pode ser visto na figura 2.

3. Controle do metabolismo do glicogênio

A regulação do metabolismo de glicogênio tem sido estudada há décadas com foco nas atividades das enzimas responsáveis pelo controle da síntese e degradação do glicogênio, ou seja, pelas enzimas glicogênio sintase e glicogênio fosforilase, respectivamente. Sabe-se que os processos de síntese e degradação do glicogênio ocorrem de forma simultânea, pois o conteúdo do glicogênio celular está sujeito a alterações em resposta a diversas situações ambientais que os organismos estão sujeitos. Tanto o processo de síntese quanto o de degradação do glicogênio apresentam um complexo mecanismo regulatório, envolvendo controle por fosforilação reversível e alosterismo das enzimas regulatórias, os quais funcionam de forma antagônica nas enzimas responsáveis por tais processos (TÉLLEZ-INON et al., 1969, FLETTERICK & MADSEN 1980).

A regulação da enzima glicogênio sintase representa o principal ponto de controle do metabolismo do carboidrato. Esta enzima é regulada tanto por modulação alostérica quanto por fosforilação reversível. Existe uma forma ativa desfosforilada e um número variável de formas pouco ativas fosforiladas, as quais diferem em suas dependências pela glicose-6-fosfato, seu efetor alostérico (ROSELL-PEREZ et al., 1962; TÉLLEZ-IÑÓN et al., 1969). De um modo geral, as enzimas glicogênio sintases são conservadas nas regiões centrais e apresentam as divergências sequenciais em suas regiões N- e C-terminais, justamente as regiões onde estão contidos os sítios de fosforilação regulatórios (ROACH, 1990; BAI et al., 1990; HARDY & ROACH, 1993). Uma vez fosforilada, esta enzima fica em sua forma inativa, a qual pode ser restaurada por alosterismo efetuado pela molécula glicose-6-fosfato. Em *N. crassa*, a via de sinalização do PKA exerce um importante mecanismo de controle sobre a atividade desta enzima. Linhagens mutantes do fungo apresentando deficiência nesta via de sinalização ou atividade PKA exacerbada apresentaram alterações no acúmulo de glicogênio, bem como na expressão do gene codificador da enzima glicogênio sintase (*gsn*). Estudos *in vitro* mostraram que a via de sinalização do PKA também influencia a fosforilação de GSN (FREITAS et al., 2010).

A enzima glicogênio fosforilase, diferentemente da enzima glicogênio sintase, se encontra em sua forma ativa após a fosforilação de seu único sítio de fosforilação (Ser 14), localizado na região N-terminal, pela enzima fosforilase quinase. É importante ressaltar que a enzima fosforilase quinase é ativada pela via de sinalização do PKA, fazendo com que a

responsivos a várias condições estressantes (MARTINEZ-PASTOR et al., 1996). A maioria dos genes codificadores de enzimas envolvidas no metabolismo de glicogênio em leveduras contém entre uma e várias cópias de STRE em seus promotores (VARELA et al., 1995).

Em *N. crassa*, estudos mostraram a ocorrência de acúmulo de glicogênio durante a fase exponencial de crescimento vegetativo, o qual foi acompanhado por um aumento tanto da atividade glicogênio sintase quanto nos níveis do transcrito *gsn* desta enzima e que o glicogênio é degradado quando a cultura atinge a fase estacionária de crescimento (NOVENTA-JORDÃO et al., 1996; de PAULA et al., 2002). Sob condição de choque térmico, ao contrário do que ocorre em *S. cerevisiae*, o nível do transcrito *gsn* sofre uma acentuada redução e retorna ao normal após a reincubação da cultura em sua temperatura fisiológica de crescimento (de PAULA et al., 2002). Estes resultados mostram que a regulação do metabolismo de glicogênio em *N. crassa* envolve mecanismos diferentes dos descritos na levedura *S. cerevisiae* e se aproxima melhor dos mecanismos de regulação de células de mamíferos, ou seja, sob condições estressantes ocorre degradação de glicogênio.

Com o objetivo de estudar a regulação da expressão do gene *gsn* estudos de caracterização molecular da região promotora do gene *gsn* de *N. crassa* foram realizados em nosso laboratório (FREITAS & BERTOLINI, 2004). Uma análise detalhada da região 5'-flanqueadora do gene permitiu a identificação de vários prováveis elementos transcripcionais, dentre eles grande quantidade de HSEs e dois elementos STREs, um deles *upstream* a um TATA Box canônico (STRE1) e o outro (STRE2), localizado dentro de um *intron* presente na região 5'-UTR do gene *gsn*. Também foram identificados elementos STREs na região 3' do gene. Os HSEs (*Heat-Shock Elements*) são elementos de transcrição encontrados em múltiplas cópias, nos promotores de genes responsivos ao choque térmico (BIENZ & PELHAM, 1987) e conferem uma forte indução frente ao estresse térmico, podendo agir tanto como elementos proximais ao TATA Box quanto como *enhancers* (PELHAM & BIENZ, 1982; BIENZ & PELHAM, 1986). Através de ensaios de retardamento em gel (*Electrophoretic Mobility Shift Assay*, EMSA) com duas regiões do promotor *gsn*, uma contendo elementos HSE e outra contendo a região correspondente ao STRE1, bandas de DNA de mobilidade reduzida foram observadas com extratos nucleares de células do fungo submetidas ao choque térmico. Estes resultados mostraram que proteínas ativadas nesta situação ambiental foram capazes de ligar aos fragmentos de DNA do promotor de *gsn* e demonstraram fortemente o envolvimento das regiões estudadas na regulação da expressão do gene durante a condição de estresse mencionada. Além disso, experimentos foram realizados com o objetivo de isolar proteínas (fatores de transcrição ou não) que ligam à região promotora do gene *gsn* e, desta maneira, participam na regulação da expressão do gene. Esses estudos combinaram experimentos de ensaios de retardamento em gel, acoplados à análise proteômica e identificação das proteínas por espectrometria de massas.

Os experimentos levaram à identificação de algumas proteínas do fungo, as quais podem estar envolvidas na regulação do gene na condição de choque térmico (FREITAS et al., 2008).

Dando continuidade aos estudos de regulação, uma coleção de linhagens nocauteadas em genes que codificam fatores de transcrição foi adquirida pelo nosso laboratório e sua utilização tornou-se um material interessante para o início dos estudos acerca dos mecanismos moleculares envolvidos na regulação do metabolismo do glicogênio em *N. crassa*.

4. Estrutura da cromatina e fatores de transcrição

Nas células eucarióticas, o genoma é constituído pela molécula de DNA associada a uma grande quantidade de proteínas que interagem de forma dinâmica com a cadeia nucleotídica, sendo responsável pela sua compartimentalização no núcleo bem como pelo controle da expressão gênica. Essas proteínas pertencem a duas classes distintas: histonas e proteínas cromossômicas não histonas. Histonas são pequenas proteínas básicas que interagem com o DNA sob a forma de octâmeros, os quais são constituídos por dois pares das histonas H2A, H2B, H3 e H4, formando um *core* globular que interage em 14 pontos com a molécula de DNA (LUGER et al., 1997). As histonas que compõem o nucleossomo são compostas de um domínio globular e flexível relativamente não estruturado na região N-terminal, denominadas de caudas, que sobressaem da superfície da nucleossomo. Estas regiões são altamente conservadas a partir de levedura para os seres humanos. Um alto grau de conservação destas proteínas em diferentes organismos corrobora a visão geral de que estas proteínas, até mesmo os domínios da cauda não estruturados, são suscetíveis de servir a funções críticas para a manutenção destas estruturas em diferentes organismos (ALLIS et al., 2007). Essa estrutura é extremamente conservada nos organismos eucarióticos e constitui a unidade básica da cromatina (McGHEE & FELSENFELD, 1980).

As histonas estão sujeitas a um grande número de modificações pós-traducionais tanto nas caudas N-terminais quanto no *core* globular formado por suas interações junto à dupla fita de DNA. Estas modificações ocorrem pela ação de proteínas que catalisam a adição de grupos funcionais presentes na cauda flexível N-terminal, principalmente nas caudas das histonas H3 e H4, como por exemplo, acetilação e metilação dos resíduos de lisinas. A acetilação e metilação das histonas H3 e H4 foram as primeiras modificações covalentes descritas, entretanto, outros tipos de modificações covalentes foram identificados, como por exemplo a fosforilação, ubiquitinação, sumoilação, ADP-ribosilação, biotilação, isomerização de prolina, e outras modificações (VAQUERO et al., 2003). Entre as modificações das histonas, a metilação da lisina mostra a especificidade mais notável. Os

sítios de metilação mais estudados são lisina 4, 9, 27,36 e 79 da histona H3 e lisina 20 de histona H4, cada um podendo ser encontrado de um a três grupos metila, requerendo para isto enzimas específicas e/ou complexos enzimáticos (MARTIN & ZHANG, 2005).

Alterações nas estruturas das histonas provocam alterações na estrutura da cromatina e, conseqüentemente, alteram o padrão de regulação gênica. Esse mecanismo recebe o nome de remodelagem da cromatina e é amplamente usado para descrever alterações na estrutura durante os processos regulatórios. Genes, cujas regiões promotoras estão presentes em nucleossomos podem ter suas expressões inibidas caso os nucleossomos não estejam relaxados (ALMER et al., 1986; WASYLYK & CHAMBON, 1979). No entanto, estudos envolvendo o mapeamento de genomas revelaram que a densidade de nucleossomos nas regiões promotoras é menor se comparado com as regiões codificadoras (BERNSTEIN et al., 2004; LEE et al., 2004; SEKINGER et al., 2005). Estes dados sugerem que o posicionamento dos nucleossomos no genoma está intimamente relacionado com a disposição das regiões codificadoras do genoma.

Outro mecanismo responsável pela regulação gênica nos eucariotos é a presença de determinadas sequências nucleotídicas que atuam como sítios alvos para fatores de transcrição. Estas sequências recebem o nome de elementos *cis*, os quais podem estar localizados tanto *upstream* quanto *downstream* da unidade transcricional e são responsáveis tanto pela ativação quanto pela inibição da expressão gênica. Alguns elementos podem estar localizados a centenas ou milhares de bases do sítio de início da transcrição, como por exemplo, os elementos *enhancers*. Proporcionalmente ao aumento da complexidade dos elementos regulatórios *cis*, 0,5-10% das proteínas codificadas pelo genoma dos eucariotos são dedicados a proteínas que regulam a transcrição gênica (LEVINE & TJIAN, 2003). Essas proteínas estão subdivididas em duas maiores classes: as proteínas denominadas fatores de transcrição que ligam a sequências nucleotídicas específicas por meio de domínios de ligação ao DNA e medeiam a ativação ou repressão transcricional seletiva dos genes, e uma grande quantidade de proteínas que constituem a maquinaria multiproteica da RNA polimerase II, denominadas fatores gerais da transcrição, requeridas para o reconhecimento do promotor e catálise da síntese de RNA (LEVINE & TJIAN, 2003).

A transcrição dos genes estruturais das células eucarióticas é regulada por fatores de transcrição que interagem com os elementos *cis* presentes nas regiões promotoras por meio de um domínio de ligação específico ao DNA. Estudos envolvendo fatores de transcrição têm mostrado que essas proteínas regulatórias possuem diferentes domínios relacionados a importantes funções do processo transcricional (LATCHMAN, 2005). Além do domínio de ligação ao DNA, os fatores de transcrição precisam também apresentar um domínio de ativação, requerido para estimular a transcrição. É importante ressaltar que alguns fatores

de transcrição não possuem o domínio de ativação. Neste caso, o domínio de ligação e o domínio de ativação estão localizados em diferentes proteínas no complexo de ligação ao DNA, sendo necessária a interação de ambos para que possa ocorrer a transcrição (LATCHMAN, 2007).

Para que o controle do processo transcricional seja eficiente, é necessário que ocorra uma ação combinatória de vários ativadores ou inibidores que se ligam aos elementos *cis* presente no DNA. Como consequência, um grande número de complexos regulatórios pode ser arranjado a partir de uma grande quantidade de combinações de fatores. Esta ideia ganhou força com o descobrimento de complexos multiproteicos que pode alterar a estrutura da cromatina, como por exemplo, os reguladores transcricionais globais em *S. cerevisiae*, tais como o complexo SAGA, com função da enzima histona acetiltransferase (HAT) que transfere o grupo acetil da acetil coenzima A (acetil-CoA) para os resíduos de lisina presente nas caudas das histonas H3 e/ou H4, alterando a carga deste resíduo, fazendo com que diminua sua interação ao DNA, e o complexo Swi/Snf, com capacidade de desestabilizar tanto os dímeros H2A/H2B quanto os tetrâmeros H3/H4 de histonas liberando os elementos *cis* bloqueados pelos nucleossomos (WOLFFE & HAYES, 1999). É importante ressaltar que todos os reguladores de cromatina precisam se ligar a regiões específicas da cromatina para poderem efetuar suas respectivas funções. Entretanto, esses reguladores geralmente não possuem sítios de ligação no DNA e interagem com determinadas proteínas já ligadas ao DNA para poderem desempenhar suas funções.

Em suma, para que ocorra o controle da expressão gênica, é necessário que os mecanismos regulatórios estimulem ou inibam a transcrição por meio do recrutamento da maquinaria proteica requerida pela RNA polimerase II e proteínas modificadoras de cromatina para desbloquear o promotor dos genes alvos. Uma consequência da remodelagem da cromatina é a exposição ou inibição dos elementos *cis* aos fatores de transcrição, os quais podem estar na superfície do nucleossomo ou no interior dos mesmos, fazendo com que genes específicos sejam ativados ou reprimidos no momento específico. Os fatores de transcrição interagem com os elementos *cis* presentes nos genes alvos por meio de um domínio de ligação específico ao DNA. A característica distintiva dos fatores de transcrição em relação as demais proteínas é que os fatores de transcrição possuem a capacidade de reconhecer sequências nucleotídicas específicas (LATCHMAN, 1997). Tal reconhecimento obedece a coordenadas temporais e espaciais extremamente rigorosas com o intuito de traduzir ao nível molecular respostas aos estímulos ambientais percebidos pelos organismos (LUSCOMBE et al., 2000).

O genoma dos organismos eucarióticos possui uma grande diversidade no repertório de fatores de transcrição, tanto em relação à presença, quanto em variedade de famílias destas proteínas regulatórias. Estima-se que os fatores de transcrição constituem de 0,5 a

10% do conteúdo do genoma eucariótico, variando de acordo com o nível de complexidade do organismo (LEVINE & TJIAN, 2003). Uma importante característica dos fatores de transcrição de eucariotos é a capacidade destas proteínas regulatórias de atuarem em cooperação com outros fatores de transcrição bem como com complexos proteicos para a regulação de diferentes genes (ARNOSTI & KULKARNI, 2005).

Em relação às suas características estruturais, os fatores de transcrição eucarióticos apresentam múltiplos domínios funcionais, não só para o reconhecimento e ligação à sequência alvo de DNA, mas também domínios para interações proteicas (LEVINE & TJIAN, 2003). Os fatores de transcrição são agrupados em famílias, definidas com base na sequência primária de seus domínios de ligação ao DNA. Estudos evolutivos têm mostrado que, embora o motivo de ligação a DNA seja altamente conservado entre diferentes eucariotos o restante da sequência é frequentemente muito diferente (LATCHMAN, 2007). O genoma do fungo *Neurospora crassa* possui 176 proteínas anotadas como fatores de transcrição, o que representa 1,7% dos genes codificadores de proteínas deste organismo (BORKOVICH et al., 2004). Desse total, 153 proteínas pertencem a 5 grandes famílias de fatores de transcrição (Zn_2Cys_6 , C2H2, bZIP, bHLH, GATA) e as restantes foram agrupadas em uma categoria denominada diversos pois fazem parte de famílias de fatores de transcrição menos expressivas.

Os fatores pertencentes à família *zinc finger* constituem uma das maiores famílias de fatores de transcrição em eucariotos, sendo responsáveis pela regulação de distintos processos metabólicos. A particularidade desta família é a presença de um ou mais átomos de zinco necessários para a formação da estrutura de seu sítio ativo denominado *zinc finger* (BRANDEN & TOOZE, 1999). As cadeias polipeptídicas do domínio *zinc finger* são geralmente pequenas, com um padrão regular de resíduos de cisteína e/ou histidina ao longo de sua cadeia. Esses resíduos são responsáveis pela interação da sequência polipeptídica com um ou mais átomos de zinco, proporcionando a formação de uma conformação independente do domínio proteico. Em eucariotos, as principais subfamílias dos fatores de transcrição *zinc fingers* possuem o domínio de ligação ao DNA C_2H_2 , GATA e Zn_2Cys_6 .

O clássico domínio *zinc finger*, também denominado de C_2H_2 , é constituído por uma unidade de conformação independente descrita como uma proteína globular, com um *core* hidrofóbico e uma cadeia lateral polar em sua superfície (BRANDEN & TOOZE, 1999). De uma forma geral, o domínio é caracterizado pela existência de duas folhas β anti-paralelas seguidas de um *hairpin*, com o átomo de zinco ligando a resíduos Cys das duas folhas β . Uma mesma proteína pode ter mais de um domínio *zinc finger*. Estas repetições assumem um arranjo modular que oferece um grande número de possibilidades de combinações para reconhecimento específico de DNA, além de permitir um grande aumento na especificidade

de reconhecimento dos sítios de ligação ao DNA. Essa família representa a segunda maior classe de fatores de transcrição existente no genoma de *N. crassa* (BORKOVICH et al., 2004).

A subfamília representada pelos fatores GATA, também denominados de C_4 , recebe esta denominação devido à sua capacidade de se ligar aos sítios (A/T)GATA(A/T) presentes nas regiões regulatórias do DNA. Análises por ressonância magnética nuclear revelaram que este domínio de ligação ao DNA é constituído por dois pares de fitas β seguidas por um *loop*, onde o átomo de zinco é coordenado por 4 cisteínas localizadas na fita β_1 , β_2 , e as duas restantes na α hélice (OMICINSKI et al., 1993; STARICH et al., 1998). O fungo *N. crassa* possui 6 fatores de transcrição pertencentes a esta família (BORKOVICH et al., 2004). Podemos citar como exemplo o fator de transcrição NIT-2, que regula a expressão do gene codificador da enzima nitrato redutase (*nit-3*), envolvida no metabolismo de nitrogênio no fungo ou os fatores de transcrição WC1 e WC2 envolvidos na maioria dos processos mediados por luz, bem como no controle do ritmo circadiano do fungo *N. crassa* (HARDING & TURNER, 1981; LINDEN & MACINO, 1997).

A subfamília Zn_2Cys_6 , também conhecida como *binuclear cluster* é restrita aos fungos, sendo predominante nos organismos pertencentes à família dos ascomicetos (TODD & ADRIANOPOULOS, 1997). O domínio *zinc finger* Zn_2Cys_6 consiste de um par de α -hélices que coordena dois átomos de zinco através de 6 resíduos de cisteína, uma α -hélice se liga ao DNA pelo sulco maior e a segunda interage com o esqueleto de fosfato (LUSCOMBE et al., 2000). Além disso, outra característica dos fatores de transcrição pertencentes a essa família é a presença de dois íons Zn^{2+} sendo coordenados por seis cisteínas, permitindo que o domínio de ligação ao DNA assumira sua conformação binuclear (SCHJERLING & HOLMBERG, 1996). Seus sítios de ligação ao DNA consistem em uma trinca de nucleotídeos conservada e intercalada por uma sequência variável ($CGGN_xCCG$). A diferença entre as proteínas desta família é o espaçamento entre os dois domínios (C_2H_2 e Zn_2Cys_6) que se ligam ao DNA, ou seja, o que determina a especificidade de ligação de uma proteína da família Zn_2Cys_6 com seu sítio alvo no DNA é o espaçamento entre os nucleotídeos (VASHEE et al., 1993). Os fatores de transcrição Zn_2Cys_6 representam a maior família de fatores de transcrição em *N. crassa* onde foram identificadas 77 ORFs que codificam proteínas contendo esse domínio.

Os fatores de transcrição pertencentes à família bHLH são exclusivos de eucariotos, são altamente conservados em diversos organismos, desde leveduras até mamíferos, estão envolvidos no controle de diversas vias metabólicas e estão divididas em diferentes classes, de acordo com suas funções (MURRE et al., 1994). O domínio de ligação ao DNA destas proteínas é constituído de duas α -hélices anfipáticas, separadas por uma região espaçadora em forma de alça, que pode variar entre 5 a 23 resíduos (BRANDEN & TOOZE, 1999). Além

disso, os fatores de transcrição pertencentes a esta família possuem duas regiões altamente conservadas: uma que consiste em uma região com resíduos predominantemente básicos, responsáveis pela interação dessas proteínas ao DNA, e outra formada por resíduos hidrofóbicos, responsáveis pela dimerização dos monômeros. Os dímeros são estabilizados por uma série de interações hidrofóbicas e eletrostáticas entre as α -hélices, devido à formação de um feixe de 4 hélices que mantém os monômeros unidos, constituindo uma interface de ligação ao DNA. Essa estrutura possui uma geometria precisa, que interage via o sulco maior do DNA em uma sequência consenso (CANNTG) denominada de *E BOX* (BRANDEN & TOOZE, 1999). É importante ressaltar que as proteínas que apresentam domínios de ligação bHLH não possuem a capacidade de diferenciar os sítios de ligação ao DNA sob a forma de monômeros. Assim, o mecanismo de dimerização tem a função de converter moléculas monoméricas inativas em complexos diméricos com atividade transcricional em períodos determinados durante o ciclo celular. Em *N. crassa*, 10 proteínas foram identificadas nesta família, das quais somente a codificada pelo gene *nuc-1* foi caracterizada, estando envolvida com a regulação da utilização de fosfato (KANG & METZENBERG, 1990).

Os fatores de transcrição pertencentes à família bZIP são constituídos por duas α -hélices com 60 a 80 resíduos de aminoácidos com uma região N-terminal básica responsável pela sua interação ao DNA e uma região responsável pela dimerização deste domínio localizada normalmente em sua região C-terminal (VINSON et al., 1993). A principal característica dos fatores de transcrição pertencentes a esta família é a presença e uma região responsável pela dimerização entre duas hélices. Esta região é constituída por sucessivas repetições do aminoácido leucina a cada 7 resíduos de aminoácido, permitindo a formação de uma estrutura supersecundária responsável pela interação entre as duas α -hélices distintas formando homo ou heterodímeros (LANDSCHULZ, 1988). No fungo *N. crassa* foram encontrados 17 fatores de transcrição pertencentes a esta família (BORKOVICH et al., 2004). Até o momento, pouco se sabe sobre os fatores de transcrição pertencentes a esta família no fungo *N. crassa*, pois mais da metade deles não apresenta ortólogos funcionais depositados em banco de dados (BORKOVICH et al., 2004; TIAN et al., 2011). A figura 3 mostra as estruturas tridimensionais dos domínios presentes nos fatores de transcrição descritos acima.

Como descrito anteriormente, uma coleção de linhagens do fungo nocauteadas em genes que codificam fatores de transcrição foi utilizada visando identificar fatores de transcrição provavelmente atuando como reguladores do metabolismo do glicogênio em *N. crassa*. Os estudos levaram à identificação de diversas proteínas, pertencentes a diferentes famílias (GONÇALVES et al., 2011), e os seus papéis no metabolismo de glicogênio estão sendo detalhadamente investigados. A seguir, algumas considerações serão realizadas em

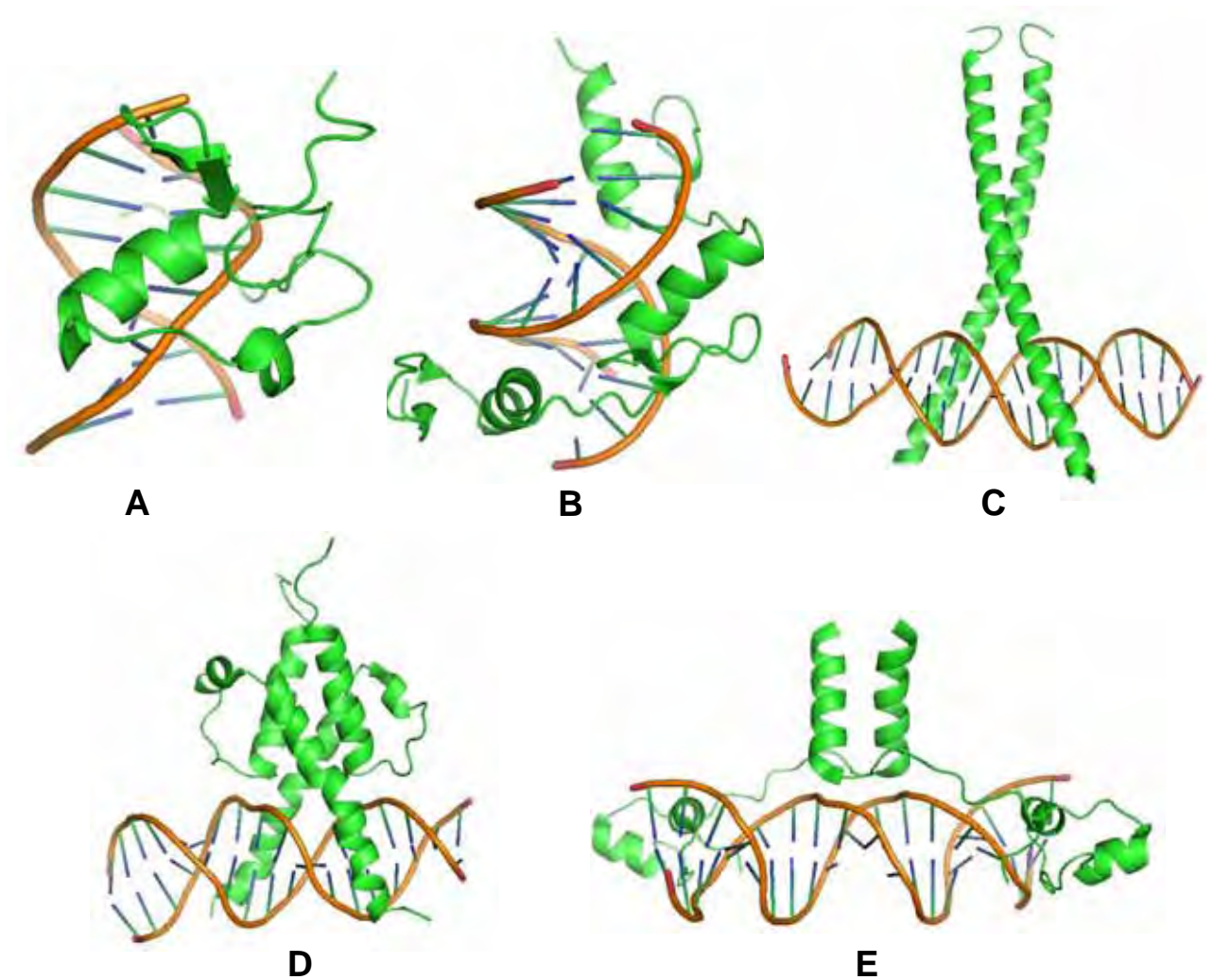


Figura 3 - Representação tridimensional dos domínios de ligação presentes nos fatores de transcrição selecionados. (a) fator de transcrição GATA-1 interagindo com o DNA (PDB 1GAT); (b) Proteína ZIF 268 formando um trímero que se liga ao DNA por meio de seus domínios C_2H_2 (PDB 1A1L); (c) Proteína CREB bZIP interagindo com o DNA (PDB 1DH3); (d) Proteína PHO4 ligada ao DNA por seu domínio bHLH (PDB 1A0A); (e) Proteína GAL4 ligada ao DNA por meio de seus domínios Zn_2Cys_6 (PDB 1D66) (<http://www.rcsb.org/pdb/home/home.do>).

relação a dois fatores de transcrição objeto de estudo, neste trabalho: o fator de transcrição XLR-1 e um fator de transcrição anotado como proteína hipotética.

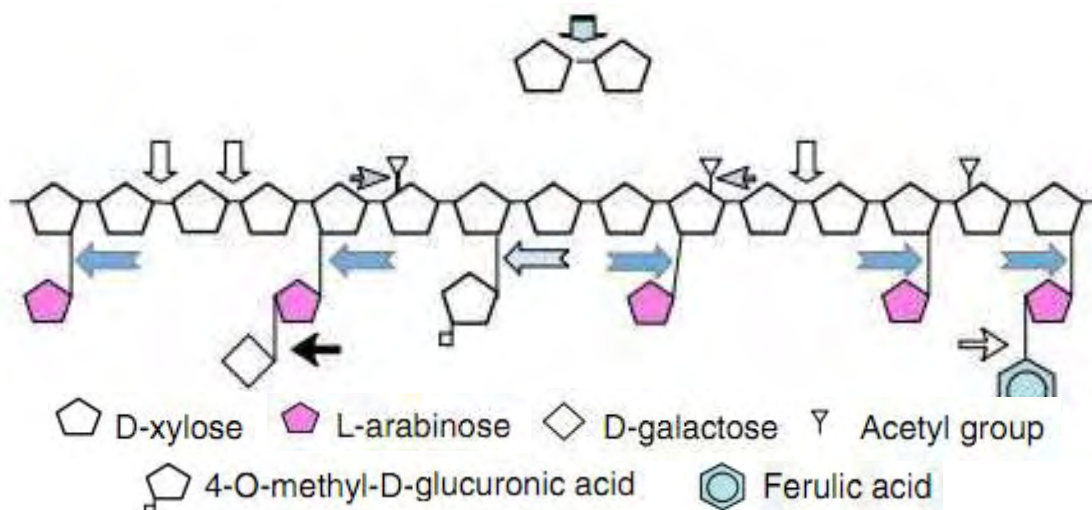
5. O fator de transcrição XlnR em fungos filamentosos

O ativador transcricional XlnR é um fator de transcrição envolvido na regulação de enzimas envolvidas na degradação da celulose e hemicelulose em fungos filamentosos. A hemicelulose é um heteropolissacarídeo constituído por pentoses, hexoses e ácidos urônicos, e representa o segundo polisacarídeo mais abundante na natureza (COLLINS et al., 2005). Atualmente este polisacarídeo vem sendo reconhecido como um potencial substrato para a produção de combustíveis líquidos, bem como outros produtos de interesse biotecnológico (POLIZELI et al., 2005). Este polissacarídeo é classificado de acordo com os principais resíduos de açúcar que constituem sua cadeia principal. O mais comum é a xilana, formada principalmente por resíduos de D-xilose unidos por ligações β -1,4, sendo encontrada em cereais e madeira. É importante ressaltar que os açúcares constituintes da cadeia principal das hemiceluloses frequentemente apresentam modificações em seus grupos laterais. Podemos citar como exemplo os resíduos de xilose, os quais podem apresentar várias modificações, como os ácidos α -D-glucurônico e o α -4-O-metil-D-glucurônico, unidos via ligações glicosídicas α -1,2, ou α -L-arabinofuranose, unida por ligações glicosídicas α -1,2 ou α -1,3 (TIMELL, 1967; WILKIE, 1979). Além disso, resíduos aromáticos, como o ácido ferrúlico, também pode estar ligado às frações de hemicelulose e pectina (BELIËN et al., 2006). Estes polissacarídeos são capazes de fazer ligações cruzadas, assim como, com compostos aromáticos poliméricos da lignina, resultando em um aumento na rigidez da parede celular (de VRIES & VISSER, 2001). A figura 4 mostra um esquema da estrutura da xilana.

Para a degradação da hemicelulose, é necessário inicialmente, a ação das endo-1,4-endoxilanasas, as quais clivam as ligações glicosídicas internas ao longo do heteropolímero, diminuindo assim o grau de polimerização do substrato. Além das endoxilanasas, é necessário a ação de enzimas β -D-xilosidases, as quais são responsáveis pela clivagem de xilobiose. Estas enzimas atacam as extremidades não redutoras de xilooligosacarídeos resultantes da atividade das endo-1,4-endoxilanasas (SORENSEN et al., 2003). Além disso, para a clivagem total da cadeia lateral, ou seja, conversão do polímero em monossacarídeos é necessário à ação conjunta de várias enzimas acessórias devido à sua complexidade estrutural (POLIZELI et al., 2005).

A hidrólise da hemicelulose por fungos filamentosos requer a produção várias enzimas xilanólíticas, as quais são induzidas pela presença de seus substratos e regulada principalmente ao nível transcricional por meio de fatores de transcrição. Em fungos

filamentosos, sabe-se que a ativação dos genes codificadores de enzimas xilanolíticas requer a ação do ativador transcricional XlnR (van PEIJ et al., 1998a). O mecanismo de ação desta proteína foi bem caracterizado em organismos do gênero *Aspergillus* e *Trichoderma* (STRICKER et al., 2008). É uma proteína reguladora que se liga ao elemento *cis* 5'-GGCTRA-3' presente na região promotora de genes codificadoras de enzimas envolvidas principalmente na degradação de hemicelulose, tais como genes que codificam endoxilanases e β -xylosidases (TAMAYO et al., 2008; de VRIES et al., 1998; van PEIJ et al., 1998a, 1998b; HASPER et al., 2002; MARUI et al., 2002, RAUSCHER et al., 2006). É importante ressaltar que o papel de XlnR como um ativador transcricional não é restrito apenas a genes que codificam endoxilanases, estendendo-se também à regulação das enzimas acessórias envolvidas na degradação de hemicelulose, como por exemplo, α -glucuronidase e celobiohidrolases (GILKENS et al., 1999; de VRIES & VISSER, 1998).



SYMBOL	ENZYME	EC-NUMBER	LINKAGE HYDROLYSED
↓	Endo-1,4- β -xylanase	3.2.1.8	β -1,4
⌞	β -D-xylosidase	3.2.1.37	β -1,4
→	α -L-arabinofuranosidase Arabinoxylan arabinofuranohydrolases	3.2.1.55	α -L-1,2; α -L-1,3 and α -L-1,5
⇌	α -D-glucuronidase	3.2.1	α -1,2
⇒	feruloyl esterase	3.2.1	ester bond
→	α -D-galactosidase	3.2.1.22	α -1,6
⇌	asetilyxylan esterase	3.1.1.6	ester bond

Figura 4 - Representação esquemática da xilana junto ao repertório enzimático envolvido em sua degradação (ARO et al., 2005).

A presença de xilose ou xilana induz a transcrição de *xlnR/xyr1*, cujo produto leva à ativação de genes que codificam tais hidrolases. Mutações deletérias em *xlnR* promove uma redução total ou parcial na transcrição de alguns genes ativados por xilose e/ou xilana, da mesma forma que a expressão de alguns genes regulados pela proteína XlnR é aumentada quando o gene foi superexpresso (GIELKENS et al., 1999; van PEIJ et al., 1998b). Por outro lado, a ação do fator de transcrição XlnR é reprimida pelo sistema de repressão catabólica mediado pela proteína CreA. Este fator de transcrição pertence à família C₂H₂ e liga-se ao elemento *cis* 5'-SYGGRG-3' presente na região promotora de seus genes alvo (KULMBURG et al., 1993). Na presença de glicose (açúcar preferencial), ocorre a transcrição de *creA*, levando à inibição ou redução da transcrição de genes ativados por XlnR. Sabe-se que a proteína CreA em *Aspergillus* controla a transcrição de genes que codificam arabinases e algumas enzimas xilanolíticas, como endoxilanases e β-xilosidases (PEREZ-GONZALEZ et al., 1998, RUIJTER et al., 1997; KUMAR & RAMÓN, 1995).

Em *N. crassa*, o fator de transcrição XLR-1 (*xy*/*an* degradation *r*egulator-1) essencial para a degradação de hemicelulose é um ortólogo de XlnR/Xyr1 de espécies de *Aspergillus* e *Trichoderma*, respectivamente. Deleção de *xlr-1* aboliu o crescimento em xilana e xilose, mas o crescimento em celulose e atividade celulolítica foram levemente afetados (SUN et al., 2012).

Objetivos

Este trabalho teve como objetivo principal caracterizar funcionalmente dois fatores de transcrição identificados previamente como reguladores do metabolismo de glicogênio no fungo *N. crassa*. A indentificação destas proteínas foi uma consequência de uma análise sistemática utilizando uma coleção de linhagens do fungo individualmente mutadas em ORFs codificadoras de fatores de transcrição. Um fator de transcrição corresponde à proteína ortóloga de XlnR, descrita como reguladora da expressão de genes codificadores de enzimas xilanolíticas e celulolíticas e o outro fator de transcrição é o produto da ORF NCU04390, uma proteína sem função conhecida até o momento. De forma mais específica, os objetivos do trabalho foram:

- Tentar estabelecer uma correlação entre metabolismo de glicogênio e controle metabólico frente a fontes de carbono alternativas. Realizar uma caracterização funcional do fator de transcrição XLR-1 no fungo *N. crassa*, a fim verificar seu envolvimento com o controle do metabolismo de glicogênio no fungo. Além disso, as análises funcionais foram realizadas com o objetivo de verificar se o mecanismo de ação desta proteína regulatória em *N. crassa* segue os mesmos padrões observados de seus ortólogos em outros fungos filamentosos.
- Tentar atribuir uma função regulatória do fator de transcrição hipotético produto da ORF NCU04390 na regulação do metabolismo de glicogênio. Determinar em escala genômica genes possivelmente regulados pela proteína hipotética por análises de CHIP-Seq uma vez que não se conhece o papel funcional desta proteína. Realizar análises de interação DNA-proteína *in vitro* utilizando a técnica de EMSA para validar os resultados obtidos no sequenciamento.

Materiais e Métodos

1. Linhagens

1.1. *Neurospora crassa*

Linhagens selvagem

FGSC#9718 (a, $\Delta mus51:Hyg^r$), **FGSC#9718** (A, $\Delta mus51:Hyg^r$)

Linhagens mutantes

FGSC#11004 (a, $\Delta NCU08000$), FGSC#11039 (a, $\Delta NCU01097$), FGSC#11062 (A, $\Delta NCU09739$), FGSC#11067 (a, $\Delta NCU06971$), FGSC#11089 (A, $\Delta NCU04851$), FGSC#11102 (A, $\Delta NCU01629$), FGSC#11123 (A, $\Delta NCU00808$), FGSC#11126 (A, $\Delta NCU01154$), FGSC#11134 (A, $\Delta NCU04390$), FGSC#11139 (a, $\Delta NCU04731$).

As linhagens foram obtidas junto ao *Fungal Genetics Stock Center* (FGSC), Kansas City, MO, USA.

1.2 *Escherichia coli*

DH10 β (Invitrogen)

Rosetta (DE3) pLysS (Novagen)

ArcticExpress (DE3) (Agilent)

2. Meios de cultivo

2.1 *Neurospora crassa*

2.1.1. Meio mínimo de Vogel (VM)

O meio de cultivo utilizado para o crescimento e manutenção das linhagens do fungo *N. crassa* foi o meio mínimo de Vogel (VM) (Vogel, 1956). Esse meio foi preparado contendo sacarose, sais de Vogel e biotina. A solução de sais de Vogel foi preparada 50X concentrada, a partir da dissolução dos seguintes sais, na ordem indicada, em 700 ml de água destilada:

510 g de citrato de sódio.H₂O

250 g de KH₂PO₄

100 g de NH₄NO₃

10 g de MgSO₄.7H₂O

5 g de CaCl₂.2H₂O (previamente dissolvido em água destilada)

2,5 mL de solução de biotina 10%

5 mL de solução de elementos traços

O volume foi completado para 1 L com água destilada e 300 µl de clorofórmio foi adicionado como agente conservante. A solução estoque de sais de Vogel foi armazenada a 4°C. A solução de elementos traços foi feita pela dissolução dos seguintes compostos em 90 mL de água:

5 g de ácido cítrico.H₂O

5 g de ZnSO₄.7H₂O

1 g de F(NH₄)₂.(SO₄)₂.6H₂O

0,25 g de CuS₄.5H₂O

0,05 g de MnSO₄.H₂O

0,05 g de H₃BO₃

0,05 g de NaMoO₄.2H₂O

O meio mínimo de Vogel foi preparado contendo 1X de sais de Vogel acrescido de 2% de sacarose ou 2% de frutose ou 2% de xilose ou 2% de xilana como fontes de carbono. Para o crescimento em meio sólido o meio foi acrescido de 2% de ágar.

2.1.2. Meio utilizado para cruzamento (SC)

Para o cruzamento das linhagens do fungo *N. crassa* foi utilizado o meio SC (WESTERGAARD & MITCHELL 1947). A solução de sais SC foi preparada 50X concentrada, a partir da dissolução dos sais listados abaixo em 500 mL de água destilada. O volume foi completado para 600 mL com água destilada e 200 µL de clorofórmio foram adicionados como agente conservante.

60 g de KNO₃

42g de K₂HPO₄

30 g de KH₂PO₄

30 g de MgSO₄.7H₂O

6 g de NaCl

6 g de CaCl₂.2 H₂O (previamente dissolvido em água destilada)

6 mL de solução de elementos traços

3 mL da solução estoque de biotina

O meio SC foi preparado contendo 1X de sais SC acrescido de sacarose 0,3% e ágar 2%.

2.1.3. Top ágar

Ágar 1,5%

Sorbitol 2M

FGS 2X (solução estoque 10X: glicose 0,5%, frutose 0,5%, sorbose 20%)

2.1.4. Base ágar

VM 1X

Ágar 1,5%

FGS 1X

4 µg/mL de higromicina (Invitrogen)

2.2. *Escherichia coli*

2.2.1. 2YT + 0,2% de glicose

Bacto-triptona 1,6%

Extrato de levedura 1,0%

NaCl 0,5%

Glicose 0,2%

2.2.2. Terrific Broth (TB)

Bacto-triptona 1,2%

Extrato de levedura 2,4%

Glicerol 0,4%

KH₂PO₄ 17 mM

K₂HPO₄ 72 mM

3. Propagação das linhagens de *N. crassa* e obtenção dos conídios

Para a obtenção dos micélios utilizados nos experimentos de choque térmico, o inóculo foi feito em garrafas de vidro contendo 50 ml de meio sólido e deixado a 30°C na estufa por 48 horas e, em seguida, expostos à luz até um total de 9 a 10 dias. Após esse período, os conídios foram coletados pela adição de aproximadamente 10 ml de água destilada estéril à garrafa seguida da filtração em gaze estéril para retenção das hifas e os conídios foram contados em câmara de Neubauer.

4. Choque térmico

Uma suspensão de conídios (densidade de 10^7 células/ml) foi inoculada em 60 ml de meio VM (Erlenmeyers de 250 ml), e colocada a 30°C sob agitação constante de 250 rpm por 24 horas em um agitador New Brunswick. Após esse período, o micélio correspondente a um dos frascos de cada linhagem foi coletado por filtração a vácuo em papel filtro, e metade do material coletado foi rapidamente congelado em nitrogênio líquido e armazenado a -80°C (amostra controle, não submetida a choque térmico). A outra metade foi transferida para um terceiro frasco contendo VM líquido, pré-aquecido a 45°C, e deixado nesta temperatura por 30 minutos. Em seguida o micélio foi coletado por filtração, rapidamente congelado em nitrogênio líquido e armazenado a -80°C (amostra submetida a choque térmico).

5. Crescimento do fungo utilizando pentoses como fontes de carbono

Para este ensaio foram utilizadas as linhagens FGSC#9718 e a FGSC#110067. Uma suspensão de conídios (densidade de 10^7 células/ml) foi inoculada em 250 ml de meio VM acrescido de frutose 2% (Erlenmeyer de 1 L) e colocada a 30°C sob agitação constante de 250 rpm por 24 horas em um agitador New Brunswick. Após esse período, o micélio foi coletado por filtração a vácuo em papel filtro e uma pequena amostra foi rapidamente congelada em nitrogênio líquido e armazenado a -80°C (amostra controle). O micélio coletado por filtração foi então dividido ao meio e uma parcela da amostra foi inoculada em 250 mL de meio VM acrescido de xilose 2% (Erlenmeyer de 1 L) e a outra parcela foi inoculada em 250 mL de meio VM acrescido de xilana 2% (Erlenmeyer de 1 L). As amostras foram incubadas a 30°C, sob agitação constante de 250 rpm. Foram realizadas coletas após 30 minutos, 1 hora, 2 horas, 6 horas e 12 horas de incubação. As amostras foram coletadas por filtração, rapidamente congeladas em nitrogênio líquido e armazenado a -80°C.

6. Preparo do extrato celular e determinação do conteúdo de glicogênio

Para o preparo do extrato celular, aproximadamente 200 mg do micélio congelado a -80°C foram macerados em nitrogênio líquido até sua total pulverização. Em seguida, transferiu-se esse material para um tubo eppendorf de 1,5 ml contendo 1 ml de solução tampão gelada (Tris-HCl 50 mM, pH 8,0, NaF 50 mM, EDTA 1 mM, PMSF 1 mM, benzamidina 1 mM, TLCK 2 mM, aprotinina 1 µg/mL e pepstatina A 1 µg/mL). A mistura foi fortemente agitada em *vortex*, e em seguida, foi centrifugada a 10.000 x *g* por 10 minutos em centrífuga Eppendorf de bancada, recolhendo o sobrenadante em um tubo eppendorf novo. Uma alíquota foi utilizada para a determinação das proteínas totais do extrato celular

pelo método de Hartree modificado utilizando soroalbumina bovina (BSA) como padrão (HARTREE, 1972).

O conteúdo de glicogênio foi determinado pela digestão do glicogênio com amiloglicosidase após precipitação com etanol absoluto como descrito por Hardy & Roach (1993), com algumas modificações. Primeiramente, foram coletados 100 µl do extrato celular bruto e adicionado 25 µl de TCA 50%. Em seguida, o material foi centrifugado a 5000 x g por 10 minutos a 4°C. O sobrenadante foi transferido para um tubo eppendorf limpo, sendo adicionado 500 µl de etanol absoluto e deixado 30 minutos a -80°C para a precipitação do glicogênio presente na fração solúvel. Finalmente, a solução foi centrifugada a 17000 x g por 10 minutos para a coleta do glicogênio, o qual foi lavado duas vezes com 500 µl de etanol 66%, levemente seco ao ar, e ressuspenso em 400 µl de solução acetato (acetato de sódio 50 mM, CaCl₂ 5 mM, pH 5,2). O glicogênio foi digerido com as enzimas α-amilase (1 µl de uma solução 10 mg/ml) e amiloglicosidase (1 µl de uma solução de 30 mg/ml) durante 12 a 15 horas. O conteúdo de glicose livre foi determinado pelo kit *Glicose PAP Liquiform* (Labtest) segundo as instruções do fabricante. O conteúdo de glicogênio presente em cada amostra foi calculado de acordo com uma curva padrão de glicogênio e expresso em relação à quantidade de proteína total de cada amostra.

A curva padrão de glicogênio foi preparada a partir de uma solução de glicogênio puro (Sigma) em solução acetato (acetato de sódio 50 mM, CaCl₂ 5 mM, pH 5,2) na concentração de 10 mg/ml.

7. Análise da expressão gênica

7.1. Extração de RNA total

O RNA total das linhagens selecionadas foi extraído pelo método de cloreto do lítio (Sokolovsky *et al.*, 1990). Para isto, uma pequena quantidade do micélio congelado foi macerada com nitrogênio líquido e o pó resultante foi transferido para tubos eppendorf contendo uma mistura de 750µl de tampão de lise (NaCl 0,6M, EDTA 10 mM, Tris-HCl 0,1M, pH 8,0 e SDS 4%) e 750µl de fenol saturado em Tris-HCl 0,1 M, pH 8,0. A suspensão foi agitada durante 15 minutos à temperatura ambiente e a fase aquosa foi coletada a 10.000 rpm por 10 minutos a 4°C em centrífuga refrigerada de bancada. O RNA contido na fase aquosa foi precipitado com 0,75 volumes de uma solução de LiCl 8 M e mantido a 4°C durante uma noite. Após esta etapa, o material foi centrifugado a 13.000 rpm por 10 minutos a 4°C, o precipitado contendo os ácidos nucleicos foi ressuspenso em 300 µl de água e novamente precipitado com acetato de sódio 3 M e etanol absoluto gelado a -80°C, durante 30 minutos. O RNA total foi coletado a 13.000 rpm, por 15 minutos a 4°C, lavado duas vezes

com uma solução de etanol 70% gelada a 13.000 rpm por 5 minutos a 4°C, seco e solubilizado em 100 µl de água previamente tratada com DEPC. A concentração do RNA foi determinada por absorvância a 260 nm em espectrofotômetro Beckman, modelo DU 640 e sua pureza analisada através da relação de absorvâncias 260/280 nm.

7.2. Análise da expressão gênica por *Northern blot*

A análise por *Northern blot* foi realizada segundo protocolo descrito por SAMBROOK & RUSSELL (2001). Para isto, aproximadamente 15 µg de cada amostra de RNA total foi previamente desnaturada em tampão de desnaturação (MOPS 1X, formamida deionizada 50%, formaldeído 20%, glicerol 10%, azul de bromofenol 0,025%, brometo de etídeo 0,05 µg/ml) a 65°C durante 10 minutos. As amostras desnaturadas foram submetidas à eletroforese em gel de ágarose 1,5% contendo formaldeído 0,6 M a 65V por aproximadamente 4 horas em tampão MOPS 1X (MOPS 20 mM, acetato de sódio 5 mM e EDTA 1 mM, pH 7,0). O RNA total, após eletroforese, foi transferido por capilaridade para membrana de nylon neutra (Hybond N, Amersham) utilizando solução tampão SSC 2X (NaCl 0,3 M e citrato de sódio 3 mM, pH 7,0) durante a noite. Após esta etapa, os RNAs foram fixados na membrana por *crosslinking* (Ultraviolet Crosslinker UVP mod. CL1000) durante 2 minutos e, em seguida, a membrana foi lavada com solução tampão SSPE 2X (NaCl 0,3 M, NaH₂PO₄ 2 mM e EDTA 2 mM, pH 7,0) para a total remoção do formaldeído. Após as lavagens, a membrana foi pré-hibridizada a 42°C, durante 4 h com 5 ml da solução *ULTRAhyb*TM (Ambion) em forno de hibridização Hybaid.

Para as análises da expressão do gene *gsn* foi utilizada como sonda um fragmento do gene *gsn* de aproximadamente 1350 pb subclonado no vetor *pMOSBlue* (construção existente no laboratório). Em relação às análises da expressão de endoxilanasas, as ORFs NCU07225 e NCU02885 foram amplificadas por PCR a partir do cDNA do fungo *N. crassa*, descrito no item 9. Estas sondas foram marcadas radiativamente com α[³²P]dATP por *random priming* utilizando o kit *NEBlot*TM (Biolabs) e purificadas por filtração em resina Sephadex G-50 (Sigma) equilibrada em tampão NT (Tris-HCl 1 mM, pH 8,0, NaCl 1 mM e EDTA 2 mM, pH 8,0) a fim de eliminar os nucleotídeos não incorporados. Em seguida, as sondas foram desnaturadas a 100°C durante 10 minutos e 100 µl da sonda desnaturada (aproximadamente 10⁸ cpm total) foi adicionada à solução de pré-hibridização *ULTRAhyb*TM (Ambion). As hibridizações foram realizadas a 42°C durante por 16 horas. Após a hibridização a membrana foi lavada em diferentes condições de estringência antes de ser exposta a filmes radiográficos (Kodak T-Mat G/RA).

Condições de lavagem:

SSPE 2x - 15 minutos a 42°C.

SSPE 0,5x; SDS 0,1% - 15 minutos a 42°C.

SSPE 0,1x; SDS 0,1% - 15 minutos a 42°C.

8. Análises das linhagens mutantes por citometria de fluxo

Para a obtenção dos conídios das 10 linhagens mutantes no fungo foi utilizado à mesma metodologia descrita no item 3 de Materiais e Métodos. Um total de 5×10^4 conídios foram coletados por centrifugação e suspensos em 0,5 ml da solução gelada de Vindelov (Triton X-100 0.1%, citrato de sódio 0.1%, RNase A 0.9 U/ml, iodeto de propídio 50 µg/mL) e mantidas no escuro por 1,5 h a temperatura ambiente. Um total de 10.000 células em suspensão foram analisadas em relação a seu tamanho, utilizando o parâmetro *Forward Scatter* (FSC), sua complexidade, utilizando o parâmetro *Side Scatter* (SSC) e a incorporação do iodeto do propídio (parâmetros: comprimento de onda da excitação em 488 nanômetro e comprimento de onda da emissão em 585 nanômetros) utilizando o citômetro FACSCalibur™ equipado com o software CellQUEST™ (Becton Dickinson, San Juan, CA, USA) no laboratório do Prof. Rui Curi, no Instituto de Ciências Biomédicas da Universidade de São Paulo, com a colaboração do Pós-Doutorando Dr. Augusto Ducati Luchessi.

Como controle dos experimentos descrito acima, conídios provenientes da mesma suspensão conidial foram também analisadas em relação à viabilidade celular. Para isto os conídios coletados foram diluídos em placas de 96 *wells* onde em cada poço foram colocados 180 µL de meio VM líquido com exceção da primeira coluna onde foi colocado 200 µL de suspensão conidial numa densidade de 10^8 conídios/mL. Diluições seriadas de foram realizadas a partir da amostra inicial até o total de 10^1 conídios/mL. Posteriormente, 5 µL de cada diluição foram transferidos para placas de Petri contendo o meio VM acrescido de sorbose 2%, glicose 0,05% e frutose 0,05%. As placas foram incubadas a 30°C por 16 horas e o crescimento colonial das linhagens mutantes foram comparados ao apresentado pela linhagem selvagem.

9. Clonagem molecular das ORFs NCU06971 e NCU04390 por RT-PCR

A estratégia utilizada para o isolamento do cDNA das ORFs NCU06971 e NCU04390 do fungo *N. crassa* foi por RT-PCR a partir do RNA total. Inicialmente, 4 µg do RNA total da linhagem selvagem do fungo foi tratado com a enzima DNase RQ1 RNase-Free DNase (Promega), para eliminar possíveis contaminações por DNA genômico presente nas amostras de RNA. Em seguida, a síntese do cDNA foi efetuada, utilizando-se o kit comercial *Superscript® III First-strand Synthesis for RT-PCR* (Invitrogen), bem como oligo (dT)₂₀

(Invitrogen) como *primer* reverso, de acordo com as instruções do fabricante. Para a reação de amplificação das ORFs citadas acima, o produto da reação de síntese do cDNA foi utilizado como molde e a enzima utilizada foi a DNA polimerase Phusion® (Thermo Scientific), em um termociclador *Mastercycler gradient* (Eppendorf), de acordo com o protocolo do fabricante. As reações de PCR foram realizadas em um termociclador *Mastercycler gradient* (Eppendorf), no volume final de 50µl, nas seguintes condições: 35 ciclos de 98°C por 10 segundos (desnaturação), 63°C por 30 segundos (anelamento dos *primers*) e 68°C por 2 minutos (extensão), após desnaturação inicial de 96°C de 30 segundos. As sequências do oligonucleotídeos utilizados nas amplificações são mostradas na tabela 1. Os códons iniciadores (ATG) e terminadores (TGA e TAA) estão destacados em negrito.

Tabela 1 - Oligonucleotídeos utilizados para a amplificação das ORFs.

Oligonucleotídeos	Sequências	Enzimas de restrição
6971-F	5'- CATATG TTGTCTAATCCGCTTCAC-3'	<i>NdeI</i>
6971-R	5'- GTCGACTCAGAGGGCCAAGCCAGTTCC -3'	<i>SalI</i>
6971Δ160-R	5'- GTCGACTCAGGCTTGGCTCGGGTTTTTGTG -3'	<i>SalI</i>
4390-F	5'- CATATG GCTTCGGAGCTTACGCC-3'	<i>NdeI</i>
4390-R	5'- GAATTCTTACTGATTCCCCGGACCCG -3'	<i>EcoRI</i>
4390Δ158-R	5'- GAATTCTTAATGATCTGCGGCACTCGAAG -3'	<i>EcoRI</i>
4390Δ483-R	5'- GAATTCTTATGGGCTGAAATGTA CTCGCG-3'	<i>EcoRI</i>

*Os codons de iniciação (ATG) e de término (TGA e TAA) da síntese protéica estão salientados em negrito.

Os produtos das reações de amplificação foram separados por eletroforese em gel de ágarose 0,8% em tampão TAE 1X concentrado (Tris-HCl 40 mM, pH 8,0, ácido acético 19 mM e EDTA 1 mM) contendo 0,5 µg/mL de brometo de etídeo. As bandas correspondentes aos produtos de amplificação foram purificados utilizando o kit *GeneClean® II* (Bio 101) e clonados nos vetores *pMOSBlue* (*GE Healthcare*) ou *pRS314* (SIKORSKI & HIETER, 1989) originando as seguintes construções plasmidiais: *pMOS-6971Δ160*, que possui fragmento de DNA codificador da proteína truncada no resíduo de aminoácido 160, *pMOS-4390Δ158*, que possui fragmento de DNA codificador da proteína truncada no resíduo de aminoácido 158, *pMOS-4390Δ483*, que possui fragmento de DNA codificador da proteína truncada no resíduo de aminoácido 483 e *pRS314-4390*, que possui fragmento de

DNA codificador da proteína inteira. Para a clonagem da ORF inteira NCU06971 foi realizado a digestão com as enzimas *NdeI* e *EcoRI* do produto de PCR purificado e o fragmento liberado foi inserido no vetor pET28a previamente digerido com as mesmas enzimas.

10. Transformação de células competentes de *E. coli*

As construções obtidas na etapa anterior foram utilizadas para transformar células competentes de *E. coli* DH10B. Cerca de 5 µL de cada construção foram adicionados a 100 µL de células competentes e incubados em gelo por 30 minutos. Após esse período, as suspensões celulares foram transferidas rapidamente para um banho a 42°C por 2 minutos para a realização do choque térmico. Após isso foi adicionado 400 µL de meio 2YT a cada suspensão celular. Cada reação de transformação foi então mantida a 37°C por 60 minutos sendo plaqueadas em meio 2YT sólido contendo o antibiótico específico. Por fim, as placas foram submetidas a 37°C, por aproximadamente 16 horas para o aparecimento dos transformantes.

11. Mini-extração de DNA plasmidial

Os transformantes foram inoculados pela transferência das colônias da placa de petri para um tubo de ensaio contendo 5 mL de meio 2YT+0,2% de glicose mais o antibiótico específico e mantidas a 37°C a 250 rpm por 5 horas para extração do DNA plasmidial pelo método de *boiling* (HOLMES & QUIGLEY, 1981). De forma resumida as células foram coletadas pela centrifugação de 3 mL de meio de cultura (14000 rpm por 1 minuto). O precipitado foi então ressuspenso em 350 µL de STET (sacarose 8%, Tris 10 mM pH 8,0, EDTA 50 mM pH 8,0, Triton X-100 0,5%) acrescido de lisozima e, em seguida, a amostra foi aquecida a 100°C por 40 segundos em banho Maria. Após isso as amostras foram centrifugadas a 14.000 rpm por 10 minutos a 4°C em uma microcentrífuga de bancada. A porção líquida da amostra foi transferida para um novo tubo contendo 40 µL de acetato de sódio 3M, pH 5,2 e 420 µL de isopropanol. As amostras foram centrifugadas a 14.000 rpm por 10 minutos a 4°C para a precipitação do DNA plasmidial. O precipitado foi ressuspenso em 100 µL de H₂O e 100 µL de fenol saturado em Tris-EDTA foi adicionado. A amostra foi agitada em vortex e centrifugada a 14.000 rpm por 10 minutos a 4°C e o sobrenadante foi transferido para um novo tubo contendo 40 µL de acetato de amônio 7,5M e 300 µL de etanol absoluto. A amostra foi centrifugada a 14.000 rpm por 5 minutos a 4°C para a precipitação do DNA plasmidial e o precipitado foi lavado 2 vezes com aproximadamente

500 µL de etanol 70%. Após esta etapa o precipitado foi seco ao ar e ressuspenso em 30 µL de H₂O.

12. Produção de proteínas recombinantes em *E. coli*

12.1. Subclonagem das ORFs em vetor de expressão

Após a confirmação dos clones positivos por análise de restrição, os insertos foram removidos dos plasmídios citados anteriormente e transferidos para o vetor de expressão pET-28a (Novagen), o qual produz a proteína recombinante fusionada a uma cauda contendo seis resíduos de histidina (His6). Para isso, os cassetes de DNAs clonados nos vetores pMOS-6971Δ160, pMOS-4390Δ158, pMOS-4390Δ483 e pRS314-4390 foram digeridos com enzimas de restrição apropriadas para a liberação de seus insertos, os quais foram inseridos no vetor pET-28a após sua digestão com as mesmas enzimas originando as construções pET-6971Δ160, pET-4390Δ158, pET-4390Δ483 e pET-4390. Os produtos das reações de ligação foram usados para transformar células competentes de *E. coli* linhagem DH10β e os transformantes foram selecionados em meio 2YT sólido, acrescido de glicose 0,2% contendo canamicina 30 µg/mL. Os clones positivos foram identificados após análise de restrição dos respectivos DNAs plasmidiais com as enzimas de restrição específicas e as sequências nucleotídicas dos insertos foram confirmadas por reações de sequenciamento de DNA.

12.2. Ensaios de indução e análise de solubilidade

Para os ensaios de indução e análise da solubilidade, colônias correspondendo aos transformantes de *E. coli* Rosetta(DE)pLys-S de todas as construções plasmidiais foram inoculadas em 5 mL de meio TB, canamicina 30 µg/mL e cloranfenicol 34 µg/mL e incubadas a 37°C/200 rpm por 16 horas para serem utilizadas como pré-inóculos. Foram inoculados 500 mL de meio TB 1 mL dos respectivos pré-inóculos acrescidos de canamicina 30 µg/mL e cloranfenicol 34 µg/mL e incubadas a 37°C/250 rpm até atingir uma DO_{600nm} entre 0,6 e 0,8. Alíquotas de 2 mL de cada uma das culturas foram removidas e as células foram coletadas por centrifugação a 14.000 rpm por 2 minutos a 4°C, solubilizadas em 200 µL de tampão de lise. Em seguida, foi adicionado IPTG na concentração final de 0,4 mM em cada uma das culturas, para a indução da expressão das respectivas proteínas recombinantes e as culturas foram mantidas a 18°C a 200 rpm por 16 horas. Ao final da indução, as células foram coletadas por centrifugação a 8.000 rpm por 5 minutos a 4°C e mantidas em banho de gelo.

Para verificar a produção das proteínas recombinantes, os precipitados celulares foram solubilizados em 5 mL de tampão de lise (Tris-HCl 50 mM pH 8,0, NaCl 20 mM, glicerol 5%, 0,5% Tween-20, 0,5% Triton X-100, EDTA 5 mM, PMSF 1 mM, benzamidina 10 mM) e as células foram rompidas em sonicador (SONICS, Vibra Cell) com 3 pulsos de 30 segundos e intervalos de 60 segundos, em banho de gelo (potência 30%). As frações solúveis foram coletadas por centrifugação a 20.000 rpm por 10 minutos a 4°C e transferidas para tubos limpos, enquanto que os precipitados insolúveis foram solubilizados em 3 mL de tampão de lise.

As proteínas totais dos extratos celulares obtidos foram quantificadas por Bradford (Bio-Rad), utilizando BSA como padrão. Aproximadamente 20 µg de proteína total de cada um dos extratos totais obtidos das amostras NI (não induzida), S (sobrenadante) e P (precipitado) foram separados por SDS-PAGE 12% no sistema Mini-Protean III (Bio-Rad) e analisados após coloração com Coomassie Blue.

12.3. Análise das proteínas recombinantes por *Western blot*

Análises de *Western blot* foram realizadas para a confirmação da produção das proteínas recombinantes. Inicialmente as proteínas foram fracionadas por eletroforese em gel SDS-PAGE e em seguida foram transferidas para uma membrana de nitrocelulose em tampão Tris, 20 mM, pH 8,0, glicina 150 mM e metanol 20% a 100V, durante 1 hora utilizando o sistema TransBlot (Bio-Rad). Após a transferência, as membranas foram bloqueadas com uma solução de 5% de leite em pó desnatado em TBST (Tris, 20 mM pH 7,5; NaCl 0,5 M; Tween-20 0,05%). A membrana foi lavada uma vez com tampão TBST por cinco minutos à temperatura ambiente e incubada com o anticorpo monoclonal de coelho anti-His conjugado com fosfatase alcalina (Sigma), na diluição de 1:10.000, em TBST contendo 5% de leite, durante 1 hora. Posteriormente, a membrana foi lavada com TBST três vezes por cinco minutos cada, a temperatura ambiente e revelada com tampão de revelação (NaHCO₃ 100 mM, MgCl₂ 1 mM, substratos para fosfatase alcalina NBT e BCIP).

12.4. Purificação das proteínas recombinantes por cromatografia de afinidade

Após a determinação das melhores condições para a expressão das proteínas na forma recombinante, foi efetuada a indução dos clones produtores das cinco proteínas para a produção das respectivas proteínas em maior escala (1 L). As células de *E. coli* foram coletadas por centrifugação e o precipitado foi ressuspensão em 80 ml de tampão de lise (Tris HCl, 50 mM pH 8,0, NaCl 300 mM, imidazol 500 mM, glicerol 10%, PMSF 1 mM e benzamidina 10 mM) e as células foram rompidas por ultrassom. Após esta etapa, o lisado

celular foi centrifugado com o intuito de separar o sobrenadante do precipitado. Por fim, o sobrenadante foi filtrado em um sistema de filtração Millipore utilizando uma membrana de 0,22 μm . Em seguida, as proteínas recombinantes foram purificadas por meio de cromatografia de afinidade em coluna HisTrap HP (GE) em um aparelho Äkta Prime (GE). Para a purificação das cinco proteínas na forma recombinante, equilibrou-se a coluna com tampão contendo Tris HCl, 50 mM pH 8,0, NaCl 300 mM, imidazol 50 mM e glicerol 10%, e todo o volume do sobrenadante filtrado foi passado pela coluna. Após esta etapa, as proteínas aderidas à coluna foram eluídas com o mesmo tampão e um gradiente de imidazol de 50-750 mM e fracionadas por SDS-PAGE a 200 V por 1 hora utilizando géis de 12% (Laemmli, 1970). As melhores frações contendo as proteínas recombinantes foram unidas e dialisadas contra 1L de tampão Tris-HCl, 50 mM, pH 7,6, 1 mM β -mercaptoetanol, 1 mM PMSF e glicerol 20% por 24 horas. Após isso as proteínas foram estocadas a -80°C .

13. Determinação dos sítios de ligação ao DNA por CHIP-Seq

A técnica de imunoprecipitação seguida do sequenciamento de DNA de alto desempenho (ChIP-Seq) foi utilizada para a determinação dos elementos de DNA alvo da proteína hipotética codificada pela ORF NCU04390. Esta etapa do trabalho foi realizada no laboratório do Prof. Dr. Michael Freitag no "Department of Biochemistry and Biophysics, Oregon State University", Corvallis, OR, EUA. Aproveitando a oportunidade do estágio no laboratório do Dr. Freitag, estes experimentos foram realizados também com os fatores de transcrição codificados pelas ORFs NCU01629 e NCU03043, ambas anotadas como proteínas hipotéticas e identificadas como reguladores do metabolismo de glicogênio por uma Doutoranda do nosso grupo. Neste trabalho, estamos apresentando os resultados relacionados à ORF NCU04390.

13.1. Preparo dos *split markers* para a produção de proteínas fusionadas a GFP

Pelo fato da proteína codificada pela ORF NCU04390 estar anotada no genoma como proteína hipotética, foi inserido uma cauda GFP em sua região C-terminal para viabilizar a análise por ChIP. Foi utilizado o vetor pZErO-GFP-hph (HONDA & SELKER, 2009) como DNA molde para a síntese por PCR do cassete contendo o DNA codificador da cauda GFP a ser inserido no genoma do fungo. Este fragmento de DNA possui 2.281 pb e foi amplificado utilizando o kit Phusion Flash 2X *Master Mix* (Finnzymes) utilizando os oligonucleotídeos utilizados utilizando foram 10XGly-F e LoxP-R (Tabela 2), nas seguintes condições: 35 ciclos de 98°C por 5 segundos (desnaturação), 58°C por 5 segundos (anelamento dos *primers*) e

Tabela 2 - Oligonucleotídeos utilizados para a síntese dos *splitmarkers*.

Oligonucleotídeos	Sequências
4390GLY-F	5'- GGAGCCTAATGGACTTCGTG-3'
4390GLY-R	5'- CCTCCGCCTCCGCCTCCGCCGCCTCCGCC CTGATTCCCC GGACCCGGTT -3'
4390LOX-F	5'- TGCTATACGAAGTTATGGATCCGAGCTCGGGGAGGCTAGG AGATAGGCG -3'
4390LOX-R	5'- ACTCAGGAGTGCTTGCTGCT -3'
10XGly-F	5'- GGCGGAGGCGGCGGAGGCGGAGGCGGAGG -3'
loxP-R	5'- CGAGCTCGGATCCATAACTTCGTATAGCA -3'
hph_SM-R	5'- TCGCCTCGCTCCAGTCAATGACC-3'
hph_SM-F	5'- AAAAAGCCTGAACTCACCGCGACG-3'

*As sequências nucleotídicas acrescentadas na região 5'-flanqueadora nos primers 4390GLY-R e 4390LOX-F estão salientadas em negrito.

68°C por 1 minuto (extensão), após desnaturação inicial de 98°C de 10 segundos. O mapa do plasmídeo pZErO-GFP-hph está mostrado na figura 5.

Em seguida foram sintetizados oligonucleotídeos para a amplificação de uma região *upstream* (4390GLY-F e 4390GLY-R) e uma região *downstream* (4390LOX-F e 4390LOX-R) ao *stop codon*, com aproximadamente 1 kb cada. É importante ressaltar que foi adicionada a sequência nucleotídica 5'-**CCTCCGCCTCCGCCTCCGCCGCCTCCGCC**-3' na região 5' *overhang* do oligonucleotídeo 4390GLY-R e a sequência nucleotídica 5'-**TGCTATACGAAGTTATGGATCCGAGCTCG**-3' na região 5' *overhang* do oligonucleotídeo 4390LOX-F, para a construções dos *splitmarkers* pela técnica de *Fusion-PCR*. A reação de amplificação tanto da região *upstream* quanto *downstream* ao *stop codon* utilizando os respectivos oligonucleotídeos foram realizadas nas mesmas condições para a amplificação da sequência nucleotídica codificadora de GFP descrito anteriormente.

Para a síntese do *splitmarkers* 5' foram utilizados os primers 4390GLY-F e LoxP-R bem junto a aproximadamente 300 ng dos fragmentos de PCR obtidos pela amplificação do vetor mais o *amplicom* 5' ao *stop codon*. Para a síntese dos *splitmarker* 3' foram utilizados os primers 4390LOX-F e hph_SM-R junto a aproximadamente 300 ng dos fragmentos de PCR obtidos pela amplificação do vetor mais o *amplicon* 3' ao *stop codon*. Após a obtenção dos *splitmarkers*, 400 ng de cada um foram utilizados para a transformação de conídios da linhagem selvagem (FGSC#9718, a) do fungo *N. crassa*.

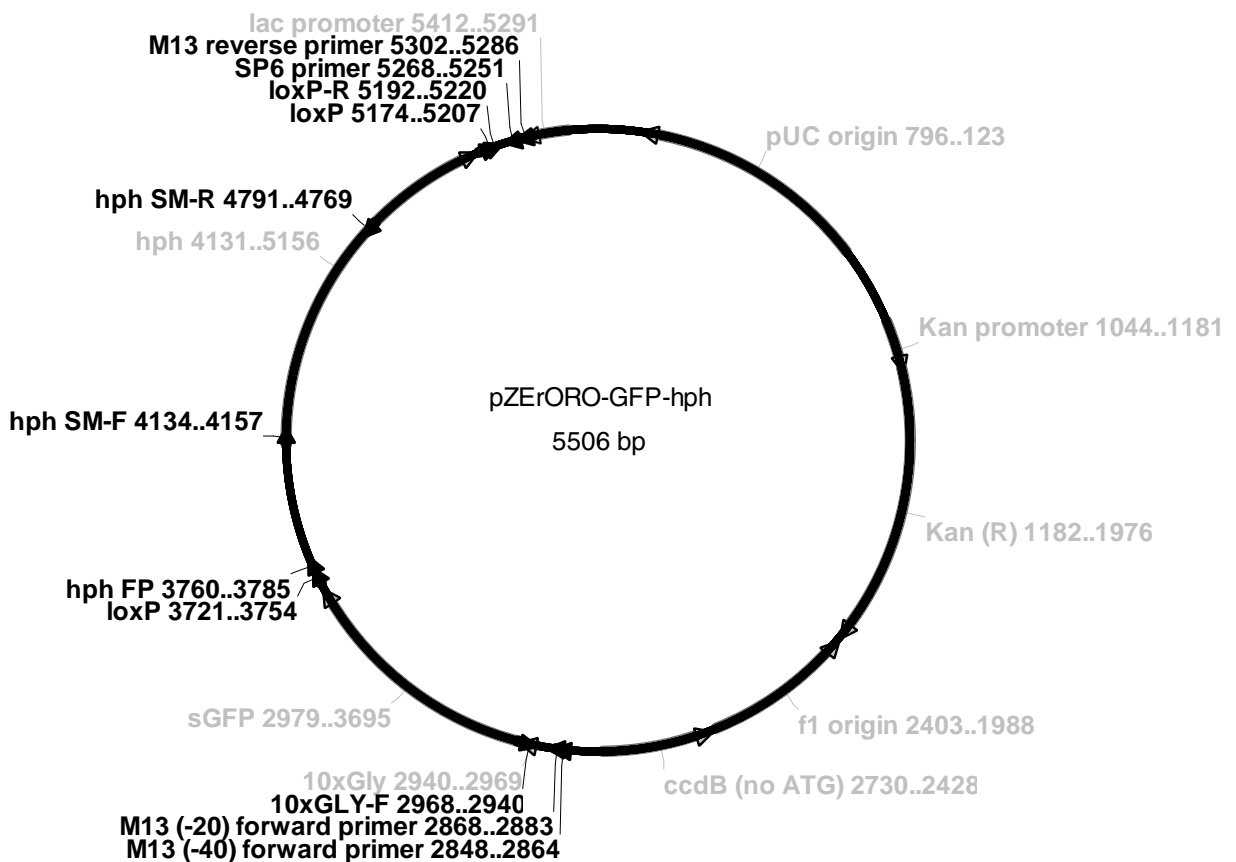


Figura 5 - Mapa do plasmídio pZErO-GFP-*hph* (HONDA & SELKER, 2009) gerado pelo software ApE v.2.0.37 (<http://biologylabs.utah.edu/jorgensen/wayned/ape>).

13.2. Transformação de conídios do fungo *N. crassa*

A inserção das construções de DNA em *N. crassa* foi realizada através da eletroporação de conídios. Os conídios da linhagem FGSC#9718 (a) utilizados para a transformação foram obtidos como descrito anteriormente no item 3. Após sua coleta, este material foi centrifugado a 3000 rpm utilizando um tubo de 15 mL do tipo *Falcon*, o sobrenadante foi descartado e os conídios foram ressuspensos em 1,5 mL de uma solução de sorbitol 1M. Para a transformação do fungo *N. crassa* foram adicionados 400 ng de cada um dos *splitmarkers* a 50 µL da suspensão conidial deste material em um tubo tipo *Eppendorf* estéril, e este material foi mantido em gelo por 10 minutos. Em seguida transferiu-se o conteúdo do tubo para uma cubeta de eletroporação estéril e eletroporadas em um

eletroporador Gene Pulser II (BioRad) nas seguintes condições: 1500 Volts, 600 Ω , 25 μ Faradays.

Após a eletroporação, foi adicionado 1 mL de sorbitol 1M à solução eletroporada no interior da cubeta e 1/3 deste volume foi transferido para um tubo estéril contendo 5 mL de VM 1X sem fonte de carbono. Este tubo foi agitado a 200 rpm a 30°C por 1 hora para a recuperação dos conídios. Após esta etapa, foram adicionados 5 ml de *top ágar* pré aquecido a 55°C. Os tubos foram então agitados por inversão e dispensados em placa contendo *base ágar* contendo 4 μ g/ml de higromicina (Sigma) e sorbose para induzir o crescimento colonial. Este material foi deixado a 30°C por 2 dias para o aparecimento dos transformantes na forma colonial. As colônias isoladas foram transferidas para tubos de ensaio contendo 1 mL de meio mínimo de Vogel + sacarose 2% acrescidos de 4 μ g de higromicina B. As culturas foram crescidas até o aparecimento de conídios, os quais foram inoculados em culturas líquidas para a obtenção de micélios utilizados na extração de DNA genômico. O restante dos conídios foi ressuspenso em sorbitol 1M e armazenado a -80°C, podendo ser utilizado em novas reações de transformação.

13.3. Extração de DNA genômico

A extração do DNA genômico foi realizada segundo POMRANING et al. (2008). Conídios coletados de acordo com o item 3 foram inoculados em tubos de ensaio contendo VM líquido contendo sacarose 1,5% e mantidos por 48 horas a 30°C sem agitação. Esses tubos foram agitados em vortex para a submersão total das hifas. Os micélios foram coletados por palitos de madeira estéreis e transferidos para tubos de polipropileno, sendo então congelados a -20°C por 2 horas e liofilizados por 16 horas. Após a liofilização foi inserido no interior do tubo uma espátula e a amostra foi agitada em vortex até a sua pulverização. Em seguida foi adicionado 600 μ L de uma solução de sal-detergente (para 500 mL: 2g de desoxicolato de sódio; 58,5g de Brij; 58,44g de NaCl) em cada tubo. As amostras foram agitadas em vortex por aproximadamente 30 segundos e incubadas à temperatura ambiente por 10 minutos. Após esse período, as amostras foram agitadas novamente e o conteúdo foi transferido para um tubo *Eppendorf* de 1,5 mL e centrifugado a 13.000 rpm por 10 minutos à temperatura ambiente. O sobrenadante foi transferido para um novo tubo de 1,5 mL contendo 1 mL de uma solução gelada de TCA-Na 4,5 M em etanol. As amostras foram agitadas gentilmente por inversão e armazenadas por 16 horas a -20°C. Após esse período as amostras foram centrifugadas a 13.000 rpm por 5 minutos a 4°C e o sobrenadante foi desprezado. Em seguida o precipitado foi ressuspenso com uma solução de 200 μ L acetato de amônio 10 mM contendo 0,3 μ g/mL de RNase A e agitado em vortex rapidamente. Após esta etapa as amostras foram incubadas a 50°C por 45 minutos sendo

agitadas em vortex a cada 15 minutos a fim de ressuspender totalmente o DNA. Foi adicionado em cada amostra 200 µL de clorofórmio. Essas amostras foram agitadas em vortex sendo em seguida centrifugadas por 5 minutos a 13.000 rpm em temperatura ambiente. O sobrenadante foi transferido para um novo tubo contendo 900 µL de acetato de amônio em isopropanol (adicionar 42,5 ml de isopropanol em 7,5 mL de acetato de amônio 5M). Essas amostras foram misturadas por inversão do tubo e centrifugadas a 13.000 rpm por 5 minutos a temperatura ambiente e o sobrenadante foi descartado. Para a lavagem do precipitado, foram adicionados 300 µL de etanol 70% ao precipitado, e as amostras foram agitadas rapidamente em vortex, centrifugadas a 13.000 rpm por 5 minutos, descartando o sobrenadante em seguida. As amostras foram secas ao ar e ressuspensas em 50 µL de H₂O.

13.4. Confirmação por *Southern blot*

Para a confirmação da introdução da sequência codificadora da cauda GFP na ORF codificadora da proteína hipotética, 4 µg de DNA genômico foram digeridos com a enzima de restrição *SacI* por 16 horas a 37°C. Posteriormente, as amostras digeridas foram submetidas à eletroforese em gel de ágarose 0,8% em uma voltagem de 150 V durante 6 h. Após eletroforese, o gel de ágarose foi transferido a um recipiente contendo HCl 0,2 M por 20 minutos para sua depurinização. Em seguida o gel foi transferido a outro recipiente contendo NaOH 0,4 M por 20 minutos. Os fragmentos de DNA foram transferidos por capilaridade para membrana de Nylon carregada positivamente (Hybond N+, GE Healthcare), utilizando solução de NaOH 0,4 M por 16 horas. Após esse período a membrana foi removida e então, exposta a luz ultravioleta por 1 minuto em um "crosslynker" (Ultraviolet Crosslynker UVP, modelo CL-1000).

13.5. Condições de hibridização e detecção

Inicialmente as membranas foram bloqueadas com 20 mL de uma solução de pré-hibridização *Zeta wash solution* (para 1 L: 80 mL de Na₃PO₄ 1M, 50 mL de NaCl, 8 mL de EDTA 0,5 M, 40 mL de SDS 10%) em tubos a 65°C por 15 minutos. A sonda utilizada foi o fragmento codificador da cauda GFP e foi marcada radiativamente com α^[32P]dATP por *random priming* utilizando o kit *NEBlot*TM (Biolabs). Após esta etapa, 100 ng da sonda de DNA marcada radioativamente foi desnaturada por 5 minutos a 100°C e adicionada ao tubo contendo solução de pré-hibridização. As amostras foram mantidas em forno de hibridização a 65°C por 16 horas. Após isso, a solução de pré hibridização contendo a sonda foi removida, e a membrana foi lavada 4X, a 15 minutos cada. Após as lavagens a membrana

foi envolvida em um filme plástico e transferida para um cassete PhosphorImager e revelada, após 16 horas, no *scanner* Storm820 (GE).

13.6. Cruzamento das linhagens de *N. crassa* e seleção dos transformantes

Os cruzamentos foram realizados de acordo com o procedimento descrito por DAVIS & DE SERRES (1970). Inicialmente a linhagem receptora FGSC9718 (A) foi inoculada em placas contendo meio SC e incubadas à temperatura ambiente no escuro por 3 dias. Após esse período, uma suspensão conidial do transformante selecionado foi inoculada sobre a linhagem receptora e as placas foram incubadas na posição invertida por 12 dias à temperatura ambiente. Após este período observou-se a formação do peritécio, e os ascósporos foram liberados nas tampas das placas. Os ascósporos foram coletados com 1 mL de água estéril, introduzidos em um tubo *Eppendorf* de 1,5 mL e centrifugados a 1.000 rpm por 5 minutos. Após a centrifugação 800 µL do sobrenadante foi desprezado e 20 µL da suspensão de ascósporos foram plaqueados em uma placa contendo meio VM 1X sólido contendo sacarose 1,5% acrescido de higromicina 4 µg/µL de meio. Estas placas foram incubadas a 65°C por 50 minutos para a ativação dos ascósporos, bem como inativação das hifas e conídios. Após esta etapa, as placas foram mantidas à temperatura ambiente por 16 horas. Finalmente, com o auxílio de um estereomicroscópio, ascósporos em germinação foram removidos das placas, com o auxílio de uma espátula e inoculados em tubos de ensaio contendo meio sólido VM 1X, sacarose 1,5%, acrescidos de higromicina 4 µg/µL de meio e mantidos por 2 dias a 30°C.

13.7. Imunoprecipitação da cromatina (ChIP)

Os conídios do fungo *N. crassa* foram obtidos de acordo com o item 3. Uma suspensão de conídios (densidade de 10^7 células/ml) foi inoculada em 30 ml de meio VM (Erlenmeyers de 125 mL), e colocada a 30°C sob agitação constante de 250 rpm por 24 horas em um agitador New Brunswick. Após esse período, o micélio foi coletado por filtração e dividido em duas partes equivalentes. Uma foi transferida para um Erlenmeyer de 125mL contendo formaldeído 1% em 30 mL de PBS 1X, permanecendo sob agitação de 100 rpm à temperatura ambiente por 15 minutos. A parte outra foi transferida para outro frasco contendo VM líquido, pré-aquecido a 45°C, e deixado nesta temperatura por 30 minutos. O micélio foi coletado por filtração e transferido para um Erlenmeyer de 125 mL contendo formaldeído 1% em 30 mL de PBS 1X, permanecendo sob agitação de 100 rpm à temperatura ambiente. Em seguida, foram adicionados 1,5 mL de glicina 2.5 M e os micélios submetidos e não submetidos ao choque térmico foram coletados por centrifugação. Cerca

de 300 mg dos materiais coletados foram lavados com PBS 1X e transferidos para um tubo *Eppendorf* contendo 0,5 mL de tampão de lise acrescido de inibidores de proteases (para 10 mL: 8 µL de H₂O; 500 µL de Hepes-KOH 1M pH 7,5; 180 µL de NaCl 5M; 250 µL de butirato de sódio 2M; 20 µL de EDTA 0,5M; 1 mL de Triton X-100 10%; 100 µL de desoxicolato de sódio 10%; 100 µL de PMSF 0,1M; 10 µL de leupeptina 10 mM; 10 µL de E-64 1mM; pepstatina 1µg/mL). Os micélios coletados foram lisados após 30 ciclos de sonicação realizados em um sonicador *Bronson Sonifer 450* (*output 2, duty cycle 80*) e, em seguida, foram mantidos no gelo por 5 minutos.

Para a fragmentação da cromatina, foram realizados 6 ciclos de 20 pulsos de sonicação num sonicador *Bronson* (*output 1,2, duty cycle 80*) com intervalos de aproximadamente 3 minutos entre cada ciclo. Os materiais lisados foram centrifugados em uma centrífuga de bancada a 14.000 rpm por 5 minutos a 4°C e os sobrenadantes foram transferidos para um novo tubo *Eppendorf*. Cerca de 20 µL deste material foi reservado e armazenado a -20°C para servir como amostra controle (*input*). Em seguida, foram adicionados 20 µL de proteína A/G ligada a *beads* de ágarose (Invitrogen) previamente equilibrada em 200 µL tampão de lise. Este material foi mantido a 4°C por 2 horas sob agitação em mesa agitadora. Após isso, o extrato celular contendo as *beads* foi centrifugado a 5.000 rpm por 1 minuto e o sobrenadante foi transferido para um novo tubo. Foram adicionados 2 µL do anticorpo anti-GFP Ab290 (Abcam) em cada uma das amostras, as quais foram mantidas a 4°C sob agitação em mesa agitadora por cerca de 16 horas. Após esse período, o material foi centrifugado e o sobrenadante foi desprezado. Foi adicionado 1 mL de tampão de lise gelado sem inibidores de proteases ao material precipitado e deixado por 10 minutos a 4°C sob agitação branda em mesa agitadora. Após esse período as amostras foram centrifugadas a 5.000 rpm por 1 minuto e o sobrenadante foi desprezado. Esta última etapa de lavagem com tampão de lise e coleta das amostras por centrifugação foi repetido mais uma vez. Em seguida, foi adicionado 1 mL de tampão de lise acrescido de NaCl 0,5 M às amostras, as quais foram mantidas por 10 minutos a 4°C sob agitação branda em mesa agitadora, sendo então centrifugadas a 5.000 rpm por 1 minuto a 4°C e o sobrenadante descartado.

Foram adicionados 1 mL de tampão de lavagem (para 200 mL: 167,6 mL de H₂O; 2 mL de Tris-HCl pH 8,0; 10 mL de LiCl 5M; 10 mL de NP-40; 10 mL de desoxicolato de sódio 10% e 400 µL de EDTA 0,5 M) em cada amostra, as quais foram mantidas sob agitação branda por 10 minutos a 4°C. Após esse período, o material foi precipitado por centrifugação e os sobrenadantes foram descartados. Foi adicionado aos materiais precipitados 1 mL de TE e os tubos foram mantidos sob agitação branda por 10 minutos a 4°C. Estes materiais foram então precipitados por centrifugação e os sobrenadantes foram descartados. Finalmente, foram adicionados aos precipitados 62,5 µL de TES (para 50 mL: 41,5 mL de H₂O; 2,5 mL de

Tris-HCl pH8,0; 1 mL de EDTA 0,5M; 5 mL de SDS 10%) e as amostras foram mantidas a 65°C por 10 minutos, sendo agitado em vortex a cada 2 minutos. Os materiais foram precipitados por centrifugação a 100 rpm por 5 minutos a 4°C e o sobrenadante foi transferido um novo tubo. Foram então acrescentados mais 62,5 µL TES aos materiais precipitados e lavados da mesma forma. Os sobrenadantes foram então unidos aos 62,5 µL TES, anteriormente reservado. Finalmente, foram adicionados 105 µL a cada amostra separada no dia anterior (*input*), sendo então mantidas a 65°C por 16 horas.

Após essa etapa foi adicionado ao material 125 µL de H₂O e 2,5 µL de RNase A (10 mg/mL) e as amostras foram incubadas por 2 horas a 50°C. Em seguida, foram adicionados 6,25 µL de proteinase K (20 mg/mL) em cada amostra, as quais foram mantidas a 50°C por duas horas. Após esta etapa foram adicionados 250 µL de fenol:clorofórmio:álcool isoamílico (25:24:1) em cada amostra, as quais foram agitadas em vortex, centrifugadas a 13.000 rpm por 5 minutos e as fases aquosas foram transferidas para um tubo *Eppendorf* contendo 250 µL de clorofórmio. As amostras foram agitadas rapidamente em vortex, centrifugadas a 13.000 rpm por 5 minutos a 4°C e a fase aquosa foi transferida para um tubo novo. Foram adicionados 25 µL de acetato de sódio 3M pH 5,2, 865 µL de etanol absoluto e 1 µL de uma solução de glicogênio 20%. As amostras foram mantidas por 16 horas a -20°C. Por fim, as amostras foram centrifugadas a 13.000 rpm por 10 minutos a 4°C e o precipitado foi ressuspenso em 30 µL de H₂O.

13.8. Preparo da biblioteca de CHIP-Seq

Para a realização das análises de sequenciamento de alto desempenho, foi necessário que os produtos da imunoprecipitação tivessem suas extremidades processados por enzimas modificadoras, permitindo a ligação de adaptadores específicos para a síntese da biblioteca de CHIP-Seq. Esta etapa foi realizada de acordo com POMRANING, SMITH e FREITAG, (dados não publicados). De maneira geral, os produtos das imunoprecipitações foram tratados com as enzimas T4 DNA polimerase para o preenchimento das extremidades 5', *Klenow* para o preenchimento das extremidades 3' por *fill in* e T4 polinucleotideo kinase para a fosforilação da extremidade 5' da *New England Biolabs* (NEB). Para isso, foram utilizados 200-400 ng de DNA e acrescentados de 5 µL de tampão da enzima T4 DNA polimerase 10X com 10 mM de ATP, 1 µL de dNTPs 20 mM, 2,5 µL da enzima T4 DNA polimerase (3 U/µL), 0,5 µL de *Klenow* (exo⁺) (1 U/µL) e 2,5 ul da enzima T4 polinucleotideo kinase (10 U/ µL) e o volume foi completado para 50 µL com H₂O. Este *mix* foi mantido a 20°C por 30 minutos e o DNA foi purificado utilizando o kit de purificação QIAquick (Qiagen) e eluída com 32 µL de 10 mM Tris-HCl (pH 8,5). Em seguida foi adicionado uma andenina na extremidades 3' dos fragmentos de DNA tratados anteriormente. Para isso, foram

adicionados aos 32 μL de DNA purificados 5 μL de tampão 1 da enzima *Klenow* 10X, 10 μL de ATP 1 mM e 3 μL de *Klenow* (exo-) (5 U/ μL). A reação foi incubada a 37° C por 30 minutos. Após isso o DNA foi purificado pelo *kit* PCR MinElute (Qiagen) e eluído com 10 μL de 10 mM Tris-HCl (pH 8,5). Foram sintetizados dois oligonucleotídeos complementares, na concentração de 10 mM e quantidades iguais de cada um dos oligos foram misturadas e desnaturadas por calor a 98°C durante 5 minutos e transferidas para temperatura ambiente para a formação de um complexo de DNA dupla fita.

Adaptadores dupla-fitas foram adicionados, na razão molar de 10:1, aos produtos de ChIP contendo as extremidades tratadas contendo adenina adicionada na posição 3' *overhang*. Assim, 5 μL de tampão da enzima T4 DNA ligase 10X (NEB), 2,5 μL da enzima T4 DNA ligase (1 U/ μL), e 1 μL dos adaptadores foram adicionados a 10 μL do DNA tratado no item anterior. O volume foi ajustado para 50 μL com H₂O. Esta reação foi mantida à temperatura ambiente por 15 minutos. Esta reação foi fracionada em eletroforese de gel de ágarose 2% NuSieve para a separação dos fragmentos de DNA ligados aos adaptadores dos adaptadores não ligados e foram excisados do gel as regiões que continham os fragmentos entre 200 e 500 pb. O DNA foi purificado utilizando o kit QIAquick (Qiagen) e eluído em 30 μL de 10 mM Tris-HCl, pH 8,5.

Por fim, para o preparo das bibliotecas de ChIP-Seq foram realizadas reações de PCR utilizando 1 μL do produto de ChIP ligado aos adaptadores, 25 μL do *Phusion DNA polymerase Mix Master* (Finnzymes), 1 μL de oligonucleotídeos *forward* e *reverse* e 22 μL de H₂O. Reações de PCR de 21 ciclos foram realizadas (10 s, a 98 °C, 30 s, a 65 °C, 30 s, a 72°C), após uma desnaturação inicial de 98°C, com o intuito de evitar a amplificação seletiva de regiões específicas. Isto é seguido por uma extensão de 5 min a 72°C. Os produtos de PCR foram quantificados por absorvância a 260 nm e 5 μL da reação foi fracionado em eletroforese em gel de ágarose 1,5% aga a fim de verificar o tamanho dos fragmentos da biblioteca de Chip-Seq.

13.9. Sequenciamento de DNA

O sequenciamento de alto desempenho foi realizado em um sequenciador HiSeq 2000 (Illumina/Solexa – San Diego – CA) no *Center of Genomic Research and Bioinformatic* (CGRB) *Core Lab* da Oregon State University (Corvallis, OR, EUA) utilizando o *kit* GD-401-2510 TruSeq SR Cluster Kit v2.5- cBot- HS e FC-401-1002 TruSeq SBS Kit - HS (50 ciclos).

13.10. Análise computacional dos resultados de ChIP-Seq

Para a análise dos dados gerados pelo sequenciador foi necessário a conversão dos arquivos fornecidos pelo sequenciador, no formato Fastq, para o formato BAM, bem como a realização dos alinhamentos dos dados obtidos com o genoma de referência para a determinação exata dos sítios de ligação ao DNA. Para isso, foi utilizado o *software* Bowtie, o qual possui um algoritmo muito eficiente que funciona como um alinhador de sequências curtas de DNA sobre o genoma do organismo de interesse (formato FASTA). Tais dados foram convertidos inicialmente para o formato SAM seguido da conversão para o formato BAM, o qual é aceito para a validação dos picos. Os picos de ChIP-Seq foram visualizados utilizando o *software* Integrative Genome viewer (<http://www.broadinstitute.org/igv/>), o qual possibilita a verificação de picos binomiais. A validação dos picos de ChIP-Seq foi realizada pelo *software* MACS 1.4.2 (<http://liulab.dfci.harvard.edu/MACS/>). Este *software* modela o comprimento dos fragmentos de ChIP sequenciados, e os utiliza para melhorar a resolução espacial dos possíveis sítios de interação da proteína de interesse com o genoma.

Por fim, para determinação dos *motifs* de DNA no qual a proteína hipotética possivelmente interage com o genoma, foram selecionadas as sequências nucleotídicas nas quais os 50 picos de DNA com o maior *score* obtido pelo *software* MACS v1.4.2 foram obtidos e suas regiões conservadas foram analisadas utilizadas pelas ferramentas *online* Scope *motif finder* v 2.1.0 (<http://genie.dartmouth.edu/scope/>) e *BioProspector* (<http://robotics.stanford.edu/~xslu/BioProspector/>).

14. *Electrophoretic mobility shift assay* (EMSA)

14.1. Obtenção e marcação das sondas

Para a verificação de interação DNA-proteína *in vitro* foram realizados ensaios de retardamento em gel (EMSA) utilizando as proteínas codificadas pelas ORFs NCU04390 e NCU06971 produzidas na forma recombinante, tanto inteira quanto suas formas truncadas. Fragmentos de DNA (sondas) contendo os elementos *cis* de DNA específico para a proteína XLR-1 (5'-GGCTGA-3'), bem como para a proteína hipotética codificada pela ORF 4390 obtidos por meio do ChIP-Seq (5'-TACATA-3') presentes na região 5'-flanqueadora de possíveis genes alvos foram obtidos a partir de reações de PCR. Tais reações foram realizadas com um volume final de 50 µL utilizando 1,0 µL de oligonucleotídeos específicos a 100 pmol/µL, 5 µL de tampão de PCR 10X, 2,5 µL de MgCl₂ 50 mM, 1 µL de dNTP (10 mM), 0,5 µL da enzima *Taq DNA polimerase* (Invitrogen), 400 ng de DNA genômico e 0,5 µL de α-[³²P]dATP para sua marcação radioativa. Para estas reações de PCR foram realizados 35 ciclos de 96°C por 1 minuto (desnaturação), 58°C por 1 minuto (anelamento dos

oligonucleotídeos) e 72°C por 1 minuto (extensão). As sequências nucleotídicas dos oligos utilizados estão na Tabela 3.

Para a purificação das sondas de DNA, as reações de PCR foram inicialmente precipitadas com 1/10 do volume de acetato de sódio 3M, pH 5,2 e 500 µL de etanol absoluto gelado e purificadas em após fracionamento em eletroforese em gel de ágarose *low melting point* a 2% (Sigma) em tampão TAE 1X. Os fragmentos foram excisados do gel de ágarose e transferidos para tubos *Eppendorf* de 1,5 mL e aquecidos a 65°C até o total derretimento do gel. Após esta etapa foram adicionados 200 µL de fenol em cada tubo e a mistura foi agitada rapidamente em vortex e centrifugadas a 13.000 rpm por 5 minutos à temperatura ambiente. O sobrenadante foi transferido para um novo tubo *Eppendorf* contendo 500 µL de uma solução fenol:clorofórmio:álcool isoamílico na proporção 25:24:1. Os tubos foram novamente agitados no vortex e centrifugados a 13.000 rpm por 5 minutos à temperatura ambiente. O sobrenadante foi transferido para um novo tubo contendo 1/10 do volume do sobrenadante de acetato de sódio 3M, pH 5,2 e 500 µL de etanol gelado. Os tubos foram deixados a -80°C por cerca de 16 horas e centrifugados a 14000 rpm a 4°C por 30 minutos para a precipitação do DNA marcado radioativamente. Em seguida o precipitado foi lavado com etanol 70%, seco ao ar e ressuspensão em 50 µL de água. Após isso, a radioatividade incorporada ao DNA foi medida por cintilação líquida em um cintilador Beckman LS6500. A concentração do DNA foi determinada por absorvância utilizando um comprimento de onda de 260 nm em um espectrofotômetro modelo DU640.

Tabela 3 - Oligonucleotídeos utilizados para a síntese das sondas de DNA para os ensaios de EMSA.

Oligonucleotídeos	Sequências
XLNR-FP2	5'-TGAGGGTGAGAAAGTTGC-3'
XLNR-RP2	5'-TATTTCTGAACGGA ACTCC-3'
7225-F	5'- AGTGTGCTAAAAATCGGCTTC -3'
7225-R	5'- GTAAAAGCTTCCGTTCTATC -3'
2855-F	5'-AAATGGAAGCTCGATGGAAG -3'
2855-R	5'- CCATTCTACACTACCCCGGC-3'
4390deb-F	5'-TACTACGATTTGGACCTCTC-3'
4390deb-R	5'-GTCTTACGGTTGTTGGTCTC-3'

Foram sintetizados competidores específicos para as sondas citadas acima a partir de reações de PCR, sem a adição do α -[³²P]dATP. Os oligonucleotídeos utilizados para estas reações bem como as condições das reações de PCR foram as mesmas. Como competidor inespecífico foi utilizado o copolímero poli(dI-dC) (GE Healthcare).

14.2. Ensaios de retardamento em gel

Os ensaios para a verificação de possíveis interações proteína-DNA *in vitro* foram realizadas em 30 a 100 μ L de uma reação contendo tampão de ligação 1X (Tris-HCl 125 mM, pH 7,9, KCl 100 mM, EDTA 0,5 mM, DTT 0,05 mM e glicerol 50%) contendo uma mistura de inibidores de proteases (PMSF 0,025 mM, benzamidina 5 μ g/mL, antipaína A 5 μ g/mL, pepstatina 5 μ g/mL e NaF 5 μ g/mL), 2 μ g/reação de competidor inespecífico poli(dI-dC) e quantidades variáveis das proteínas produzidas na forma recombinante. Inicialmente, diferentes concentrações das proteínas recombinantes foram testadas em tampão de ligação à temperatura ambiente junto a 0,6 ng das sondas de DNA marcadas radioativamente (10^4 com) por 20 minutos. Para os ensaios de competição, o competidor específico foi inicialmente adicionado à reação de ligação e após 10 minutos foram adicionadas as sondas marcadas. O competidor específico foi adicionado a excesso molar de 10 a 40X superior em relação à sonda marcada. Após as reações de ligação, as reações foram submetidas à eletroforese em gel de poliacrilamida 5% (29:1) não desnaturante, em tampão TBE 0,5X a 15°C, 300 V e 10 mA por 4 ou 5 horas. Ao término da eletroforese, o gel foi seco e autorradiografado em filme Kodak T-Mat G/RA.

Resultados

1. Fatores de transcrição selecionados

Este trabalho é decorrente dos estudos realizados no Mestrado, quando uma coleção de linhagens mutantes do fungo *N. crassa* contendo genes codificadores de fatores de transcrição individualmente nocauteados foi adquirida pelo laboratório. A coleção foi utilizada com o objetivo de identificar fatores de transcrição que regulam o metabolismo do glicogênio no fungo *N. crassa* (GONÇALVES et al., 2011). O conteúdo de glicogênio de todas as linhagens mutantes foi determinado, antes e depois do choque térmico, revelando a existência de 10 linhagens com um padrão de acúmulo do carboidrato distinto da linhagem selvagem. A partir deste resultado, estas 10 linhagens mutantes do fungo foram selecionadas para que estudos mais aprofundados fossem realizados na tentativa de compreender melhor como tais fatores de transcrição regulam o metabolismo do glicogênio. Os valores do conteúdo de glicogênio antes e após choque térmico destas linhagens são representados na figura 6A. Um fato de extrema importância é que dos 10 fatores de transcrição selecionados, apenas 2 deles possuem funções conhecidas, o codificado pela ORF NCU08000, anotado como fator de transcrição cutinase 1 α (CTF1 α) e o codificado pela ORF NCU06971, anotado como ativador transcricional XlnR. O fator de transcrição da cutinase 1 α (CTF1 α) em *Hypocrea jecorina*, o qual é ortólogo à proteína FarA em *A. nidulans*, participa na ativação de genes envolvidos na degradação da cutina, composto lipídico da parede celular vegetal que constitui um polímero de muitos ácidos graxos de cadeias longas, presente no limbo foliar (Hynes et al., 2006). Os demais fatores de transcrição estão anotados como proteínas hipotéticas, portanto não possuem ortólogos funcionais depositados em banco de dados.

2. Análise da expressão dos genes *gsn* e *gpn* por Northern blot nas linhagens mutantes

Após a seleção das 10 linhagens mutantes que apresentaram alterações em relação ao acúmulo de glicogênio antes e/ou depois do choque térmico, foram realizadas análises da expressão dos genes *gsn* e *gpn*, antes e após choque térmico, com o objetivo de verificar se tais fatores de transcrição controlam o metabolismo do glicogênio por meio do controle da transcrição destes genes. O RNA total foi extraído a partir de micélios congelados das linhagens selecionadas e utilizados para análise, utilizando o RNA da linhagem selvagem FGSC# 9718 como referência. As análises foram realizadas pela técnica de *Northern blot* utilizando o RNA ribossômico 18S como controle endógeno. Pelo menos três experimentos independentes foram realizados para a confirmação dos resultados. Na figura 6B, podemos

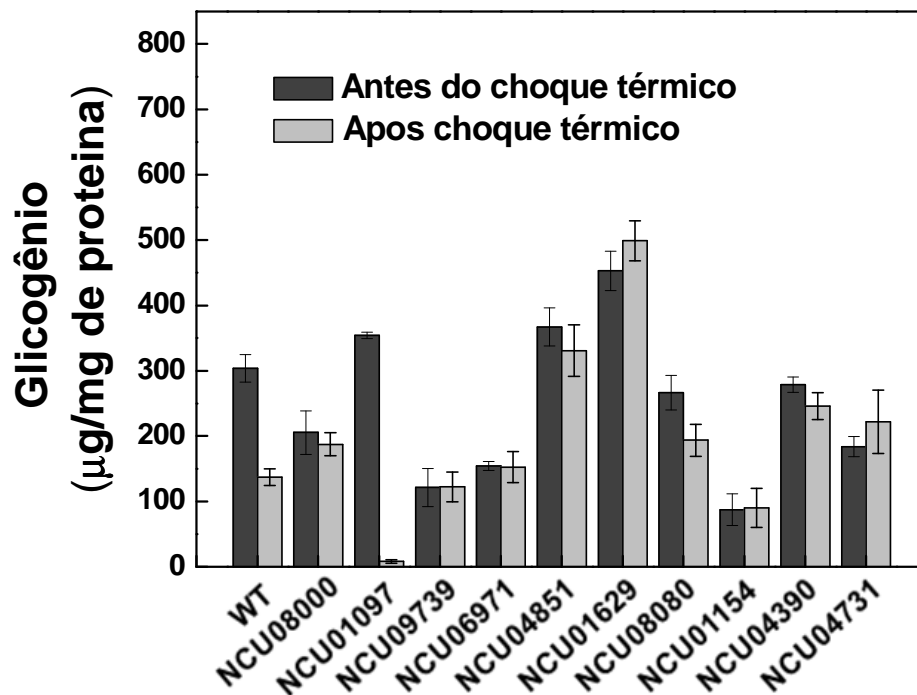
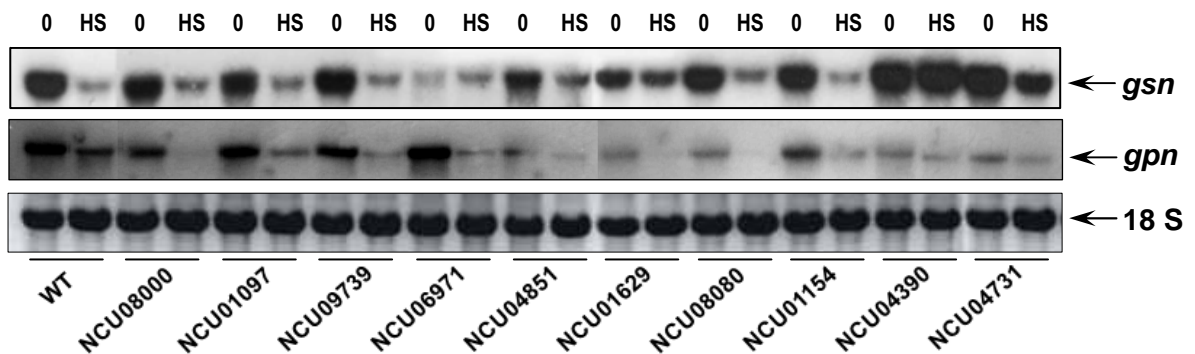
A**B**

Figura 6 - (A) Análise do conteúdo de glicogênio nas linhagens selvagem (WT) e mutantes selecionadas antes e após o choque térmico. As amostras foram crescidas em VM a 30°C por 24 horas e transferidas para meio VM previamente aquecido a 45°C e mantidas por 30 minutos nesta condição. O conteúdo de glicogênio foi quantificado como descrito em Materiais e Métodos. **(B)** Expressão dos genes *gsn* e *gpn* nas linhagens selvagem (WT) e mutantes selecionadas antes e após o choque térmico por *Northern blot*. Perfil do RNAr 18 S visualizado com brometo de etídio. Canaletas 0 e HS representam o RNA total extraído das linhagens do fungo antes e após serem submetidas ao choque térmico (45°C por 30 minutos), respectivamente. As amostras de RNA foram separadas por eletroforese em gel de ágarose contendo formaldeído, transferidas para uma membrana de nylon neutra e hibridizadas com fragmento dos genes *gsn* e *gpn*, marcados radioativamente com $\alpha^{32}\text{P}$ -dATP, e utilizados como sondas.

observar a presença de uma única banda de hibridização com 3,0 kb, a qual corresponde aos tamanhos dos transcritos *gsn* e *gpn*. É importante ressaltar que os níveis do RNA ribossomal 18S não apresentaram alterações tanto antes quanto depois da situação de estresse térmico. Os resultados das expressões de *gsn* e *gpn*, antes e após choque térmico, mostraram que a ausência de determinados fatores de transcrição, promove uma modulação na expressão de *gsn* e *gpn* antes e/ou após choque térmico. Este fato sugere que alguns dos fatores de transcrição analisados estão envolvidos no controle do metabolismo do glicogênio por meio do controle transcricional do gene *gsn* e/ou *gpn*. Por outro lado, apesar da ausência de alguma das proteínas regulatórias promoverem alteração do acúmulo de glicogênio, alguns mutantes não apresentaram alterações na transcrição tanto de *gsn*, quanto de *gpn*. Pelo fato da enzima GSN ser a enzima regulatória do processo de síntese de glicogênio, mas não a única envolvida no processo, possivelmente os demais fatores de transcrição estejam envolvidos no controle do metabolismo de glicogênio, atuando possivelmente na regulação de outras proteínas envolvidas nesta via metabólica.

Pela figura 6B, podemos observar que o nível transcricional do gene *gsn*, antes e depois do choque térmico, nas 10 linhagens selecionadas, é muito semelhante ao apresentado pela linhagem selvagem quando submetida às mesmas condições, exceto para as linhagens FGSC#11067, FGSC#11102, FGSC#11134 e FGSC#11139, cujas ORFs nocauteadas são NCU06971, NCU01629, NCU04390 e NCU04731, respectivamente. A partir desta figura fica evidente que os demais fatores de transcrição selecionados estão envolvidos na regulação do metabolismo, mas não no controle transcricional dos genes *gsn*. A linhagem contendo a ORF NCU06971 nocauteada, ao contrário da linhagem selvagem, apresentou uma redução drástica do transcrito *gsn* quando crescida a 30°C e, quando submetida à situação de estresse térmico os níveis transcripcionais de *gsn* permaneceram baixos, como apresentado pela linhagem selvagem sob a mesma condição. A linhagem contendo a ORF NCU01629 nocauteada apresentou uma leve redução na transcrição de *gsn* quando crescida a 30°C e um aumento dos níveis de transcrito após a condição de estresse térmico quando comparada com a linhagem selvagem. A linhagem contendo a ORF NCU04390 nocauteada não apresentou alterações da transcrição de *gsn* em relação à linhagem selvagem quando crescida a 30°C, mas quando submetida à condição de estresse térmico, os níveis transcripcionais de *gsn* permaneceram muito elevados, não apresentando redução como apresentado pela linhagem selvagem. Finalmente, a linhagem contendo a ORF NCU04731 nocauteada, assim como a linhagem selvagem, não apresentou alterações dos níveis do transcrito *gsn* quando crescida a 30°C, mas quando submetida à condição de estresse térmico houve uma diminuição dos níveis transcripcionais de *gsn*, mas não tão pronunciado como observado na linhagem selvagem.

Em relação à expressão do gene *gpn*, podemos verificar que de uma forma geral todas as linhagens apresentaram o mesmo perfil da linhagem selvagem, em ambas as condições, ou seja, uma redução no nível do transcrito *gpn* após o choque térmico em todas as linhagens mutantes. Entretanto, os níveis do transcrito apresentaram alterações entre as diferentes linhagens mutantes. Por exemplo, as linhagens contendo as ORFs NCU01097 e NCU09739 nocauteadas apresentaram níveis do transcrito semelhantes ao apresentado pela linhagem selvagem, em ambas as condições. As demais linhagens mutantes analisadas apresentaram perfis da expressão de *gpn* totalmente diferentes do apresentado pela linhagem selvagem. É interessante destacar os níveis de expressão dos dois genes nas linhagens mutantes nas ORFs NCU04390 e NCU04731, as quais apresentaram níveis exacerbados para o gene *gsn* e níveis altamente reduzidos para o gene *gpn*, tanto na temperatura normal de crescimento quanto no choque térmico.

De acordo com as análises do conteúdo de glicogênio antes e após a condição de choque térmico junto das análises da transcrição de *gsn* e *gpn* nas linhagens mutantes sob as mesmas condições, as linhagens mutantes nas ORFs NCU06971 e NCU04390 se mostraram muito promissoras no estudo do controle da via metabólica do glicogênio, pois na ausência destas proteínas ocorre alteração tanto do perfil de acúmulo do glicogênio, bem como no padrão de expressão dos genes *gsn* e *gpn*, sugerindo que estas proteínas regulatórias possivelmente apresentem funções de grande importância para a manutenção do controle da via metabólica do glicogênio. Sendo assim, estas proteínas foram escolhidas para dar continuidade nos estudos de caracterização funcional de fatores de transcrição envolvidos na regulação da via metabólica do glicogênio.

3. Análises das linhagens mutantes por citometria de fluxo

A citometria de fluxo é uma técnica de análise automatizada, extremamente sensível, através da qual uma única partícula pode ser caracterizada física e bioquimicamente em meio líquido. Esta caracterização inclui parâmetros quantitativos, como tamanho, granulosidade (ou complexidade interna) e intensidade da fluorescência, os quais podem ser detectados simultaneamente e mensurados, proporcionando análises rápidas e multiparamétricas (GIVAN, 2004). Esta técnica foi utilizada para as análises referentes à morfologia conidial, bem como análise do ciclo celular dos conídios provenientes das 10 linhagens mutantes inicialmente selecionadas. Em relação à morfologia e ciclo celular podemos notar que das linhagens analisadas, apenas a linhagem com a ORF NCU09739 nocauteada apresenta um perfil distinto da linhagem selvagem, tanto em relação à morfologia conidial (figura 7A) quanto em relação ao ciclo celular (figura 7B) quando comparada com a linhagem selvagem. Podemos notar que os conídios desta linhagem

mutante são menores e apresentam menor granulosidade em relação aos da linhagem selvagem. Na figura 7A, as áreas em azul representam regiões com grande quantidade de conídios e as áreas em verde representam regiões com uma menor densidade celular, uma vez que não é possível localizar todos os conídios analisadas no *dot-plot*. Além disso, percebemos pela figura 7B que o ciclo celular desta linhagem mutante foi deslocado para a esquerda, região também conhecida como "sub-G1", sugerindo que este fator de transcrição esteja envolvido também no controle do ciclo celular.

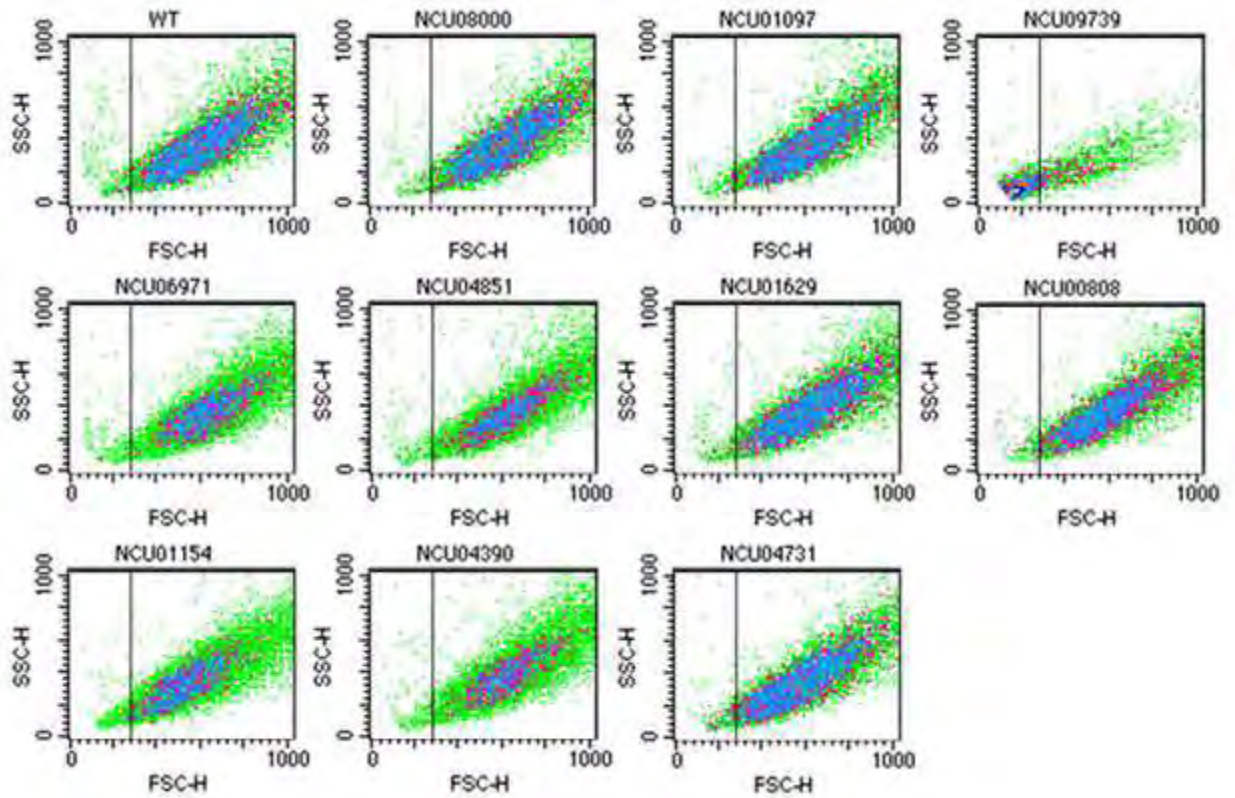
4. Produção e purificação de proteína recombinante

Análises conformacionais teóricas realizadas no servidor ELM (<http://elm.eu.org/>) mostraram que os dois fatores de transcrição escolhidos para a realização de análises funcionais, os fatores de transcrição codificados pelas ORFs NCU06971 e NCU04390, apresentam dois domínios conformacionais conhecidos e localizados nas mesmas regiões: o domínio *zinc finger* localizado na região N-terminal e um domínio específico de fatores de transcrição de fungos na região central das proteínas. Pelo fato de não haver na literatura informações relevantes em relação ao domínio específico de fungos, estas proteínas foram produzidas forma inteira e truncadas nas regiões central e C-terminal na forma recombinante em células de *E. coli*, com o intuito de analisar a importância do domínio específico de fungos em relação à interação DNA-proteína. As formas truncadas foram as seguintes: 4390 Δ 158 (aminoácidos 1-158, contém apenas o domínio de ligação ao DNA), 4390 Δ 483 (aminoácidos 1-483, contém os dois domínios, o de ligação ao DNA e o específico de fungos) e 6971 Δ 160 (aminoácidos 1-160 contém apenas o domínio de ligação ao DNA). Uma representação esquemática dos dois fatores de transcrição estudados neste trabalho, tanto na forma inteira quanto nas formas truncadas, está representada na figura 8.

4.1 Ensaios de indução e análise de solubilidade das proteínas recombinantes

Para a produção das proteínas recombinantes e verificação de suas solubilidades, as construções pET-04390 Δ 158, pET-04390 Δ 483, pET-04390, pET-06971 Δ 160 e pET-06971 foram utilizados para transformar células competentes das linhagens de *E. coli* produtoras Rosetta (DE)pLys-S e ArticExpressTM (DE3). Inicialmente foram realizados ensaios de indução em pequena escala (50 mL de cultura) com o objetivo de determinar as melhores condições de expressão. A confirmação da produção foi realizada por análises de *Western blot*, utilizando anticorpo monoclonal anti-His. Bandas correspondentes às proteínas His4390, His4390 Δ 483, His4390 Δ 158, His6971 e His6971 Δ 160 foram identificadas nos

A



B

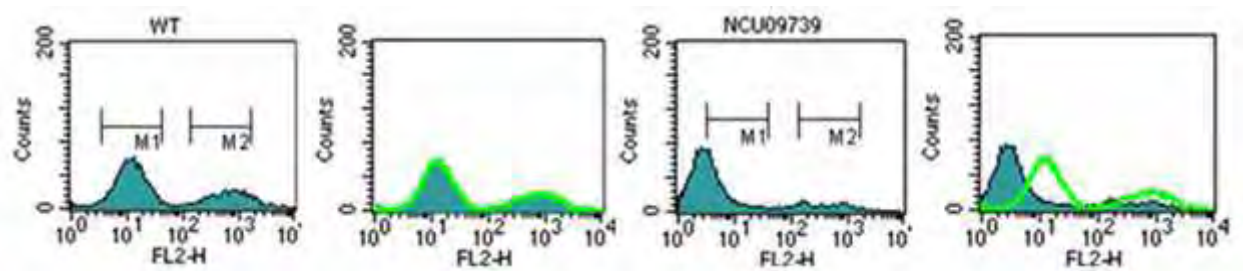


Figura 7 - Análise da morfologia dos conídios das linhagens mutantes de *N. crassa* por citometria de fluxo. **(A)** Gráfico de dispersão de pontos de acordo com tamanho (FSC) em relação à granulosidade (SSC) dos conídios. **(B)** Análise do ciclo celular dos conídios da linhagem mutante na ORF NCU09739 comparada à linhagem selvagem.

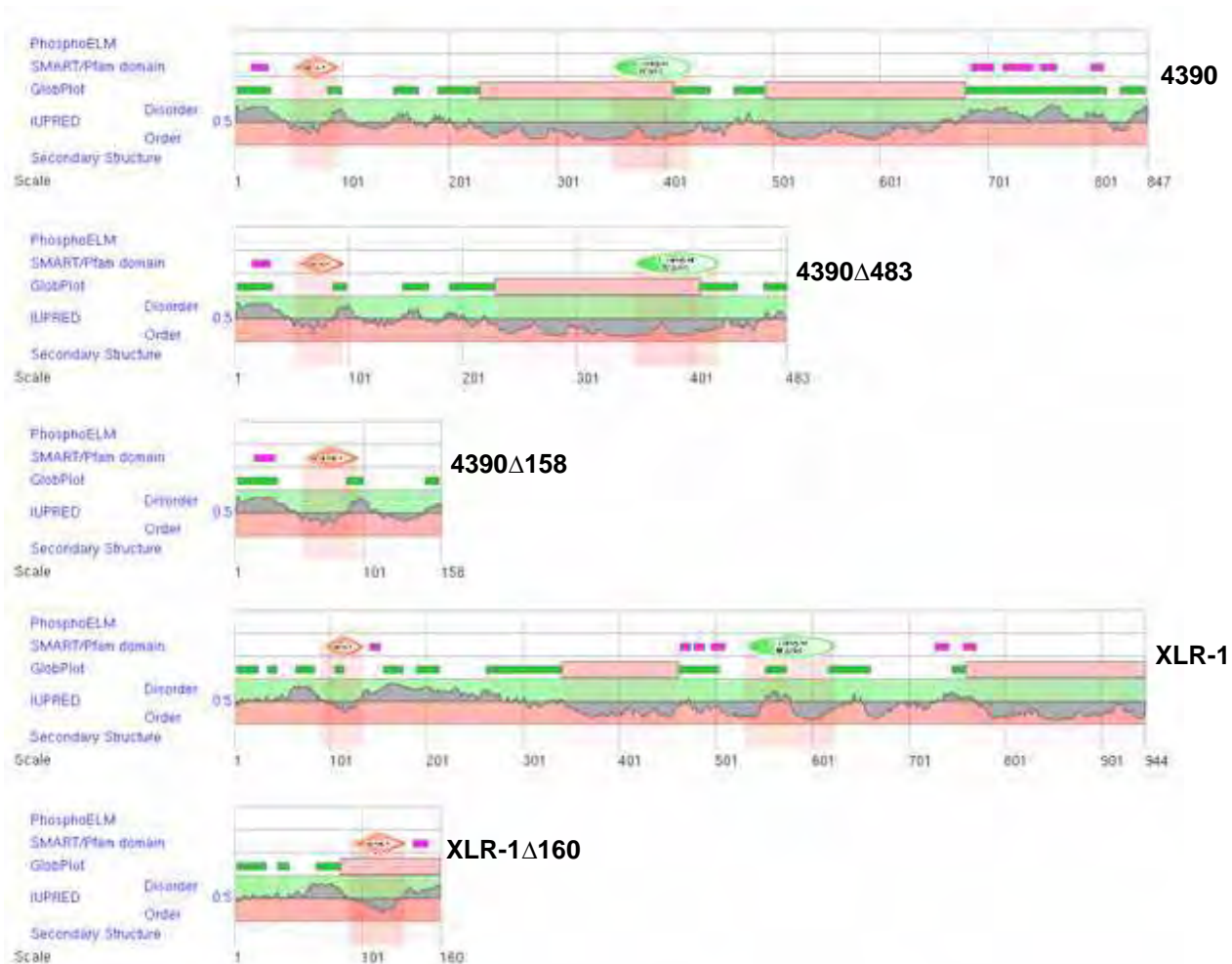


Figura 8 - Representação esquemática dos fatores de transcrição codificados pelas ORFs NCU06971 e NCU04390 inteiros e truncados utilizados neste trabalho, utilizando o servidor ELM (<http://elm.eu.org/>). O losango laranja representa o domínio de ligação ao DNA Zn₂Cys₆, a elipse verde representa o domínio específico de fungos.

extratos celulares dos respectivos clones produtores, induzidos por IPTG, fracionados em SDS-PAGE (figura 9A e 10A) e pelos resultados das análises de *Western blot* (figura 9B e 10B). Os resultados revelaram que a proteína 4390 inteira (100 kDa), apesar de apresentar bandas discretas nos géis de acrilamida, foi expressa em quantidade razoável e está presente tanto na fração solúvel (S), em menor quantidade, quanto no precipitado (P), em maior quantidade. A proteína 4390 Δ 483 apresentou bandas mais evidentes nas análises os géis de acrilamida (57 kDa) e a análise por *Western blot* revelou que apesar desta proteína estar presente nas frações solúvel (S) e insolúvel (P), sua quantidade na fração solúvel foi muito menor do que a quantidade presente no precipitado.

A proteína 4390 Δ 158, apesar de apresentar bandas discretas nas análises dos géis de acrilamida (25 kDa), foi expressa em uma quantidade razoável, estando presente na fração solúvel (S) quanto no precipitado (P), em quantidades equivalentes, após sua separação por centrifugação. Em relação à expressão da proteína 6971, foi verificado pela técnica de *Western blot* uma banda do tamanho esperado correspondente à sua massa molecular teórica (107 kDa), uma vez que esta proteína foi expressa em baixa quantidade tanto no sobrenadante (S) quanto no precipitado (P). Finalmente, foi detectado a presença da proteína 6971 Δ 160 no gel (21 kDa) em quantidade razoável na fração solúvel (S) e em maior quantidade na fração insolúvel (P). Estas análises foram confirmadas por *Western blot*.

4.2. Produção e purificação de proteínas recombinantes

Após a escolha de um clone produtor de cada transformação bem como a determinação das melhores condições para a expressão das proteínas na forma recombinante, foi efetuada a indução dos clones produtores das cinco proteínas para a produção das respectivas proteínas em maior escala (1 L), com o objetivo de se obter uma grande quantidade das proteínas de interesse para análises de interação DNA-proteína *in vitro* pela técnica de EMSA. A purificação das proteínas recombinantes foi realizada por cromatografia de afinidade utilizando colunas His-Trap (GE) em um sistema de purificação Äkta Prime (GE). As figuras 11-15 mostram as etapas de purificação das proteínas His4390 (figura 11), His4390 Δ 483 (figura 12), His4390 Δ 158 (figura 13), His6971 Δ 160 (figura 14) e His6971 (figura 15). O perfil de eluição destas proteínas está mostrado nas figuras A, e em B as análises das frações cromatográficas por SDS-PAGE. Foram eluídas 36 frações proteicas, e as frações pares foram fracionadas por SDS-PAGE utilizando gel de poli-acrilamida 12%. As frações onde as proteínas de interesse se encontraram menos contaminadas foram unidas e dialisadas e congeladas a -80°C para serem utilizadas posteriormente nas análises de interação DNA-proteína por EMSA.

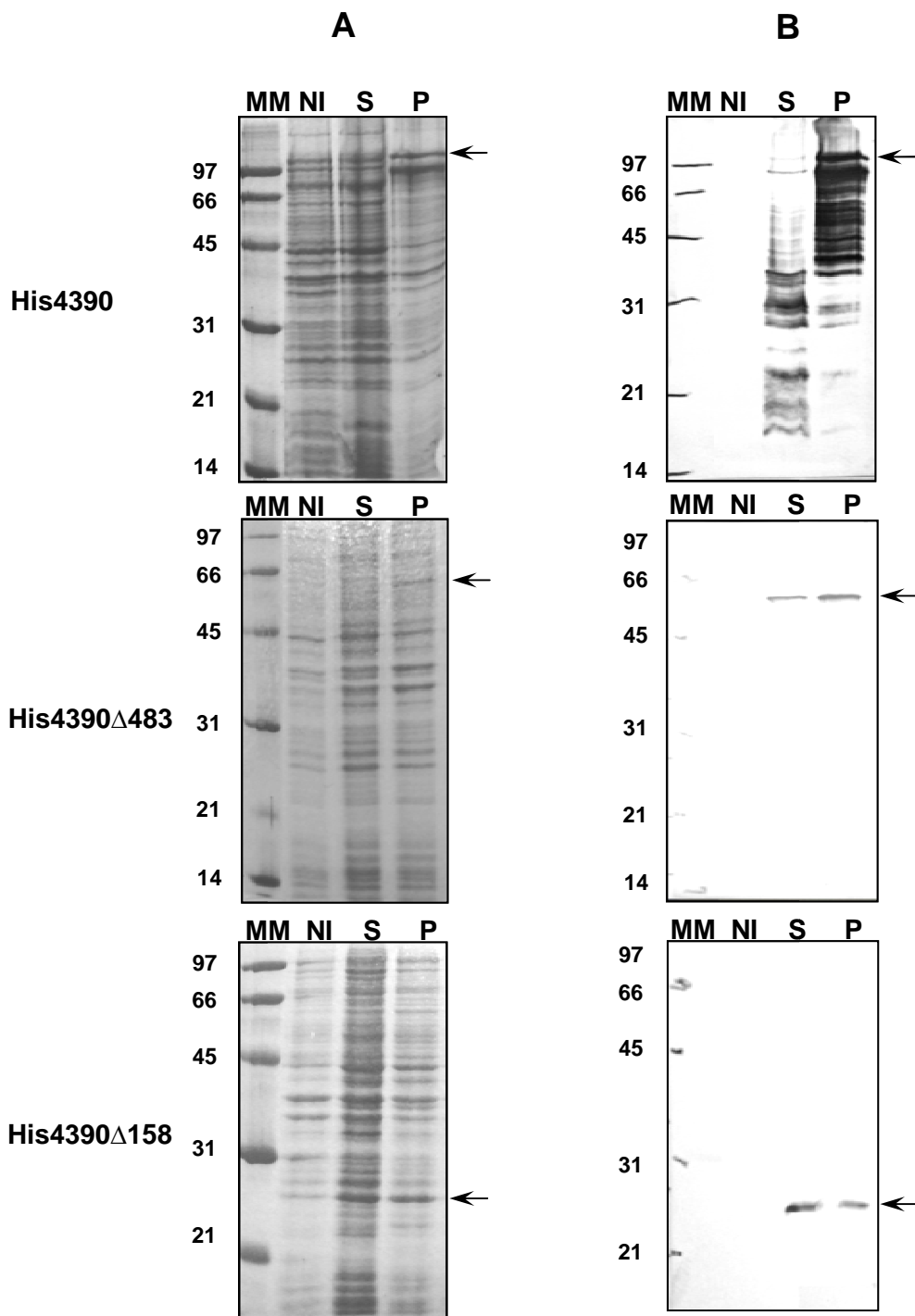


Figura 9 - Produção das proteínas His4390 Δ 158, His4390 Δ 483 e His4390 na forma recombinante em *E. coli*. **(A)** Extratos celulares brutos obtidos a partir da indução de culturas de *E. coli* transformantes, contendo as construções pET4390 Δ 158, pET4390 Δ 483, pET4390 foram analisados por SDS-PAGE no sistema Mini-Protean III (Bio-Rad) e analisados após coloração com *Coomassie-Blue Brilliant R-250* (Sigma). **(B)** Confirmação da produção das proteínas recombinantes por *Western blot* utilizando anticorpo monoclonal de coelho anti-His conjugado com fosfatase alcalina (Sigma). As setas indicam as proteínas recombinantes. **MM**, massa molecular (kDa); **NI**, extrato proteico de células não induzidas; **S**, sobrenadante; **P**, precipitado. Os números à esquerda correspondem às massas moleculares em kDa.

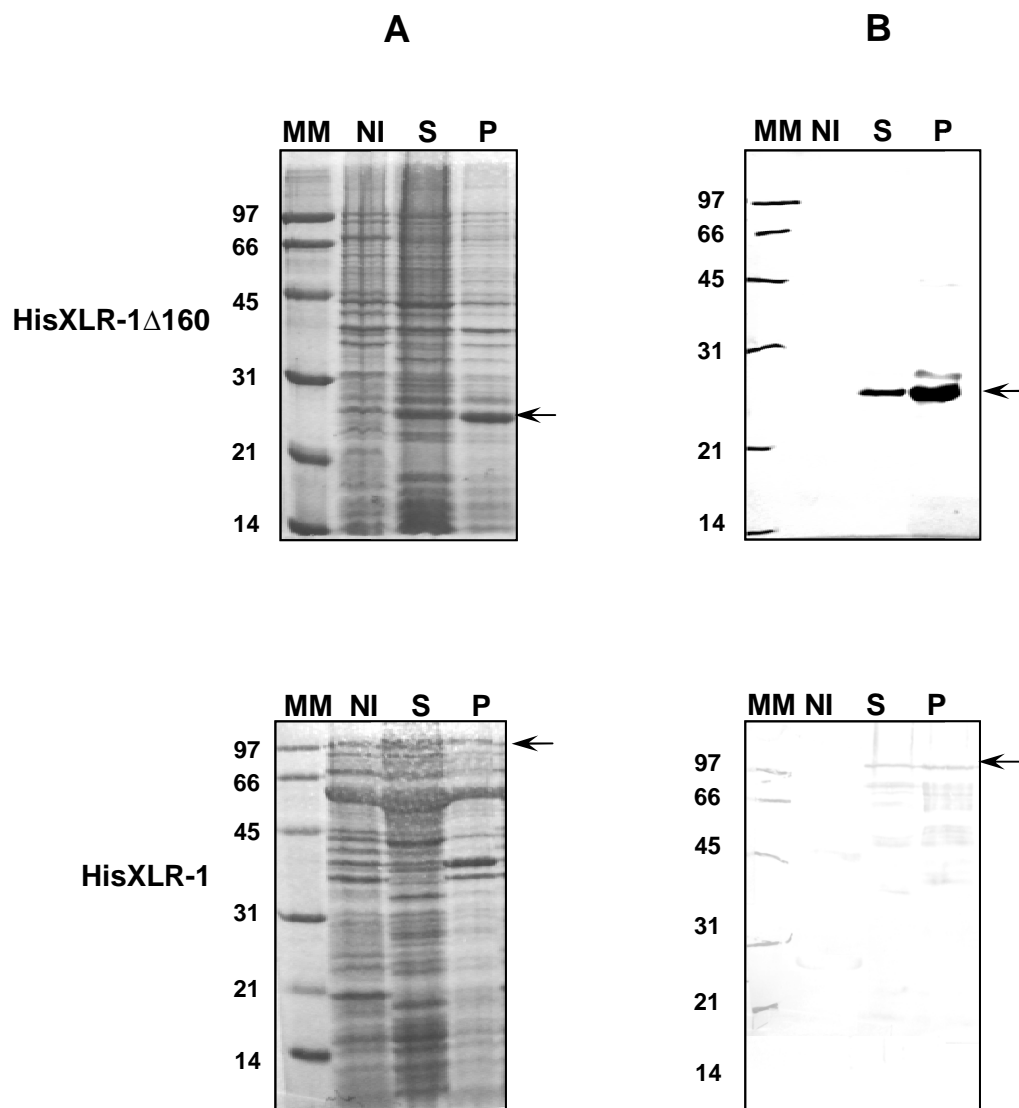


Figura 10 - Produção das proteínas HisXLR-1 Δ 160 e HisXLR-1 na forma recombinante em *E. coli*. **(A)** Extratos celulares brutos obtidos a partir da indução de culturas de *E. coli* transformantes, contendo as construções pET6971 Δ 160 e pET6971 foram analisados por SDS-PAGE no sistema Mini-Protean III (Bio-Rad) e analisados após coloração com *Coomassie-Blue Brilliant R-250* (Sigma). **(B)** Confirmação da produção das proteínas recombinantes por *Western blot* utilizando anticorpo monoclonal de coelho anti-His conjugado com fosfatase alcalina (Sigma). As setas indicam as proteínas recombinantes. **MM**, massa molecular (kDa); **NI**, extrato proteico de células não induzidas; produzida em alta quantidade a partir da fração solúvel.

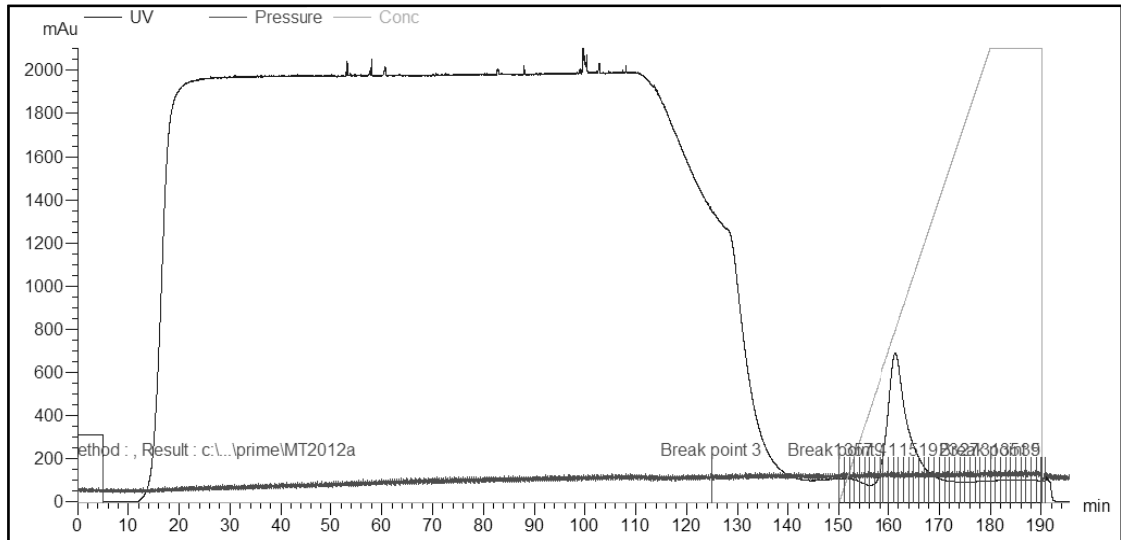
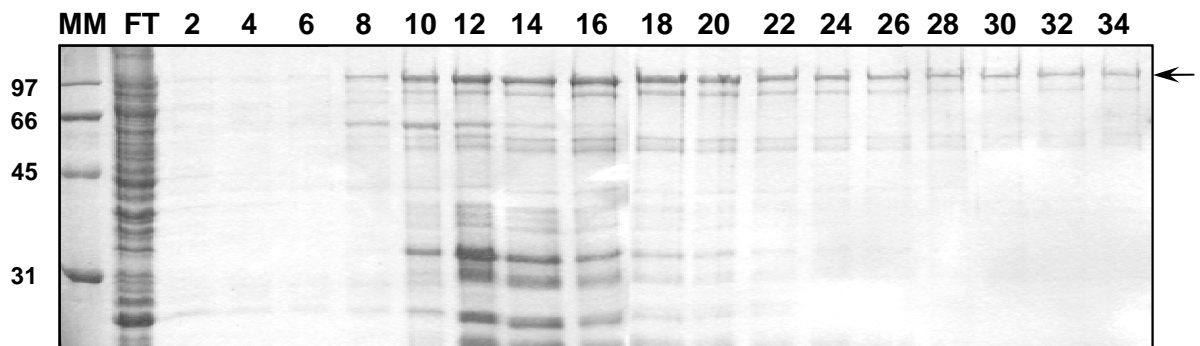
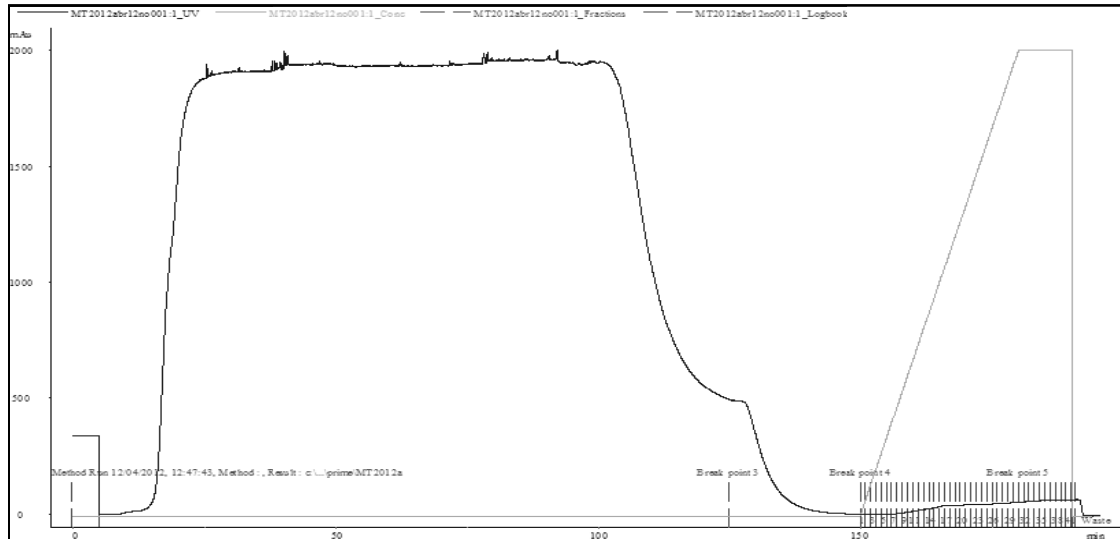
A**B**

Figura 11 – Purificação e análise da proteína His4390 produzida em *E. coli*, linhagem Rosetta, em larga escala. **(A)** Cromatograma das etapas de purificação da proteína em coluna HisTrap (GE) utilizando um gradiente linear de 50 mM a 750 mM de Imidazol. **(B)** Análise eletroforética das frações pares eluídas. MM, padrão de massa molecular, FT, *flow through*. Eletroforese realizada em gel de poliacrilamida 12% corado com Coomassie Brilliant Blue R-250. A seta mostra a proteína de interesse.

A



B

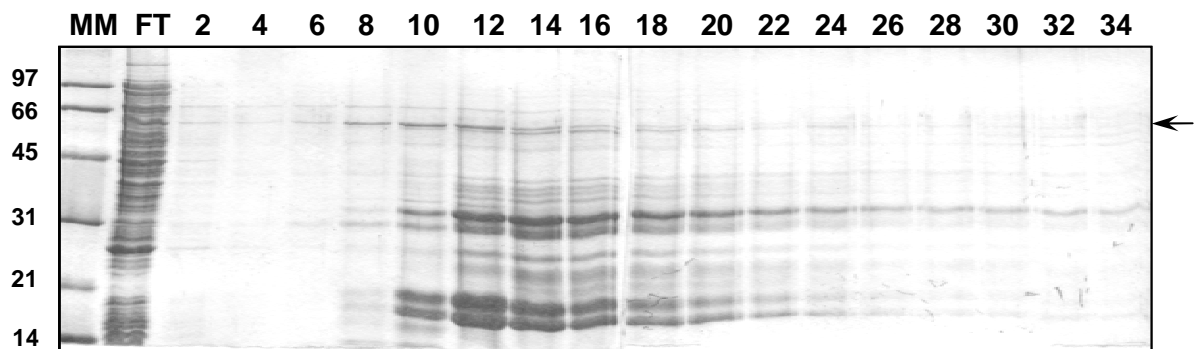
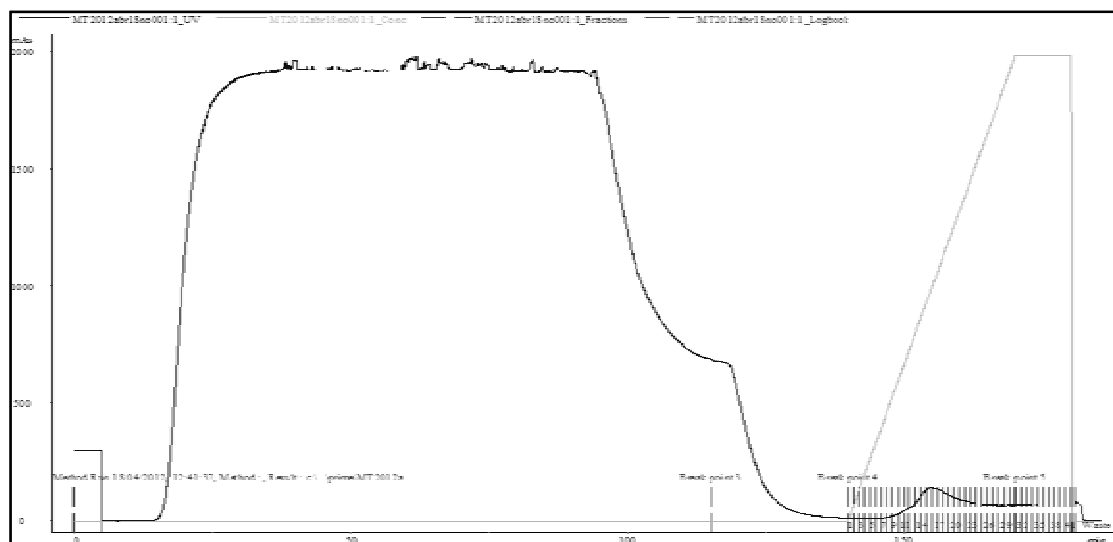


Figura 12 – Purificação e análise da proteína His4390 Δ 483 produzida em *E. coli*, linhagem Rosetta, em larga escala. **(A)** Cromatograma das etapas de purificação da proteína em coluna HisTrap (GE) utilizando um gradiente linear de 50 mM a 750 mM de Imidazol. **(B)** Análise eletroforética das frações pares eluídas. MM, padrão de massa molecular, FT, *flow through*. Eletroforese realizada em gel de poliacrilamida 12% corado com Coomassie Brilliant Blue R-250. A seta mostra a proteína de interesse.

A



B

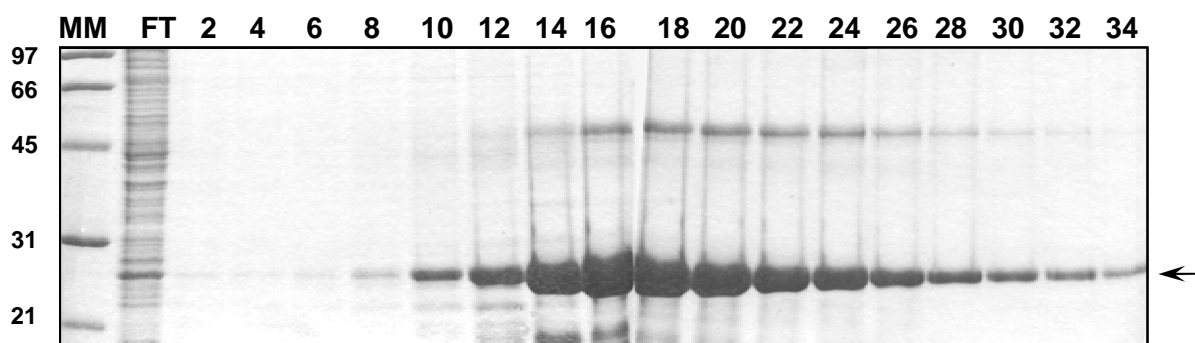
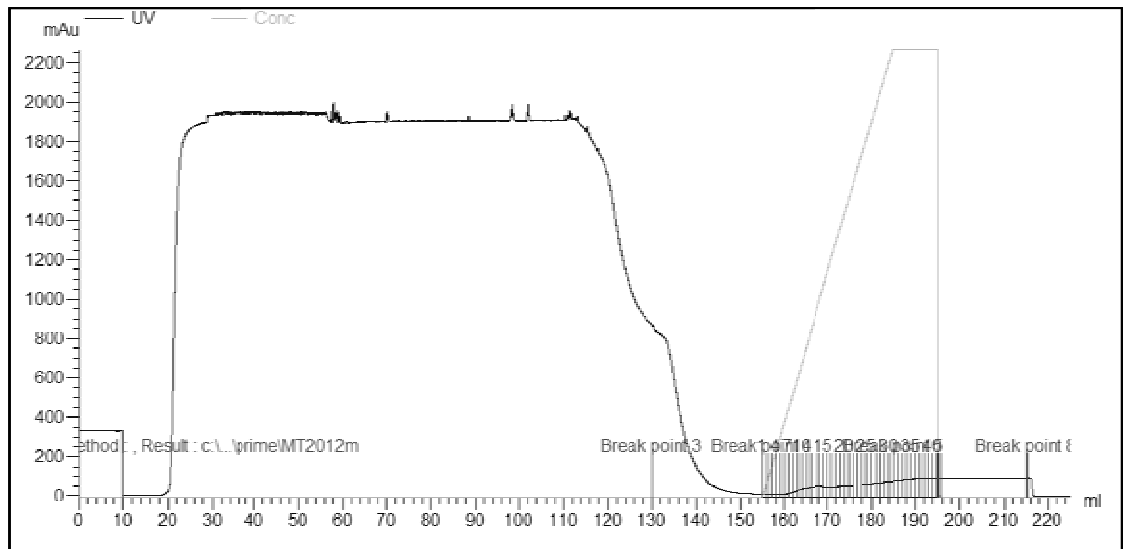


Figura 13 – Purificação e análise da proteína His4390 Δ 156 produzida em *E. coli*, linhagem Rosetta, em larga escala. **(A)** Cromatograma das etapas de purificação da proteína em coluna HisTrap (GE) utilizando um gradiente linear de 50 mM a 750 mM de Imidazol. **(B)** Análise eletroforética das frações pares eluídas. MM, padrão de massa molecular, FT, *flow through*. Eletroforese realizada em gel de poliacrilamida 12% corado com Coomassie Brilliant Blue R-250. A seta mostra a proteína de interesse.

A



B

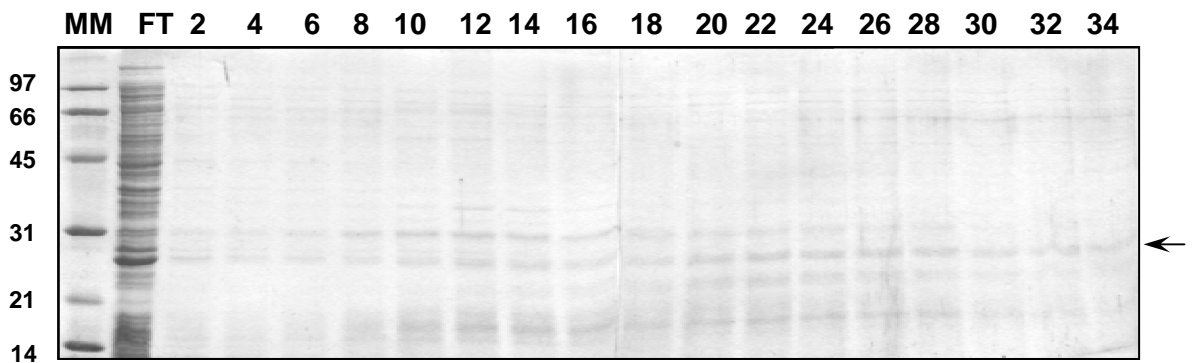
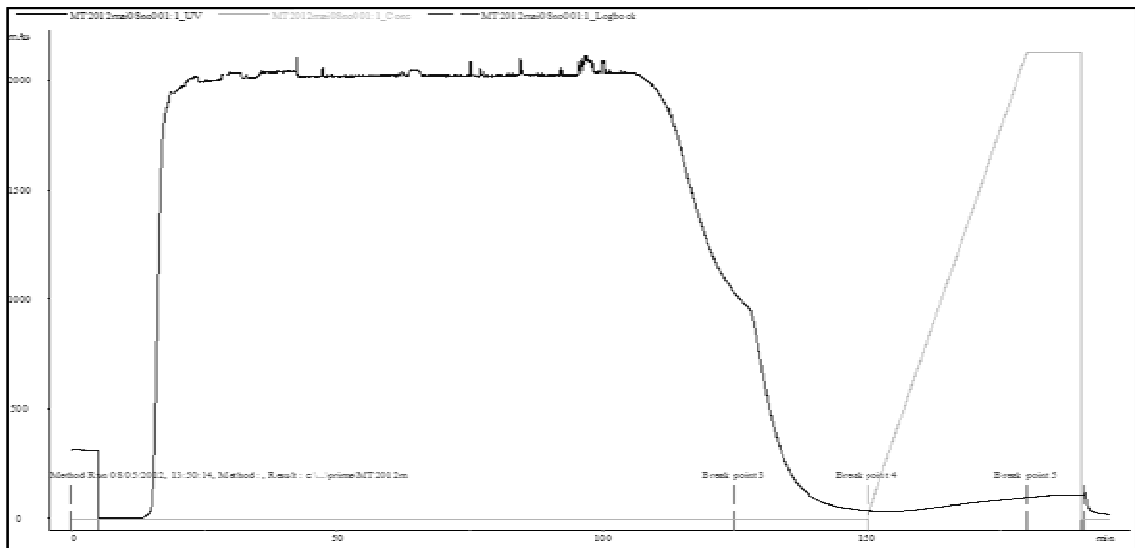


Figura 14 – Purificação e análise da proteína HisXLR-1 Δ 160 produzida em *E. coli* em larga escala. (A) Cromatograma das etapas de purificação da proteína em coluna HisTrap (GE) utilizando um gradiente linear de 50 mM a 750 mM de Imidazol. (B) Análise eletroforética das frações pares eluídas. MM, padrão de massa molecular, FT, *flow through*. Eletroforese realizada em gel de poliacrilamida 12% corado com Coomassie Brilliant Blue R-250. A seta mostra a proteína de interesse.

A



B

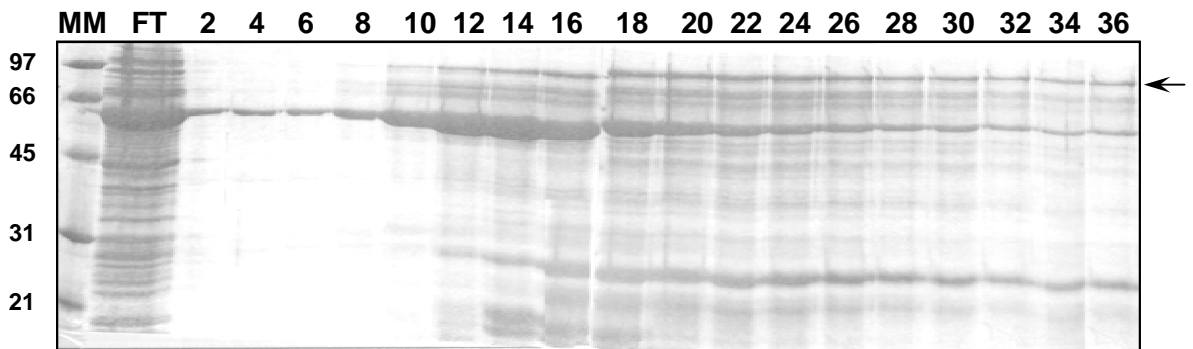


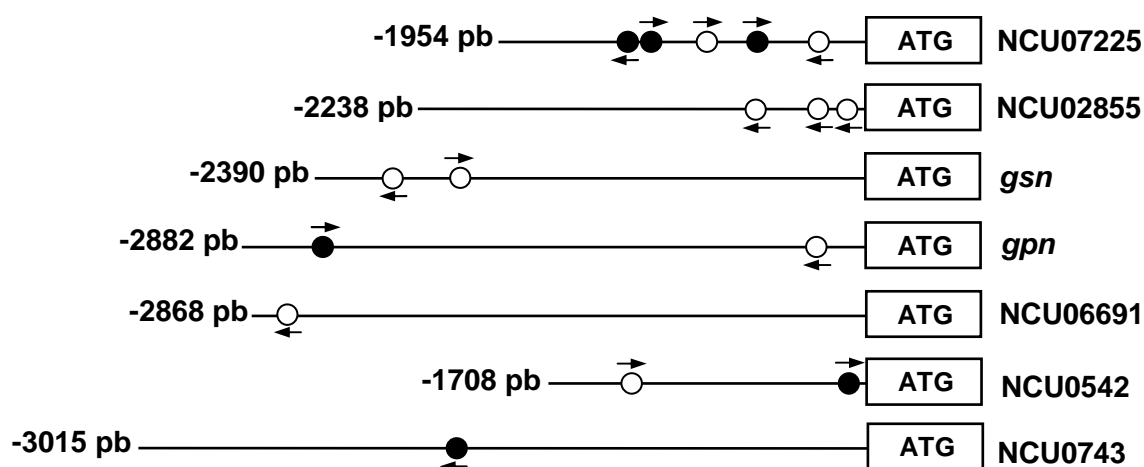
Figura 15 – Purificação e análise da proteína HisXLR-1 produzida em *E. coli* em larga escala. **(A)** Cromatograma das etapas de purificação da proteína em coluna HisTrap (GE) utilizando um gradiente linear de 50 mM a 750 mM de Imidazol. **(B)** Análise eletroforética das frações pares eluídas. MM, padrão de massa molecular, FT, *flow through*. Eletroforese realizada em gel de poliácridamida 12% corado com Coomassie Brilliant Blue R-250. A seta mostra a proteína de interesse.

5. Caracterização funcional da proteína XLR-1 no fungo *N. crassa*

A caracterização funcional do fator de transcrição XLR-1 no fungo *N. crassa* foi realizada como uma tentativa de correlacionar sua função na regulação do metabolismo de glicogênio. Inicialmente, a linhagem mutante foi confirmada pela ausência de amplificação da ORF por PCR e uma reação paralela com a linhagem selvagem. Foram realizadas buscas por elementos de DNA regulatórios para este fator de transcrição (5'-GGCTRA-3') nas regiões 5'-flanqueadoras dos genes *gsn*, *gpn* e dos genes codificadores das enzimas glicogenina e enzima ramificadora, ambas envolvidas na síntese do carboidrato, e enzima desramificadora do glicogênio, envolvida na degradação do carboidrato, o que revelou a existência de elementos de DNA regulatórios nas respectivas regiões (figura 16). Na figura 16A estão representados as localizações físicas dos prováveis *motifs* para a proteína e na figura 16B a localização exata na região. Foram encontrados dois possíveis elementos de DNA regulatório localizados na região promotora do gene *gsn* nas posições -2071 e -2252, um possível elemento localizado na região 5'-flanqueadora do gene *gpn* localizado na posição -217, um possível elemento localizado na região 5'-flanqueadora de *gnn* localizado na posição -2832, dois possíveis elementos regulatórios localizados na região 5'-flanqueadora do gene NCU05429 localizados a -140 e -1374 e um possível elemento regulatório na região 5'-flanqueadora do gene NCU00743 localizado na posição -2043. As posições são referentes ao *codon* iniciador ATG de cada gene.

Também foram realizados análises no banco de dados de *N. crassa* uma busca por proteínas ortólogas as duas endoxilanasas de *Aspergillus nidulans* ativadas pelo fator de transcrição XlnR, com o intuito de verificar se em *N. crassa* este fator de transcrição possui um papel regulatório na expressão de genes codificadores de enzimas xilanolíticas, como mostrado em outros fungos filamentosos. Foram utilizadas as sequências polipeptídicas das endoxilanasas xlnA (AN3613) e xlnB (AN9365) como *query* em análises de BlastP no banco do genoma do fungo *N. crassa*. Um dado curioso é que independente da sequência utilizada como *query* (xlnA e xlnB), foram obtidas apenas duas proteínas ortólogas em *N. crassa*, as quais apresentaram *e-value* igual a zero e identidade acima de 60% em relação às proteínas utilizadas como *query*, uma delas é codificada pela ORF NCU02885, a qual está anotada como endo-1,4- β -xylanase A e a outra é codificada pela ORF NCU07225, a qual está anotada como endo-1,4- β -xylanase 2. O alinhamento múltiplo das sequências das proteínas codificadas pelas ORFs AN3613, AN9365, NCU07225 e NCU02855 está mostrado na figura 17. Como mostrado no alinhamento, podemos verificar que as proteínas analisadas possuem um elevado grau de identidade entre si. A única peculiaridade é a presença de uma inserção na região C-terminal da proteína codificada pela ORF NCU07225. Esta região apresenta um domínio de ligação à celulose localizada entre os resíduos 204 e 232, o qual

A



B

ORF	Anotação da proteína	Posição (em nt)*
NCU07225	endo-1,4-beta-xilanase 2	-271, -557, -870, -1027, -1037
NCU02855	endo-1,4-beta-xilanase A	-37, -51, -474
<i>gsn</i>	glicogênio sintase	-2071, -2252
<i>gpn</i>	glicogênio fosforilase	-217, -2631
NCU06698	glicogenina	-2832
NCU05429	enzima ramificadora	-140, -1374
NCU0743	Enzima desramificadora do glicogênio	-2043

*em relação ao *start codon*

Figura 16 – Possíveis elementos de DNA regulatórios *cis* para a proteína XLR-1 na região 5'-flanqueadora de genes codificadores de enzimas do metabolismo de glicogênio e endoxilanases do fungo *N. crassa*. **(A)** Representação esquemática dos elementos regulatórios *cis* localizados nas regiões 5'-flanqueadoras dos genes. As sequências 5'-GGCTAA-3' estão representadas por círculos preenchidos e as sequências 5'-GGCTGA-3' por círculos vazios. **(B)** Localização exata dos possíveis elementos de DNA regulatórios para o fator de transcrição XLR-1 encontrados na região 5'-flanqueadora de genes do fungo *N. crassa*.

tem a função de ancoragem da enzima ao seu substrato.

Oligonucleotídeos específicos para estas ORFs foram desenhados e utilizados para a obtenção das sequências codificadoras destas ORFs via PCR, com o intuito de utilizá-las como sondas em análises de expressão gênica por *Northern blot*. As sequências dos oligonucleotídeos utilizados para a amplificação das sondas acima citadas estão mostradas em Materiais e Métodos. Após a identificação das proteínas ortólogas em *N. crassa* (ORFs NCU07225 e NCU02855) foram também realizadas buscas por elemento de DNA regulatório para a proteína XLR-1 nas regiões 5'-flanqueadoras destas ORFs, revelando a existência de tais elementos localizados nas regiões -271, -557, -870, -1027, -1037 para a ORF NCU07225 e -37, -51, -474 para a ORF NCU02855, como mostrado na figura 16.

```

AN3613      1  GDVITYTNGAGGSYTVQWSNVGNFVGGKGNPNPGRTRTINYGGSFNPSGNGYLAVYGWTQNP
AN9365      1  GDVITYTNGDGGSYTVEWLNVGNFVGGKGNPNPSSQTISYSGSFNPSGNGYLSVYGWTQNP
NCU02855    1  GNVINYANGASGSYSVNWQNAAGNFVAGKGNPNPGRARTITISGNFRPSGNGYLSVYGWTRNP
NCU07225    1  GSVRYTNEAGGOYTATWSGNGNVGGKGNMPTDRTINYITGTYSPNGNSYLAVYGWTRNP

AN3613     61  LIEYYIVESYGTYNPSSGCOHRGTVYSDGATYDIYTATRYNAPSIEGTATFEQFWSVRQS
AN9365     61  LIEYYIVESYGDYNPGTAGTHQGLESDGSTYDIYTATRENAPSIEGTATFTQFWSVRQS
NCU02855   61  LVEYYVVENFGSYNPSGSAQRLGVSVDGTYDIYKTTRYNQPSIDGTRTFNQYWSVRQQ
NCU07225   61  LIEYYVVENFGTYNPSIGATRLGVSITSDGGVYDIYRTRQVFNQPSIDGTATFYQYWSVRQQ

AN3613    121  KRTGGTIVTMANHFNAWAALGMRLGTHNYQIVATEGYQSSGSASITVY-----
AN9365    121  KRTSGSVTTONHFDAWSQLGMVTLGTHNYQIVAVEGYQSSGSASITVS-----
NCU02855  121  KRTGGTIVTMANHFNAWAKAGLHLGTHNYQIVATEGYQSSGSAQITVHG-----
NCU07225  121  KRTGGSVNMKNHFDAWTRSGLLTLGTHNYQIVATEGYFSSGSSITINVGASGGSGGGGGGTT

AN3613    -----
AN9365    -----
NCU02855  -----
NCU07225  181  PENPGLTTTPTPNPGNGGGGSCAAKWGQCGGQGWGATCCQSGSTCQAANQWYSQCV

```

Figura 17 – Alinhamento de sequências dos resíduos de aminoácidos das endoxilanas codificadas pelas ORFs AN3613 (endo-1,4-beta-xilanase A de *A. nidulans*), AN9365 (endo-1,4-beta-xilanase B de *A. nidulans*), NCU07225 (endo-1,4-beta-xilanase 2 de *N. crassa*) e NCU02855 (endo-1,4-beta-xilanase A de *N. crassa*). As sequências polipeptídicas das endoxilanas de *A. nidulans* foram obtidas no bando do genoma do fungo (<http://www.aspgd.org/>) e utilizadas como *query* em análises de BlastP no banco de dados do genoma de *N. crassa*.

5.1. Conteúdo de glicogênio e expressão do gene *gsn* e das ORFs codificadoras de endoxilanas em diferentes fontes de carbono

O glicogênio é o carboidrato de extrema importância para os organismos e a regulação de seu metabolismo é fundamental para que um organismo possa responder de modo rápido e eficiente a variações das condições ambientais e nutricionais. Resultados citados anteriormente mostraram que na ausência do ativador transcricional XLR-1, os

níveis de glicogênio bem como do transcrito *gsn* foram extremamente reduzidos, indicando a possível existência de um controle transcricional de *gsn* por este fator de transcrição. Além disso, análise do promotor *gsn* revelou a presença de elementos de DNA regulatórios para este fator de transcrição. Com o objetivo de iniciarmos um estudo funcional deste fator de transcrição em *N. crassa*, e de verificar se a proteína tem um papel na regulação da transcrição do gene *gsn* no fungo quando crescido em fontes de carbono alternativas, como xilose e xilana, foram realizadas análises de crescimento da linhagem selvagem do fungo e da linhagem mutante na ORF NCU06971. As células foram obtidas, como explicado em Materiais e Métodos, crescidas por 24 horas em VM acrescido de frutose 1% (açúcar não repressor), os micélios transferidos para meios contendo xilose ou xilana e coletados após 30 minutos, 1, 2, 6 e 12 horas após a transferência.

Como mostrado na figura 18A, observou-se uma redução considerável no conteúdo de glicogênio tanto na linhagem selvagem (S) quanto na linhagem mutante cuja ORF codificadora de XLR-1 está nocauteada (NCU06971^{ko}) quando crescidos em frutose, xilose ou xilana. Além disso, é importante ressaltar que na linhagem mutante houve um menor acúmulo de glicogênio em relação à linhagem selvagem, mostrando possivelmente a existência de um controle da proteína regulatória XLR-1 na via metabólica do glicogênio, independente da fonte de carbono na qual o fungo cresceu. Foram também realizadas análises da expressão do gene *gsn*, bem como dos transcritos correspondentes às ORFs NCU02855 e NCU07225 nas linhagens selvagem e mutante (figura 18B). Estas análises mostraram que a expressão do gene *gsn* na linhagem selvagem e na mutante foi reduzida em comparação com as mesmas linhagens do fungo crescidas em sacarose (compare com a figura 6, pág. 57). Um fato que chama a atenção é que o nível do transcrito *gsn* foi reduzido em ambas as linhagens, sendo observado na linhagem mutante uma redução mais acentuada. Muito provavelmente, os baixos níveis do transcrito podem estar correlacionados com o reduzido acúmulo de glicogênio observado nesta fonte de carbono.

Em relação às análises da expressão dos genes codificadores de endoxilanasas no fungo *N. crassa* (figura 18B), nota-se que os genes foram expressos a partir de 2 h de crescimento do fungo em xilana e seus transcritos permaneceram presentes até 12 h de crescimento, mas com um nível menor comparado ao tempo de 2 h. Além disso, é importante ressaltar que não se observou a presença dos transcritos referentes às endoxilanasas quando o fungo cresceu na presença de xilose como fonte de carbono. Um ponto que chama a atenção é que ambas as linhagens mostraram a presença dos transcritos, entretanto na linhagem mutante a intensidade foi reduzida em relação à linhagem selvagem. Isto sugere que o fator de transcrição XLR-1 no fungo *N. crassa* regula a expressão de genes codificadores de enzimas xilanolíticas, mas que possivelmente em *N. crassa* não seja o único regulador, pois em sua ausência, os genes codificadores de enzimas xilanolíticas

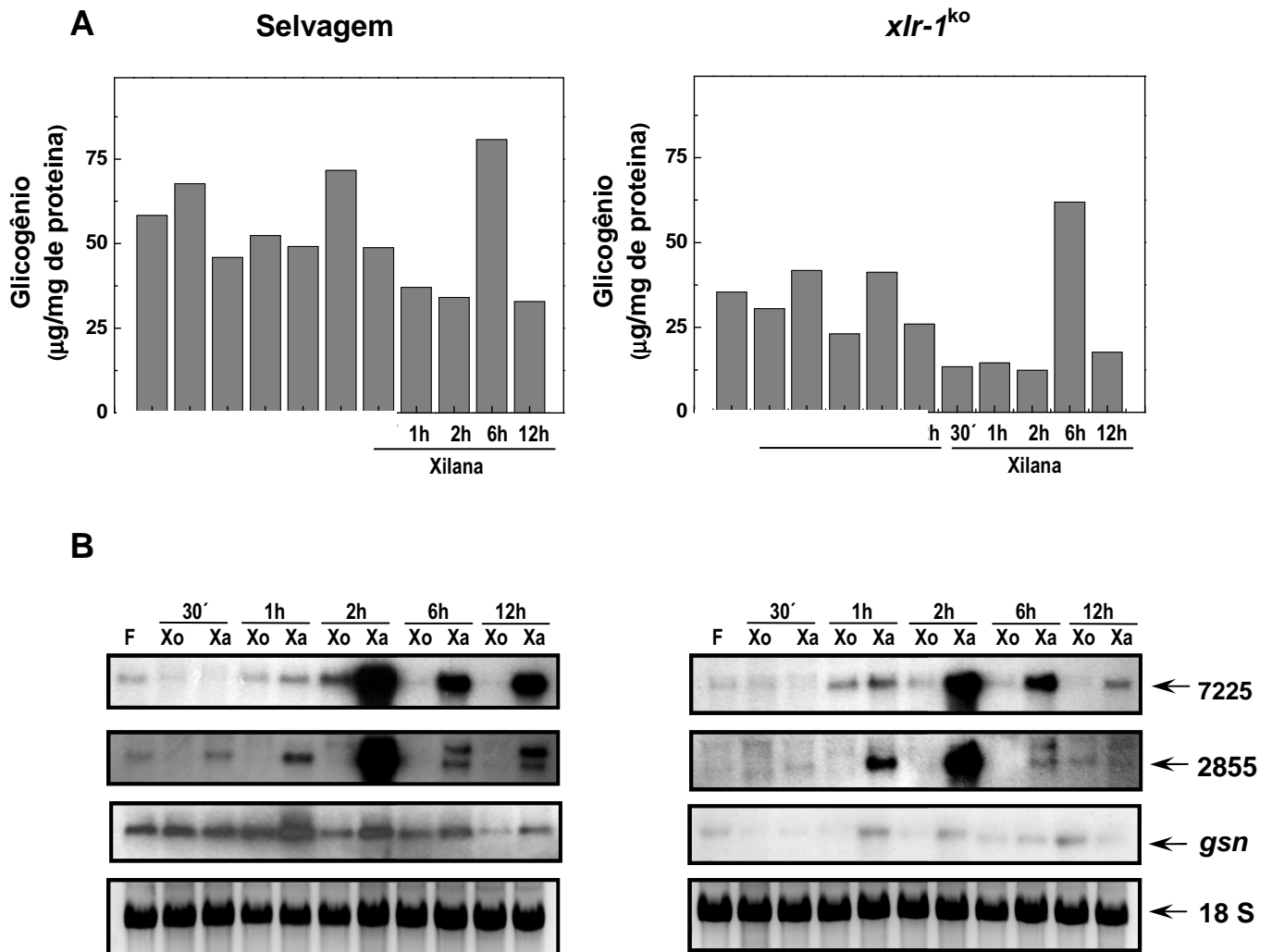


Figura 18 - Análise do conteúdo de glicogênio e da expressão gênica nas linhagens selvagem (S) e mutante em XLR-1 (*xlr-1^{ko}*) por *Northern blot* do fungo *N. crassa* crescido em diferentes fontes de carbono. O micélio foi crescido a 30°C por 24 horas em meio VM acrescido de frutose 1% (F) e transferido para meio VM contendo 1% de xilose (xo) ou xilana (xa) e coletados após 30 minutos, 1 hora, 2 horas, 6 horas e 12 horas. **(A)** Acúmulo de glicogênio. **(B)** Análise da expressão do gene *gsn* nas amostras analisadas em A. As amostras de RNA total foram separadas por eletroforese em gel de ágarose contendo formaldeído, transferidas para uma membrana de nylon neutra e hibridizadas com as seqüências nucleotídicas das ORFs NCU07225, NCU02855 e fragmento do gene *gsn* marcadas radioativamente com $\alpha^{32}\text{P}$ -dATP e utilizadas como sonda.

foram transcritos. Diferentemente, em *A. nidulans*, a ausência de XlnR leva a não expressão de genes codificadores de enzimas xilanolíticas (TAMAYO, 2008).

5.2. Análise de ligação DNA-proteína por EMSA

Os resultados obtidos com as dosagens do conteúdo de glicogênio antes e após o choque térmico, bem como da expressão dos genes *gsn* e *gpn* reforçam a ideia de que o fator de transcrição XLR-1 esteja possivelmente envolvido no controle do metabolismo do glicogênio por meio do controle transcricional de *gsn*. Devido à presença de *motifs* de DNA para este fator de transcrição na região promotora de *gsn* foram realizados ensaios de retardamento em gel a fim de verificar uma possível interação DNA-proteína utilizando um fragmento de DNA com 120 pb da região promotora de *gsn* contendo o *motif* de DNA 5'-GGCTGA-3'. Um fato curioso é que este fator de transcrição apresenta um domínio específico de fatores de transcrição de fungos entre os resíduos de aminoácido 530 e 622, o qual não se tem nenhuma informação sobre sua função na proteína. Foram produzidas na forma recombinante, a proteína XLR-1 inteira e uma forma truncada em seu resíduo 160 (figura 6), a qual possui somente seu domínio de interação ao DNA.

Ensaio de retardamento em gel utilizando quantidades variáveis de 1 a 25 µg das proteínas HisXLR-1 e HisXLR-1Δ160 produzidas na forma recombinante foram realizados com o intuito de determinar qual seria a melhor concentração para a interação destas proteínas com o fragmento do promotor *gsn* contendo o *motif* 5'-GGCTGA-3'. Após vários ensaios, não foi possível verificar a formação de complexos DNA-proteína utilizando tanto a forma inteira da proteína quanto a forma truncada. Pelo fato dos mecanismos regulatórios de XlnR em *Aspergillus spp.* serem muito bem conhecidos (TAMAYO et al., 2008), as regiões 5'-flanqueadoras dos genes codificadores de endoxilases do fungo *N. crassa* foram analisadas em busca de *motifs* 5'-GGCTRA-3', com o intuito de verificar se a ausência na formação de complexos ocorreu devido a problemas experimentais ou não. As regiões de ambos genes mostraram a existência do referido *motif*. Duas regiões foram selecionadas como sonda, uma na região 5'-flanqueadora da ORF NCU02855 (137 pb), a qual possui 1 *motif* para XLR-1 localizado na posição -474 pb do *codon* de iniciação da tradução ATG e outra na região 5'-flanqueadora da ORF NCU07225 (162 pb), a qual possui 1 *motif* localizado na posição -245 pb do *codon* de iniciação da tradução ATG. Outros ensaios de retardamento em gel foram realizados com estas regiões como sondas marcadas radioativamente, bem como com uma quantidade crescente das proteínas (1 a 25 µg) HisXLR-1 e HisXLR-1Δ160 recombinantes. Mais uma vez não foi possível verificar a formação de complexos DNA-proteína utilizando tanto a forma inteira da proteína quanto a forma truncada.

Estes resultados foram intrigantes, pois mostraram que o fator de transcrição XLR-1 não liga à região promotora do gene *gsn* e de genes codificadores de enzimas xilanolíticas. Entretanto, como mostrado na figura 18, tais genes são possivelmente controlados por XLR-1. Além disso, resultados prévios do laboratório mostraram a formação de complexos DNA-proteína quando o mesmo fragmento do promotor *gsn* foi analisado com extrato proteico total do fungo (GONÇALVES et al., 2011). Ensaio adicionais terão que ser realizados para uma caracterização funcional mais completa do papel deste fator de transcrição em *N. crassa*.

6. Caracterização funcional da proteína hipotética codificada pela ORF NCU04390

Pelos dados obtidos anteriormente, esta proteína hipotética se mostrou extremamente promissora, podendo contribuir de forma significativa para uma melhor compreensão dos mecanismos regulatórios envolvidos no metabolismo de glicogênio. Análises do acúmulo de glicogênio na linhagem FGSC#11134 (NCU04390^{ko}) mostraram a ocorrência de um aumento expressivo no acúmulo de glicogênio após a condição de choque térmico, bem como do aumento do transcrito *gsn*, comportamento extremamente diferente ao apresentado pela linhagem selvagem. Pelo fato desta proteína estar anotada como hipotética, portanto não se conhece o seu provável sítio de ligação ao DNA, foram realizadas análises de imunoprecipitação da cromatina seguida de sequenciamento de alto desempenho (ChIP-Seq), com o fungo crescido a 30°C e após condição de estresse térmico (45°C). Esta parte do trabalho foi realizado durante um estágio sanduíche.

6.1 Introdução do *tag* GFP na região C-terminal da proteína 4390

Para a realização da análise de ChIP-Seq, foi essencial a introdução de um *tag* para que, após o *crosslink* desta proteína à dupla fita de DNA e da fragmentação da cromatina por sonicação, os complexos formados pela proteína de interesse junto às suas regiões alvo no DNA fossem recuperados utilizando um anticorpo contra o *tag*. Neste caso foi escolhido a introdução do *tag* GFP na região C-terminal da proteína hipotética 4390. O esquema detalhado do procedimento utilizado está mostrado na figura 19. Para isso, foi necessário o desenho de dois pares de oligonucleotídeos denominados 4390GLY-F, 4390GLY-R (*amplicom* GLY), 4390LOX-F e 4390LOX-R (*amplicon* LOX), cada um com aproximadamente 1 kb de tamanho. É importante ressaltar que o *amplicon* GLY se localiza *upstream* ao *stop codon* e o *amplicon* LOX se localiza *downstream* ao *stop codon*. Outro detalhe extremamente importante foi o acréscimo de aproximadamente 30 *mers* nas regiões

5' dos oligonucleotídeos 4390GLY-R e 4390LOX-F. Estas sequências nucleotídicas adicionais são complementares a cada uma das extremidades do cassete de GFP-*hph*. As bases adicionadas ao oligonucleotídeo 4390GLY-R, além de se sobreporem ao *stop codon* do gene, são complementares à região codificadora do *linker* codificador de 10 Gly entre a proteína e a sequência codificadora de GFP. Já a sequência nucleotídica adicionada ao oligonucleotídeo 4390LOX-F é complementar a uma região do gene *hph* presente no cassete GFP-*hph*. Esta abordagem permitiu que as regiões acrescentadas nos oligonucleotídeos fossem capazes de se sobreporem em uma reação de *Fusion-PCR*, gerando os *split markers* 5' e 3', os quais foram utilizados para a cotransformação de conídios do fungo *N. crassa*. Estas reações de amplificação utilizaram DNA genômico com molde. O produto da reação de amplificação utilizando 4390GLY-F e 4390GLY-R foi um fragmento com 1.174 pb, correspondente ao *amplicom* 4390-GLY (dado não mostrado) e o produto da reação de amplificação utilizando 4390LOX-F e 4390LOX-R foi um fragmento com 1.002 pb, correspondente ao *amplicom* 4390-LOX (dado não mostrado). Para a amplificação do cassete de DNA contendo a sequência codificadora do *tag* GFP bem como a sequência codificadora do gene *hph* (GFP-*hph*) foram utilizados os oligonucleotídeos 10XGLY-F e LoxP-R e o vetor pZErO-GFP-*hph* como DNA molde. Foi obtido uma banda com 2,2 kb correspondente à sequência acima citada (dado não mostrado).

Os 3 fragmentos de interesse foram excisados do gel de ágarose, purificados e quantificados. A síntese dos *splitmarkers* 5' e 3' foi realizada a partir da técnica de *Fusion-PCR*. O *splitmarker* 5' foi obtido por uma reação de amplificação utilizando 1 µL do *amplicon* GLY purificado e a mesma quantidade do produto do cassete GFP-*hph*, utilizando os oligonucleotídeos 4390GLY-F e *hph*_SM-R. Já o *splitmarker* 3' foi obtido por uma reação de amplificação utilizando 1 µL do cassete GFP-*hph* purificado e a mesma quantidade do *amplicon* LOX, utilizando os oligonucleotídeos *hph*_SM-F e LoxP-R. Foram obtidos uma banda com 3.026 pb, correspondente ao *splitmarker* 5' e uma banda com 2.089 pb correspondente ao *splitmaker* 3' (dados não mostrado). Os fragmentos de interesse foram excisados do gel de ágarose, purificados e quantificados.

Após a transformação de conídios da linhagem selvagem FGSC#9718 do fungo *N. crassa*, 5 transformantes foram aleatoriamente selecionados e crescidos individualmente em tubos contendo VM acrescido de glicose 1,5% e higromicina. A presença da região codificadora do *tag* GFP foi confirmada pela técnica de *Southern blot*, após crescimento dos clones em meio líquido e extração de DNA genômico. A figura 20 mostra o resultado da análise de *Southern blot* dos 5 transformantes selecionados. Pode-se verificar a presença de uma única banda de hibridização com 3,15 kb, representando a inserção de uma única cópia no *locus* correto nos 5 transformantes analisados. O transformante 1 foi selecionado para realizar o retrocruzamento com o objetivo de obter segregantes puros (homokarion)

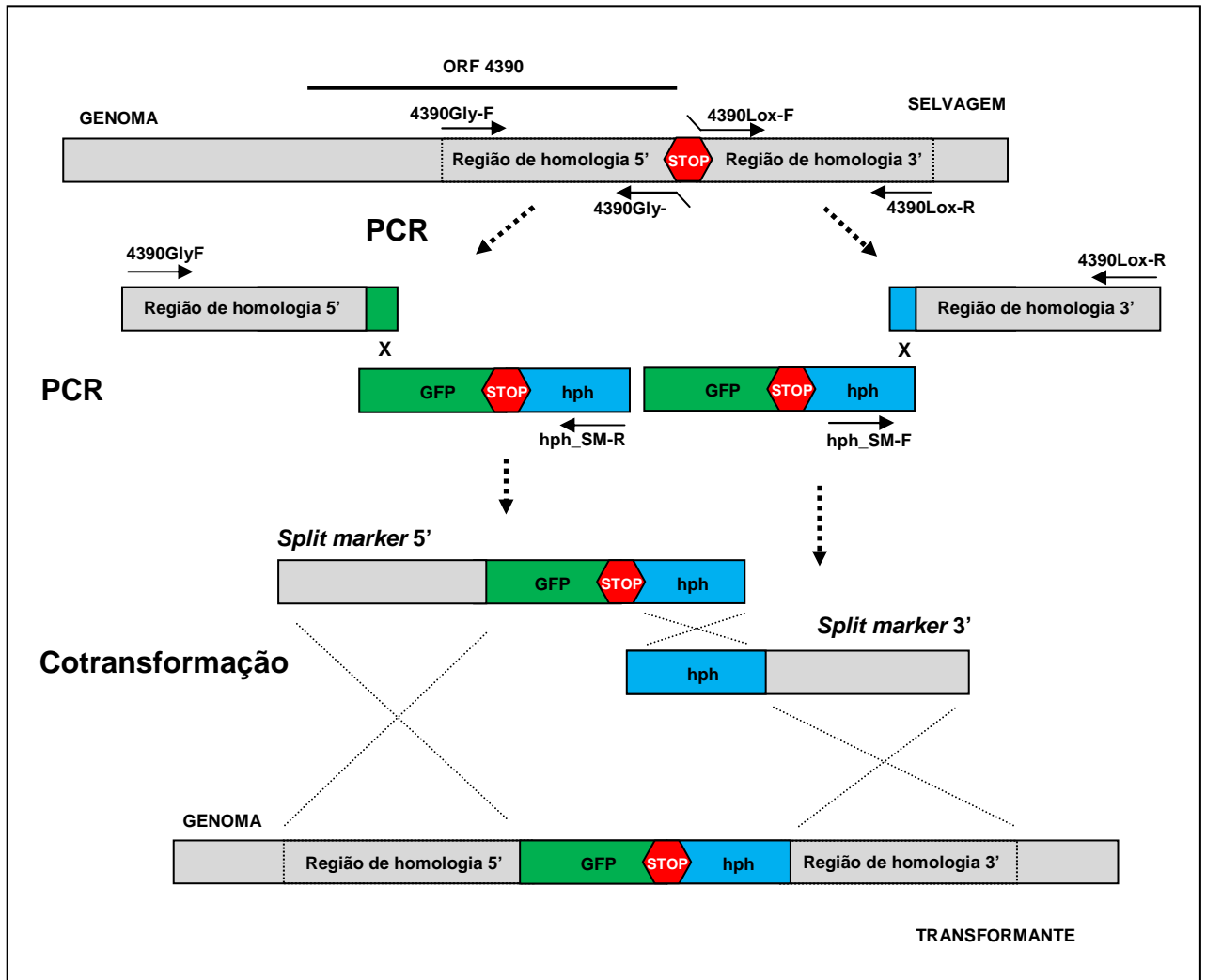


Figura 19. Representação esquemática do processo de introdução de *tag* utilizando *split markers*. Oligonucleotídeos foram desenhados para a amplificação das regiões 5' e 3' flangeadoras aos *stop codon*. Na região 5' dos primers 4390GLY-R e 4390LOX-F foram acrescentadas sequências de 30 *mers* a fim de sobrepremem ao *cassette* de DNA GFP-*hph*.

contendo a proteína fusionada com o *tag*. A linhagem selvagem FGSC#9718 (A) do fungo *N. crassa* foi utilizada com linhagem receptora, realizando então o cruzamento das linhagens como descrito Materiais e Métodos. Após 18 dias de cruzamento, os ascósporos foram coletados da tampa da placa com água estéril, submetidos a choque térmico para sua ativação e plaqueados em VM acrescido de sacarose 1,5% e higromicina para a seleção dos segregantes portadores da marca de seleção. Após 16 horas foi realizada uma busca por ascósporos germinados. Um total de 10 ascósporos foram coletados da placa e transferidos para tubos contendo VM sólido acrescido de sacarose e higromicina. Estes tubos foram mantidas a 30°C por 2 dias para promover o crescimento das hifas. Foi constatado que dos 10 transformantes coletados, 8 carregavam a marca de seleção, comprovado pelo fato destes terem crescido em meio mínimo contendo higromicina. O transformante 1 foi selecionado para as análises de ChIP-Seq.

6.2 Imunoprecipitação da cromatina

Foram realizados experimentos de imunoprecipitação da cromatina nas duas condições de crescimento do fungo *N. crassa*, a 30°C e em situação de estresse térmico, com o intuito de verificar a ocorrência de alvos diferenciais do fator de transcrição 4390 em cada uma das condições. Após o *crosslink* das proteínas à cromatina, foi realizada a fragmentação da cromatina por sonicação e uma alíquota do extrato celular (*input*) foi reservada. Após o tratamento do extrato bruto reservado com proteinase K e tratamento térmico, este material anteriormente reservado foi fracionado em gel de ágarose 0,8% para a determinação do tamanho dos fragmentos de DNA resultantes da sonicação. Foi verificado que apesar de todos os ciclos de sonicação dos micélios serem realizados, houve a ocorrência de fragmentos de DNA com diversos tamanhos e em grande quantidade (dado não mostrado), não apresentando problemas para a análise.

Após o reparo de suas extremidades e ligação dos adaptadores em suas extremidades, este material foi fracionado em eletroforese em gel de ágarose *low melting* 2% e fragmentos de DNA com tamanho entre 200 pb a 500 pb foram excisados do gel e purificados. O material purificado foi utilizado como molde para a síntese das bibliotecas de ChIP-Seq (dado não mostrado). As bibliotecas de ChIP-Seq foram quantificadas e 10 nM de cada biblioteca de ChIP-Seq gerada foi utilizada como material de análise nas reações de sequenciamento de alto desempenho.

As análises de ChIP-Seq também foram realizadas utilizando o *tag* V5, menor do que GFP, mas os resultados de sequenciamento ainda não foram disponibilizados.

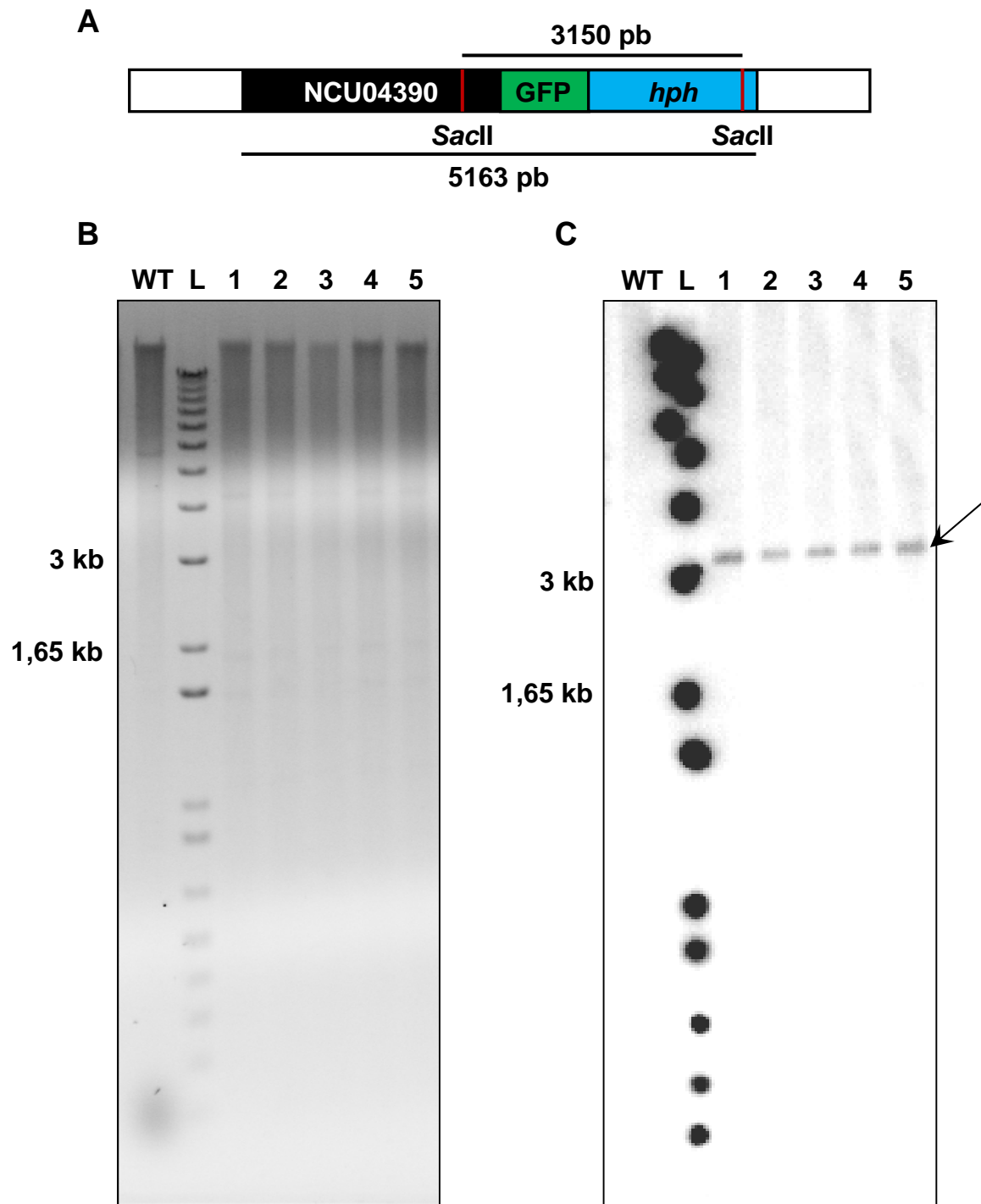


Figura 20. Análise da integração dos *splitmarkers* no genoma do fungo *N. crassa*. DNA genômico de 5 clones transformados foram analisados por *Southern blot* para a confirmação da inserção correta no genoma. (A) Modelo esquemático do gene codificador da proteína hipotética 4390 contendo a região codificadora do *tag* GFP junto a *hph*. (B) Amostras de DNA de cada clone foram digeridos com a enzima *SacII*, fracionadas por eletroforese em gel de ágarose 0,8%. (C) As amostras foram transferidas para membrana de nylon e hibridizadas com o fragmento codificador do *tag* GFP marcado radiativamente. L, DNA ladder 1 kb plus, WT, linhagem selvagem, 1 a 5, transformantes.

6.3. Sequenciamento de alto desempenho

As reações de sequenciamento de alto desempenho foram realizadas nos sequenciadores HiSeq 2000 (Illumina/Solexa). Esta plataforma de sequenciamento utiliza de um método de sequenciamento o qual ocorre em duas etapas: A primeira consiste na clonagem *in vitro* dos fragmentos de DNA gerados pela biblioteca de ChIP-Seq em uma plataforma sólida (*flow cell*) utilizando nucleotídeos não marcados, processo também conhecido como PCR de fase sólida (MARDIS, 2008; FEDURCO et al., 2006; TURCATTI et al., 2008). A superfície da *flow cell* é dividida em oito *lanes* que podem ser utilizadas para o sequenciamento de até oito bibliotecas distintas. Em cada *lane* existem oligonucleotídeos fixados à superfície da *flow cell* pela extremidade 5' deixando a extremidade 3'. Estes oligonucleotídeos possuem suas extremidades 3' livres e, por serem complementares aos utilizados na síntese das bibliotecas de ChIP-Seq, são utilizados no processo de síntese dos fragmentos imobilizados no suporte por hibridização.

A segunda etapa consiste no sequenciamento propriamente dito, na qual são utilizados os quatro nucleotídeos terminadores, com a hidroxila do C3 bloqueada e marcados com diferentes fluoróforos para o sequenciamento das colônias formadas na etapa anterior. O processo de sequenciamento corre após o anelamento de um oligonucleotídeo aos fragmentos de DNA sintetizados a partir dos oligonucleotídeos imobilizados na *flow cell*. Inicialmente ocorre a incorporação e detecção de um único nucleotídeo terminador às cadeias de DNA. Após sua detecção do nucleotídeo incorporado por uma câmera, a *flow cell* é lavada e os nucleotídeos incorporados ao DNA perdem o fluoróforo com o qual estava marcado e possuem a hidroxila de seu C3 desbloqueada, permitindo a incorporação de um novo nucleotídeo no ciclo seguinte. Esses ciclos foram repetidos 50 vezes, gerando inúmeras cópias de cada fragmento gerado na biblioteca de ChIP-Seq, as quais permanecem próximas umas das outras, formando um *cluster* de sequenciamento.

As análises de ChIP-Seq foram feitas em duplicatas para cada uma das situações de crescimento do fungo. Resultados de análise dos fragmentos sequenciados pelo software *Bowtie*, mostraram que para as amostras do fungo crescidas a 30°C foram analisadas 11518827 e 11342780 sequências curtas, com 50 pb, onde 70,97% e 65,73% delas foram alinhadas junto ao genoma do fungo *N. crassa* com sucesso. Já para as amostras do fungo submetidas à condição de estresse térmico, foram analisadas 13503675 e 9670717 sequências curtas, com 50 pb, onde 89,65% e 76,65% das leituras foram alinhadas junto ao genoma do fungo *N. crassa* com sucesso. Estes resultados foram validados utilizando o software MACS v1.4.2 e confirmados por análise visual dos picos utilizando o visualizador IGV 2.1. Apesar da diferença obtida na quantidade total de leituras realizadas entre as

réplicas, os resultados obtidos foram semelhantes, pois os resultados dos enriquecimentos dos picos com valores aceitáveis, avaliados pelo software MACS v1.4.2 foram bem próximos.

Um fato interessante é que na condição de crescimento a 30°C, não foram encontrados picos enriquecidos com *score* aceitável, sugerindo que este fator de transcrição possivelmente não interaja com a cromatina nesta condição de crescimento. Por outro lado, as análises do enriquecimento de picos obtidos nas amostras do fungo submetidas a choque térmico apresentaram 210 regiões de confiáveis de interação deste fator de transcrição ao DNA, dos quais 113 regiões estão localizadas na região 5'-flanqueadora de genes, 96 estão localizadas na região 3'-flanqueadora de genes e 4 regiões estão localizadas na sequência codificadora dos genes. Esta análise foi realizada pela validação do software MACS v1.4.2 bem como a visualização dos picos pelo software IGV, devido à capacidade de diferenciar picos propriamente ditos ao invés de ruídos.

Ao fazer uma análise mais detalhada, foi possível verificar que dos 210 genes possivelmente controlados pela proteína 4390, 128 são codificadores de proteínas hipotéticas e os restantes estão envolvidos em diferentes processos celulares. Para facilitar a compreensão, os genes selecionados possivelmente regulados pela proteína foram agrupados em 13 categorias distintas, de acordo com suas funções, como mostra a figura 21. Em relação aos genes codificadores de proteínas com funções conhecidas, 22 são envolvidos no metabolismo de carbono, 8 são envolvidos com transportadores, 8 envolvidos com biossíntese de vitaminas, aminoácidos e coenzimas, 6 envolvidos em processos celulares, 6 envolvidos no metabolismo de ácidos graxos, 5 são genes mitocondriais, 4 envolvidos no metabolismo de nitrogênio, enxofre ou fósforo, 3 envolvidos com metabolismo de RNA, 3 com metabolismo secundário, 2 envolvidos com DNA (síntese, reparo, etc.), 2 envolvidos com peroxissomo e apenas 1 envolvido no ciclo celular. Na Tabela 4 estão mostradas as proteínas relacionadas ao metabolismo de carbono. Nela podemos observar a presença de vários genes envolvidos em diferentes vias metabólicas relacionadas ao metabolismo e as enzimas relacionadas ao metabolismo de glicogênio estão destacadas.

6.4. Determinação do *motif* de DNA para a proteína 4390

Após a determinação dos genes possivelmente regulados pela proteína 4390 na condição de choque térmico, as sequências nucleotídicas presentes na região destacada pelos 40 picos com o melhor *score* foram analisadas nos servidores *SCOPE* e *Bioprosector* para a determinação do possível *motif* de DNA para a proteína 4390. Como pode ser visto na figura 22, ambas as ferramentas identificaram a região consenso que corresponde ao *motif* de DNA alvo para a proteína 4390, como algumas variações. Podemos verificar que o

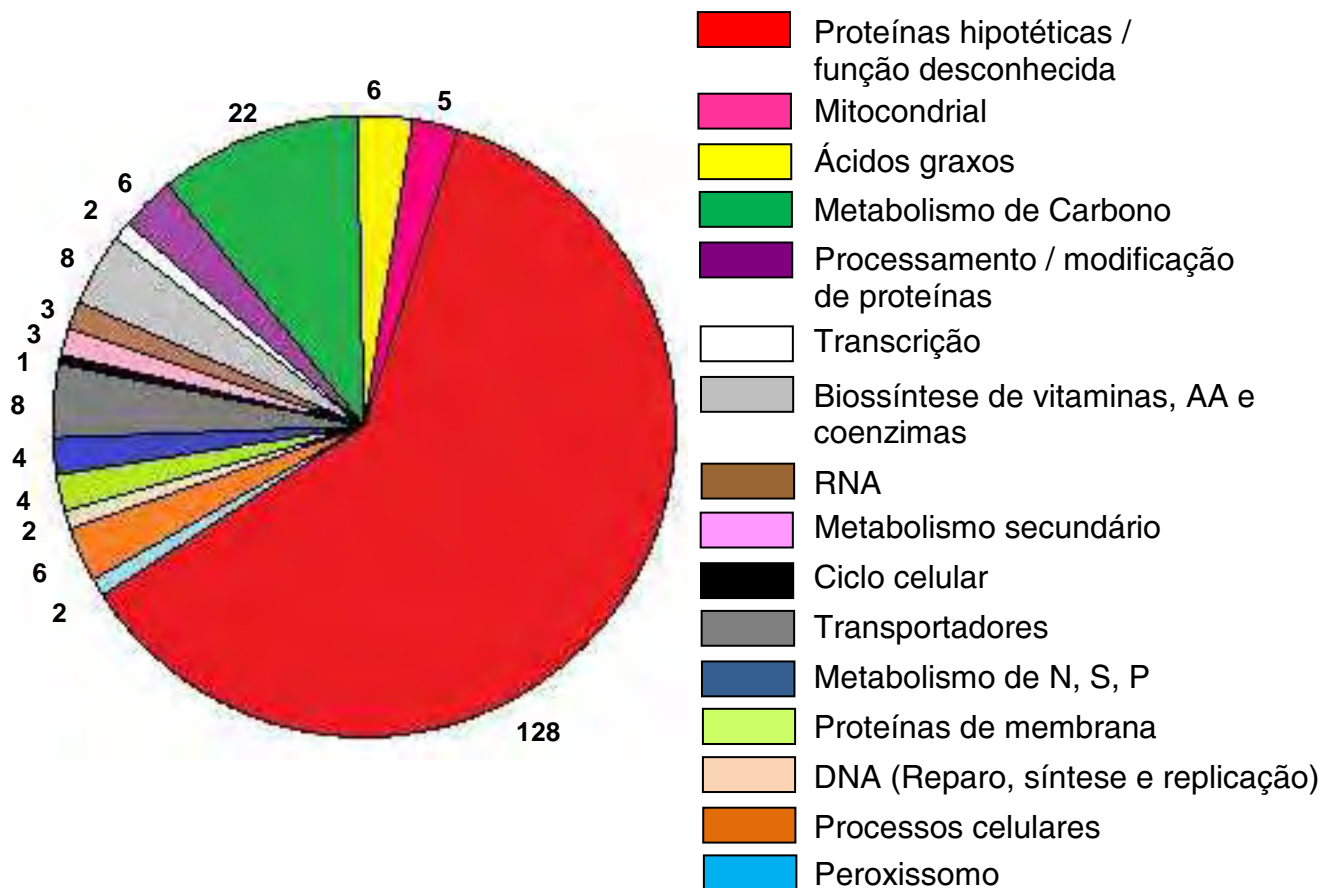


Figura 21 - Classificação funcional dos produtos dos genes possivelmente controlados pelo fator de transcrição 4390 em situação de choque térmico. Os genes selecionados foram agrupados em diversas categorias considerando as funções das proteínas codificadas por eles.

resultado obtido pela análise no *SCOPE* (22A) mostrou uma sequência nucleotídica com um *core* central conservado composto por 5'-CAT-3' contendo uma região variável *upstream* e *downstream*. Já o resultado obtido pelo *Bioprospector* (22B) resultou em uma sequência nucleotídica com um *core* 5'-TAC-3' conservado, contendo regiões variáveis. Apesar dos resultados não terem sido idênticos, é possível notar certa similaridade entre eles, sendo possível concluir que o *motif* provavelmente corresponde à sequência 5'-TACATR-3'.

Tabela 4 – Genes envolvidos no metabolismo de carbono possivelmente regulados pelo fator de transcrição codificado pela ORF NCU04390.

ORF	Anotação
NCU08750	isoamyl alcohol oxidase
NCU07351	alpha-glucuronidase
NCU07027	glycogen phosphorylase
NCU00281	UDP-glucose, esteroil transferase
NCU01120	phosphoglycerate mutase
NCU04510	aldose reductase
NCU06005	glycerol kinase
NCU00629	6-phosphofructokinase
NCU01845	phosphoglycerate mutase family domain-containing protein
NCU05037	dihydrodipicolinate synthetase
NCU10107	ribose 5-phosphate isomerase
NCU00743	glycogen debranching enzyme
NCU07042	alcohol dehydrogenase
NCU06687	glycogen synthase-1
NCU00087	6-phosphogluconolactonase
NCU00609	initiation-specific alpha-1,6-mannosyltransferase
NCU09285	zinc-containing alcohol dehydrogenase
NCU01227	succinyl-CoA ligase alpha-chain
NCU05542	xanthine phosphoribosyltransferase 1
NCU05429	1,4-alpha-glucan branching enzyme
NCU04108	isoamyl alcohol oxidase
NCU06303	alcohol dehydrogenase

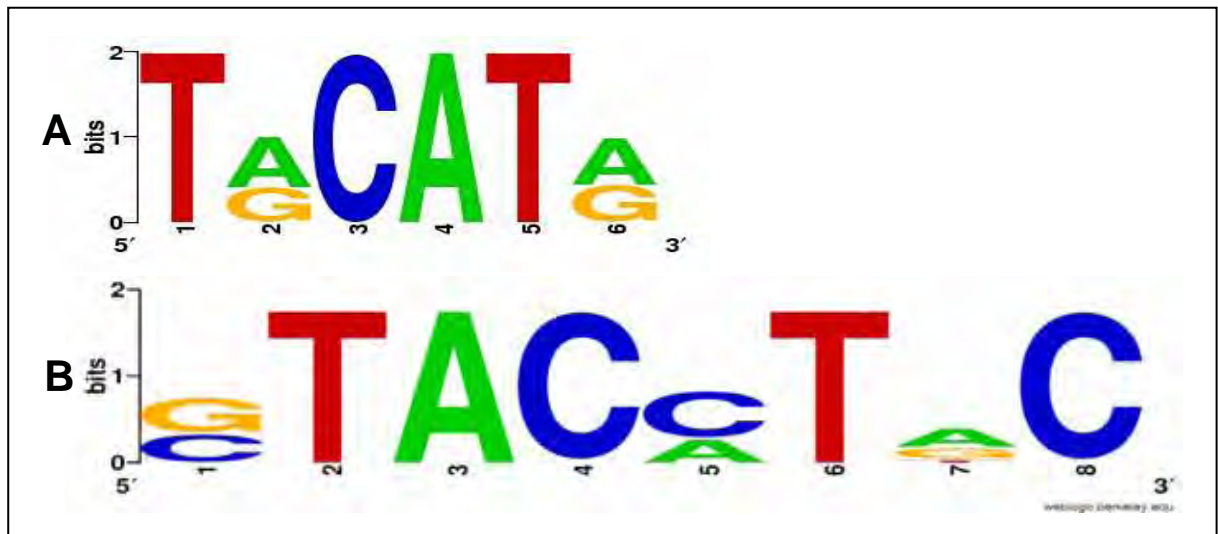


Figura 22. Análise da interação da proteína 4390 ao DNA em situação de estresse térmico revelou a sequência consenso de ligação desta proteína utilizando os servidores *SCOPE* (<http://genie.dartmouth.edu/scope/>) e *Biopro prospector* (<http://ai.stanford.edu/~xslu/BioProspector/> *SCOPE*). A altura relativa de cada nucleotídeo, mostrados na direção 5'-3' (eixo X), indica o grau de conservação da sequência com uma máxima pontuação de 2 (eixo Y).

6.5. Análise das regiões 5' e 3' flanqueadora dos genes envolvidos no metabolismo de glicogênio para *motifs* da proteína 4390

Após a obtenção da possível sequência consenso do *motif* de DNA reconhecido pela proteína 4390 com o auxílio das ferramentas *SCOPE* e *Biopro prospector*, foram realizadas buscas por tal elemento nas regiões 5' e/ou 3' flanqueadoras dos genes envolvidos no metabolismo de glicogênio coberta pelos picos utilizando todas as possíveis variações. Foram encontradas somente sítios com a combinação 5'-TACATA-3' nas regiões dos genes analisados. Além disso, um fato interessante foi a presença de múltiplos sítios nas regiões analisadas. Foram encontrados 5 *motifs* na região 3' flanqueadora do gene codificador da enzima glicogênio fosforilase (NCU07027) nas posições +2054, +2937, +3536, +3577 e +3582, 3 *motifs* na região 5'-flanqueadora do gene codificador da enzima desramificadora do glicogênio (NCU00743) nas posições -373, -457 e -461, 3 *motifs* na região 5'-flanqueadora do gene codificador da enzima ramificadora (NCU05429) nas posições -236, -665 e -1009 e 5 *motifs* na região 5'-flanqueadora do gene codificador da enzima glicogênio sintase

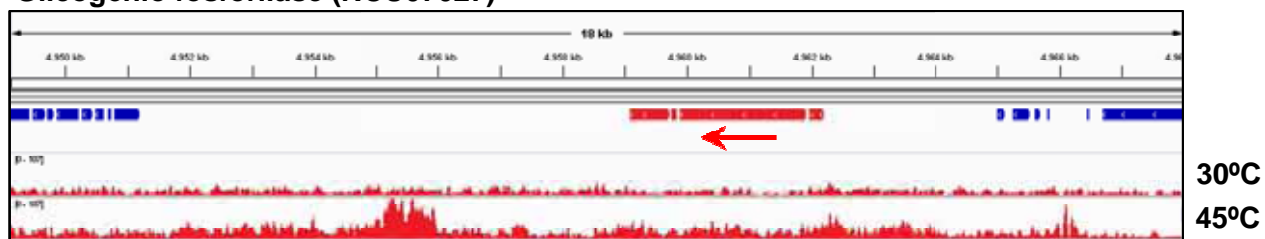
(NCU06687), nas posições -954, -1253, -1666, -2748 e -2752. Um fato extremamente curioso foi o pico que contém os possíveis *motifs* para a proteína 4390 na região 5'-flanqueadora do *gsn* se sobrepor a uma pequena ORF, anotada como uma proteína hipotética e localizada *upstream* ao *gsn*. A figura 23 mostra a posição dos picos obtidos nas análises de ChIP-Seq e as ORFs citadas acima destacadas de vermelho, obtidas pelo software IGV 2.1 (<http://www.broadinstitute.org/igv/>). A figura 24 mostra as sequências nucleotídicas das regiões contendo os *motifs* identificados.

6.6. Ensaio de ligação com a proteína His4390 recombinante por EMSA

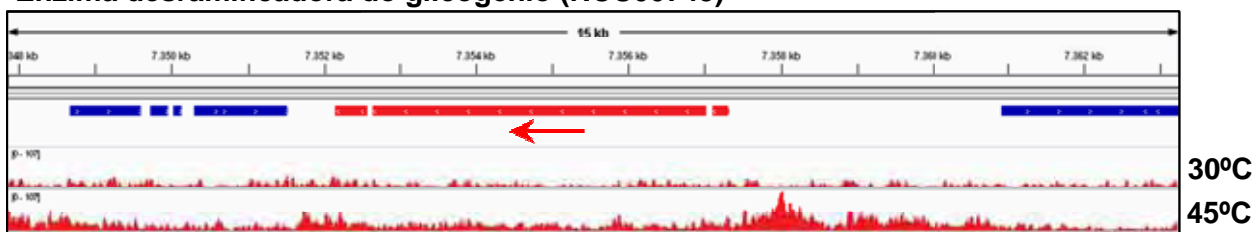
Após a obtenção da localização exata do *motif* possivelmente reconhecido pela proteína hipotética 4390 nas regiões 5' e/ou 3'-flanqueadora dos genes envolvidos no metabolismo do glicogênio, ensaios de interação DNA-proteína *in vitro* foram realizados com o intuito de validar a técnica de ChIP-Seq. Em um primeiro momento, foram analisados os *motifs* presentes na região 5'-flanqueadora do gene que codifica a enzima desramificadora do glicogênio. Foi escolhida como sonda uma região com 170 pb que compreende a sequência entre os nucleotídeos das posições -510 pb e -341 a partir do codon iniciador do gene. Esta região foi selecionada pelo fato do score do pico ter sido um dos mais altos entre os obtidos pelos demais genes envolvidos no metabolismo do glicogênio e pelo fato desta região se localizar possivelmente na região promotora deste gene, pois se encontra bem próximo ao codon iniciador do gene.

Na figura 24A temos uma representação esquemática da posição dos *motifs* para a proteína His4390 na sonda utilizada. Nestes ensaios foi analisada a interação da proteína His4390 produzida na forma recombinante, utilizando uma quantidade crescente de proteína (1 a 30 μ g) com o intuito de determinar a melhor concentração desta proteína para sua interação com a sonda marcada radioativamente (Figura 24B). Podemos verificar a formação de um complexo DNA-proteína com elevado peso molecular na canaleta 6, a qual corresponde a 6 μ g de proteína, ressaltado na figura por uma seta, o qual representa a interação da proteína 4390 com a sonda de DNA marcada radioativamente. Com o intuito de testar a especificidade da interação desta proteína à sonda marcada, foram adicionados na reação de *binding* diferentes concentrações do competidor específico (fragmento de 170 pb não marcado). Pode-se observar na figura 24C que, com a adição do competidor a partir de 10X em excesso molar, ocorre a ausência da formação do complexo DNA-proteína marcado radioativamente. Isso se deve ao fato de que a proteína His4390 recombinante interage de forma específica com o competidor específico, deixando a sonda marcada totalmente livre, sendo então detectada na região mais inferior do gel. Foram também realizadas análises

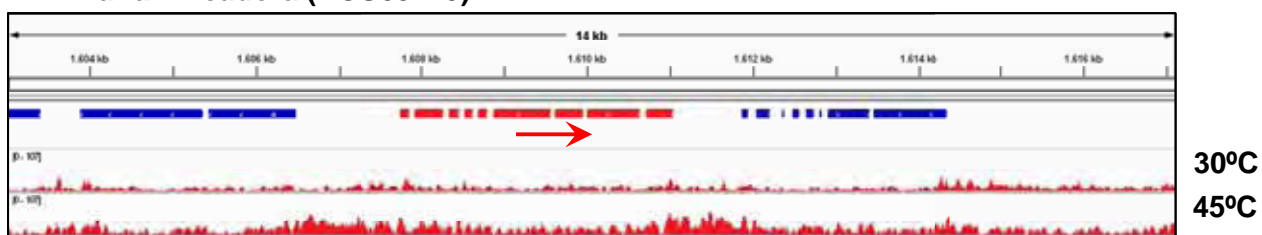
Glicogênio fosforilase (NCU07027)



Enzima desramificadora do glicogênio (NCU00743)



Enzima ramificadora (NCU05429)



Glicogênio sintase (NCU06687)

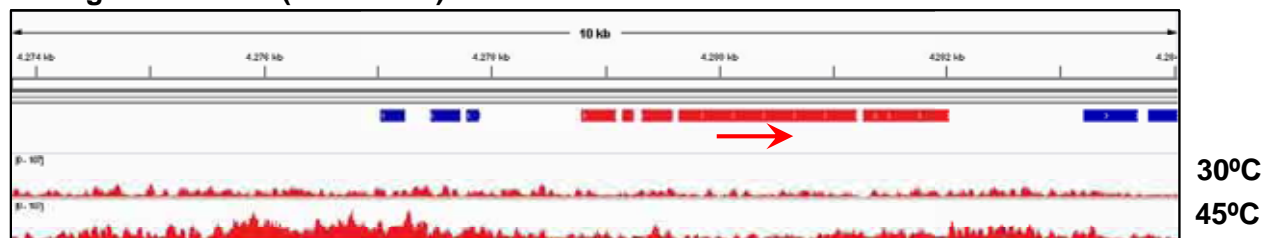


Figura 23 - Análise dos picos nas regiões 5' e 3'-flanqueadoras dos genes codificadores das enzimas envolvidas no metabolismo de glicogênio cobertas pelos picos. A orientação dos genes (5'→3') está indicada pela seta.

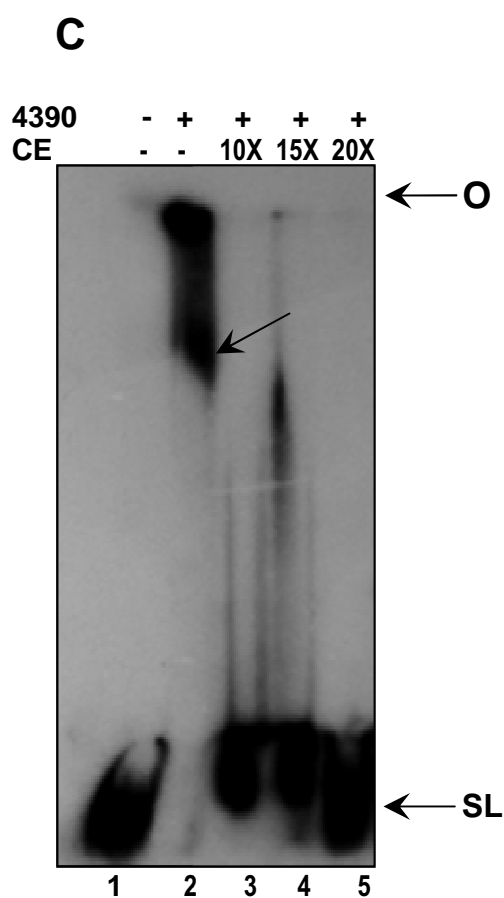
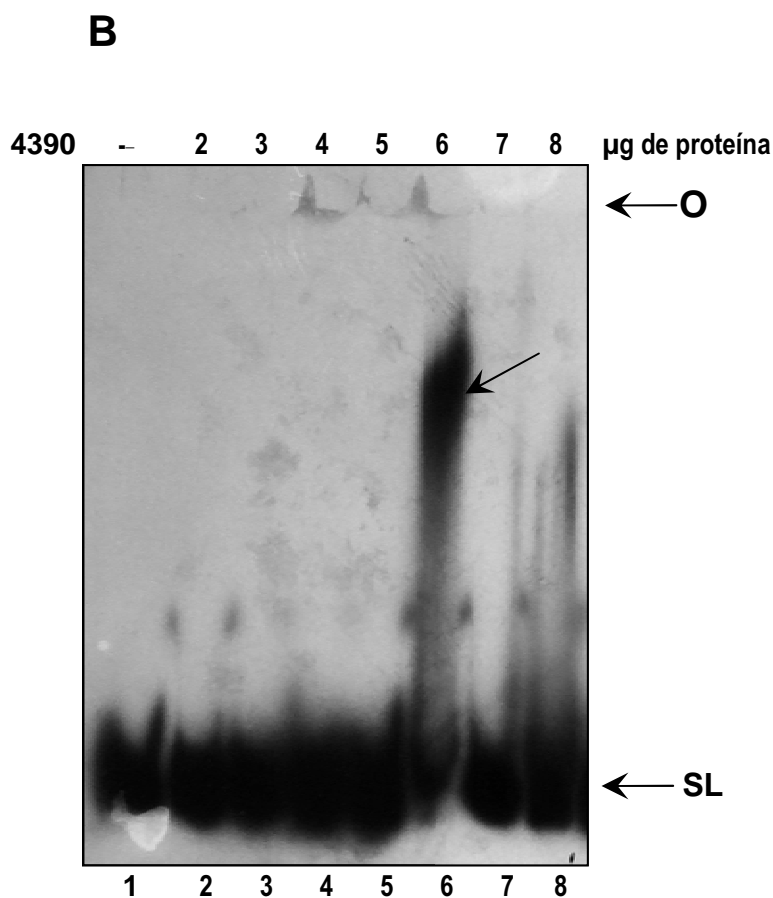
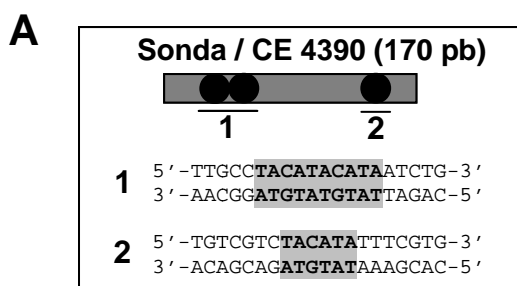


Figura 24 - Ensaio de interação DNA-proteína *in vitro* (EMSA) mostrando a interação da proteína recombinante His4390 ao fragmento da região 5'-flanqueadora do gene codificador da enzima desramificadora do glicogênio contendo 3 *motifs* TACATA para esta proteína. (A) Representação esquemática da sonda utilizada neste ensaio. (B) Quantidades crescentes de proteínas foram adicionadas à sonda. (C) Quantidades crescentes do competidor específico (CE) foram testadas com 6 μg da proteína recombinante. O, origem do gel; SL, sonda livre. O complexo DNA-proteína está indicado pela seta. Eletroforese em gel de poliacrilamida 5% não desnaturante em tampão TBE 0,5X.

de interação DNA-proteína com as duas formas truncadas da proteína hipotética 4390, a His4390 Δ 483 e His4390 Δ 158 utilizando a mesma sonda das análises com a proteína inteira His4390. Em relação à proteína His4390 Δ 158 pode-se verificar na figura 25B a ausência da formação de complexos DNA-proteína utilizando uma quantidade crescente de proteína recombinante (2 a 15 μ g). Por outro lado, podemos verificar pela figura 26C a formação de um complexo DNA-proteína nas análises de EMSA ao utilizar a partir de 3 μ g da proteína His4390 Δ 483. Foram também adicionados à reação de *binding* diferentes concentrações do competidor específico (fragmento de 170 pb não marcado) com o intuito de testar a especificidade da interação desta proteína à sonda marcada. Pode-se observar na canaleta 5 da figura 25C que com a adição do competidor a partir de 30X em excesso molar ocorre a ausência da formação do complexo DNA-proteína marcado radioativamente, possivelmente pelo fato de que a proteína His4390 Δ 483 recombinante interage de forma específica com o competidor específico, deixando a sonda marcada totalmente livre.

Estes resultados são preliminares e novos experimentos serão realizados. Entretanto, o resultado obtido confirma a ligação da proteína à região promotora do gene que codifica a enzima desramificadora da quebra de glicogênio e aponta para um papel regulatório do produto da ORF na situação de estresse térmico, um dado ainda não descrito na literatura.

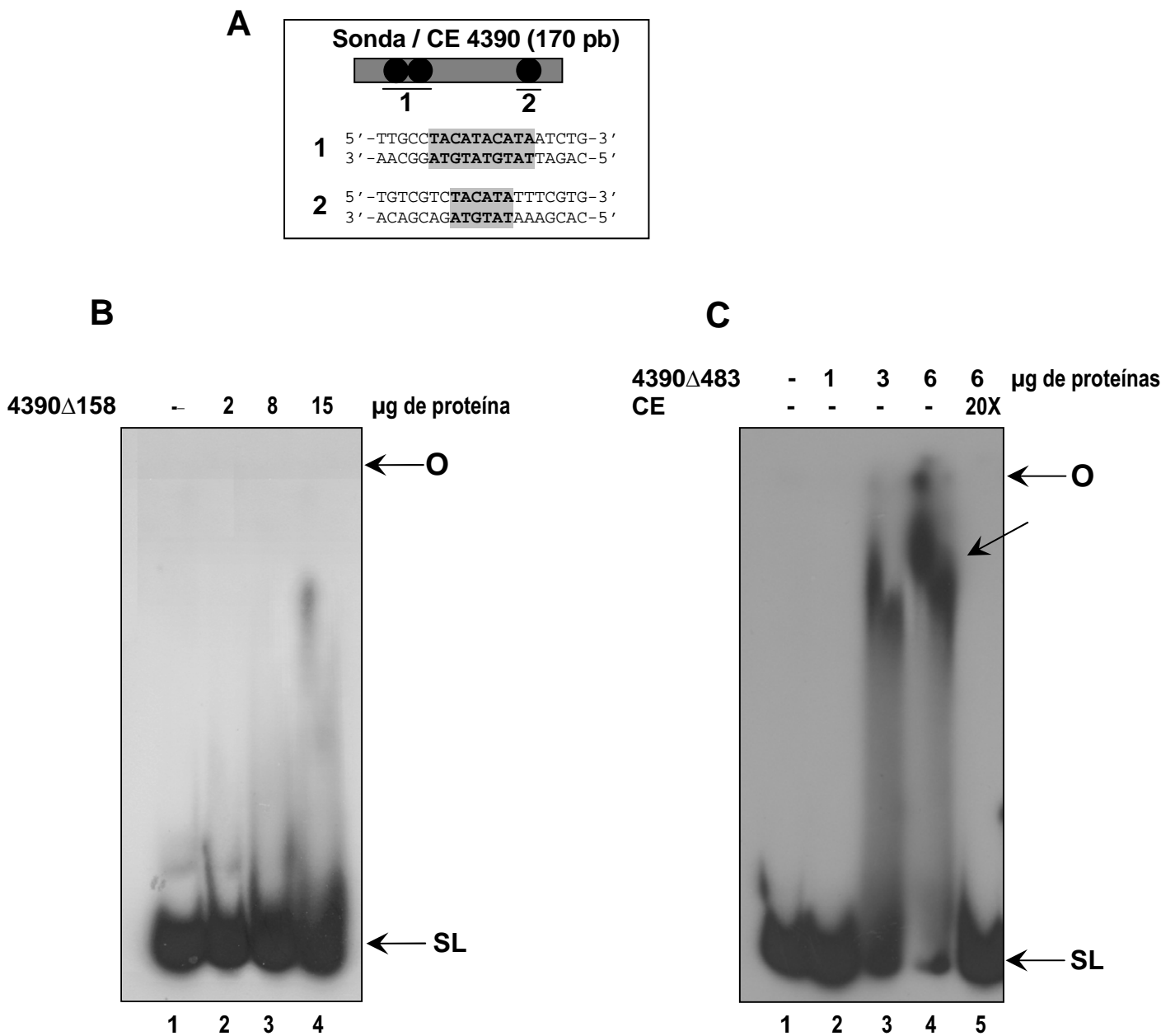


Figura 25 - Ensaio de interação DNA-proteína *in vitro* (EMSA) mostrando a interação das forams truncadas da proteína recombinante 4390 ao fragmento da região 5'-flanqueadora do gene codificador da enzima desramificadora do glicogênio contendo 3 *motifs* TACATA para esta proteína. (A) Representação esquemática da sonda utilizada neste ensaio. (B) Quantidades crescentes da proteína recombinante His4390 Δ 158 foram adicionadas à sonda. (C) Quantidades crescentes da proteína His4390 Δ 483 utilizadas nos ensaios de EMSA. Diferentes concentrações do competidor específico (CE) foram testadas com 6 μ g da proteína recombinante. O, origem do gel; SL, sonda livre. O complexo DNA-proteína está indicado pela seta. Eletroforese em gel de poliácridamida 5% não desnaturante em tampão TBE 0,5X.

Discussão

A regulação da transcrição é um dos principais pontos de controle da atividade gênica, a qual é responsável pelo controle dos processos biológicos nos organismos eucariotos. Embora a síntese proteica possa ser controlada na célula por meio de uma variedade de mecanismos, alteração do nível transcricional do gene é o processo celular mais direto e utilizado pela célula (WASSERMAN & FICKETT, 1998). Em contraste com os organismos procariotos, onde o processo de regulação transcricional pode ser compreendido por meio da indução por um único fator, nos organismos eucarióticos este processo é principalmente controlado por uma sofisticada interação de múltiplos fatores de transcrição, que interagem com elementos *cis* distribuídos ao longo de todo o genoma, fornecendo um aumento do controle específico dos genes.

Embora um grande número de projetos genomas tenha sido finalizado, somente uma pequena fração dos mecanismos regulatórios da expressão gênica nos organismos eucariotos é conhecida. Este processo representa a coordenação espacial e temporal da tradução do material genético para proteínas. Para isso, uma complexa maquinaria interpreta os estímulos internos e externos transferindo funções aos fatores de transcrição. O conhecimento dos mecanismos regulatórios é essencial para uma compreensão profunda da maioria dos aspectos fundamentais da vida, ou seja, decifrar a linguagem dos elementos *cis* é descrever sua influência no fator de transcrição e no controle do gene alvo. Um conhecimento da posição de elementos regulatórios individuais, bem como sua atuação na regulação gênica é um pré-requisito para o início da compreensão dos processos moleculares responsáveis pela manutenção da homeostase celular.

De um modo geral, este trabalho teve como objetivo caracterizar funcionalmente alguns fatores de transcrição envolvidos no metabolismo do glicogênio no fungo *N. crassa*. Com a utilização de uma coleção de linhagens mutantes em fatores de transcrição, visando buscar linhagens apresentando alterações no conteúdo de glicogênio, foi possível correlacionar o fenótipo observado de cada linhagem mutante com a ausência do fator de transcrição. Esta estratégia resultou na identificação de 10 fatores de transcrição possivelmente envolvidos na regulação do metabolismo de glicogênio no fungo (GONÇALVES et al., 2011). A partir deste resultado foram realizadas análises dos transcritos de *gsn* e *gpn* com o intuito de verificar se alguns dos fatores de transcrição previamente isolados atuavam no metabolismo do glicogênio por meio do controle transcricional destes genes.

Apesar de não se ter obtido uma correlação direta entre o acúmulo de glicogênio e o padrão de expressão destes genes, na maioria das 10 linhagens mutantes analisadas, verificamos uma correspondência entre o conteúdo de glicogênio e a transcrição dos genes *gsn* e *gpn* nas linhagens contendo as ORFs NCU06971 e NCU04390 nocauteadas. Na linhagem mutante na ORF NCU06971, foi verificado um baixo acúmulo de glicogênio

quando o fungo foi crescido a 30°C bem como uma baixa transcrição do gene *gsn* na mesma condição. Em relação à linhagem mutante na ORF NCU04390 foi verificado um elevado acúmulo de glicogênio após a condição de choque térmico, seguida por um aumento expressivo na transcrição de *gsn* sob a mesma condição. O produto da ORF NCU06971 corresponde ao fator de transcrição XlnR, descrito como regulador transcricional de genes codificadores de enzimas xilanolíticas e/ou celulolíticas em fungos filamentosos e o produto da ORF NCU04390 corresponde a uma proteína hipotética. Como consequência destes resultados, a análise funcional destas proteínas constituiu o principal objetivo deste trabalho.

Análises de citometria de fluxo utilizando conídios das 10 linhagens mutantes inicialmente selecionadas mostraram que a ausência de apenas um fator de transcrição, codificado pela ORF NCU09739, promove alterações tanto da morfologia conidial quanto na progressão do ciclo celular. O gene, no banco de dados de *N. crassa*, está anotado como *ada-7* (*all development altered-7*) e o produto gênico, muito possivelmente, está envolvido em vias regulatórias essenciais à regulação do metabolismo celular, bem como na regulação dos mecanismos celulares envolvidos na morfologia conidial e no ciclo celular do fungo *N. crassa*.

1. O fator de transcrição XLR-1 e regulação do metabolismo do glicogênio

Foram realizadas diversas abordagens metodológicas para a caracterização funcional do fator de transcrição XLR-1, com o intuito de verificar se este fator de transcrição obedece no fungo *N. crassa* os mesmos padrões de regulação observados em outros fungos filamentosos. Pelo fato deste fator de transcrição estar relacionado à ativação de genes codificadores de enzimas envolvidas principalmente na degradação de hemicelulose nos fungos filamentosos como, por exemplo, *Aspergillus* e *Trichodera* e seu mecanismo de ação ser muito bem compreendido, foram realizadas análises transcricionais de genes codificadores de duas endoxilanasas descritas serem ativadas por XlnR neste organismos. Em nossas análises, as linhagens selvagem e mutante em XLR-1 foram crescidas em meio contendo frutose e transferidas para meios contendo xilose ou xilana como fonte de carbono. É descrito que a ação do fator de transcrição XlnR, em outros fungos filamentosos, é reprimida na presença de glicose por meio do sistema de repressão catabólica (CCR), mediado pelo fator de transcrição CreA. Esta repressão promove a inibição da transcrição de genes codificadores de endoxilanasas no fungo *A. nidulans* por meio da inativação do fator de transcrição XlnR (TAMAYO et al., 2008).

Neste trabalho, foram analisados tanto o conteúdo de glicogênio, quanto o nível dos transcritos codificadores de enzimas endoxilanasas e algumas conclusões puderam ser

obtidas a partir dos resultados obtidos. Em relação ao conteúdo de glicogênio foi verificado que na ausência do fator de transcrição XLR-1 ocorre um menor acúmulo do carboidrato em questão. Além disso, os níveis do transcrito *gsn* na linhagem selvagem foram superiores ao apresentado pela linhagem mutante *xlr-1^{KO}* em todas as condições analisadas, demonstrando que na ausência deste fator de transcrição ocorre um menor acúmulo de glicogênio, possivelmente devido a uma diminuição na expressão de *gsn*, independente da fonte de carbono na qual o fungo cresceu. Análises dos transcritos de genes codificadores de endoxilanasas mostraram a presença dos transcritos para ambos os genes tanto na linhagem selvagem (maior quantidade) quanto na linhagem mutante (menor quantidade) após 2 h de exposição à xilana. Não foi possível verificar a presença de transcritos de ambos os genes quando o fungo utilizou xilose como fonte de carbono. Este dado sugere que, diferentemente do apresentado em espécies do gênero *Aspergillus* e *Trichoderma*, que o fator de transcrição XLR-1 está relacionado com o metabolismo de hemicelulose, mas possivelmente não seja o único fator de transcrição envolvido neste processo em *N. crassa*.

Como a expressão do gene *gsn* na linhagem mutante foi reduzida em relação à selvagem foi importante analisar se o fator de transcrição XLR-1 regula a expressão do gene através da sua ligação. A análise do promotor *gsn* revelou a presença de 2 *motifs* para este fator de transcrição. As análises de interação DNA-proteína *in vitro*, realizadas com as proteínas XLR-1 inteira e truncada (XLR-1 Δ 160) utilizando fragmentos de DNA do promotor *gsn* contendo prováveis *motifs* para XLR-1 não mostraram a formação de complexos DNA-proteína. Além disso, não foi observado a formação de complexos mesmo quando fragmentos de DNA dos promotores dos genes que codificam endoxilanasas foram analisados. Os nossos resultados contradizem as análises realizadas por CALERO-NIETO et al. (2008), os quais utilizaram a proteína XlnR de *Fusarium oxysporum* produzida na forma recombinante em *E. coli* e fragmento do promotor do gene *xyl-4* codificador de uma endoxilanasase. Os autores mostraram a presença de complexos DNA-proteína utilizando tanto a forma truncada de XlnR quanto a forma inteira do fator de transcrição.

Em nosso trabalho, o fator de transcrição XLR-1 também não foi capaz de ligar fragmentos do promotor de genes codificadores de enzimas xilanolíticas. Este resultante é intrigante, pois, foi demonstrado que a expressão de tais genes foi influenciada pelo fator de transcrição. Uma possível explicação é que XLR-1 requer a presença de proteína(s) parceira(s) para exercer o papel regulatório.

A proteína XLR-1 possui 945 resíduos de aminoácidos, apresentando seu domínio de ligação ao DNA do tipo *zinc-finger* em sua região N-terminal (resíduos 90 ao 133) e um domínio específico de proteínas de fungo na região central da proteína. Já o fator de transcrição XlnR de *F. oxysporum* possui 886 resíduos de aminoácidos, e segundo os servidores SMART (<http://smart.embl-heidelberg.de/>) e PFAM (<http://pfam.sanger.ac.uk/>),

este fator de transcrição é constituído apenas por um domínio específico de fungos e localizado em sua região central (figura 26). Não foi possível identificar o domínio de interação ao DNA nesta proteína. Análises de identidade entre XLR-1 e XInR mostraram que a proteína XInR possui ao redor de 64% de identidade com a proteína XLR-1, mas o alinhamento de sequências entre estas proteínas revelou que apenas 23 dos 43 resíduos envolvidos no domínio de interação ao DNA são conservados e apenas no lado C-terminal do domínio. Possivelmente, esta sequência de resíduos de aminoácidos ausente em XLR-1 e presente em XInR pode ser a responsável pela formação do complexo DNA-proteína em *F. oxysporum*, mesmo embora não configure um domínio de ligação a DNA típico desta família de fatores de transcrição.

A ausência de ligação pode ser interpretada de várias maneiras: é possível que XLR-1 necessite de proteínas parceiras ou que o *motif* para esta proteína seja diferente do descrito em outros fungos filamentosos ou que a proteína necessite de modificações pós-traducionais que a bactéria não realiza. Entretanto, SUN et al. (2012), descreveram a sequência consenso 5'-GGCTRR-3' como o possível sítio de ligação para XLR-1, o qual é o existente nas regiões promotoras dos genes analisados em nosso trabalho. Portanto, a necessidade de proteínas parceiras é o mais provável, o que pode ser corroborado pelos resultados de expressão gênica.

2. O produto da ORF NCU04390 e regulação do metabolismo do glicogênio

O fator de transcrição codificado pela ORF NCU04390 foi inicialmente selecionado devido a alteração de acúmulo de glicogênio em choque térmico apresentado pela linhagem contendo esta ORF nocauteada. Foi observado um acúmulo exacerbado do carboidrato em condição de estresse térmico em quantidade muito similar ao apresentado em crescimento vegetativo. Além disso, foi observado na linhagem mutante na ORF NCU04390 que o nível do transcrito *gsn* sob a mesma condição continuava elevado, demonstrando que a ausência deste fator de transcrição promove um hiperacúmulo de glicogênio bem como a perda do controle transcricional de *gsn*.

Com o intuito de realizar uma caracterização funcional do fator de transcrição codificado pela ORF NCU04390 foram realizadas análises de CHIP-Seq, uma vez que esta proteína esta anotada no genoma do fungo como proteína hipotética. A técnica de CHIP vem sendo amplamente utilizada para o isolamento de fragmentos de DNA ligados por proteínas ou complexos proteicos *in vivo* (MARDIS, 2007). Ao associar esta técnica ao sequenciamento de alto desempenho (CHIP-Seq), foi possível a realização de um mapeamento ao nível genômico de sequências nucleotídicas possivelmente reconhecidas por este fator de transcrição. O formaldeído é um dos reagentes mais comuns utilizados

para a realização de reações cruzadas (*crosslink*) entre DNA e proteínas. Por ser uma molécula muito pequena, ele penetra facilmente nas células e inativa rapidamente as proteínas ao reagir com as aminas primárias presentes nas proteínas e nos ácidos nucleicos (APARÍCIO et al., 2004), permitindo a realização de uma “fotografia” das interações DNA-proteína ocorridas em determinadas condições de crescimento do organismo. Pelo fato do formaldeído promover também o *crosslink* entre proteínas, ele permite a identificação de proteínas que interagem ao DNA indiretamente, ou seja, interagindo junto a outras proteínas (KIM et al., 2006).

Análises de CHIP-Seq realizadas com este fator de transcrição mostraram que na temperatura normal de crescimento não se observou a formação de picos referentes a interações DNA-proteína em nenhum gene. Por outro lado, foram encontradas 210 proteínas possivelmente reguladas pela proteína hipotética 4390 em condição de estresse térmico. Esse fato sugere que esta proteína esteja relacionada ao controle transcricional de genes na condição de estresse térmico. Pelas análises realizadas não foi possível determinar se este fator de transcrição está envolvido na regulação de genes sob outros tipos de estresse (osmótico, pH, oxidativo). Análises futuras serão realizadas com o intuito de investigar melhor a relação deste fator de transcrição e o controle transcricional de alguns genes em tais condições de estresse. Pela figura 21 foi possível verificar que a maioria dos genes possivelmente regulados por este fator de transcrição são proteínas hipotéticas. Atualmente sabe-se a função de somente 40% das proteínas codificadas pelo genoma do fungo *N. crassa* (WANG et al., 2011). Em relação às proteínas contendo funções conhecidas foi verificado que o maior grupo das proteínas possivelmente reguladas por este fator de transcrição é o das proteínas envolvidas no metabolismo de carbono, dado de extrema importância desta análise.

Com o intuito de comprovar os resultados obtidos na análise de CHIP-Seq, análises de interação DNA-proteína *in vitro* pela técnica de EMSA foram realizadas utilizando um fragmento de 170 pb da região 5'-promotora do gene codificador da enzima desramificadora do glicogênio. Este fragmento foi escolhido devido à sua proximidade ao codon codificador da metionina inicial deste gene (-341 pb) e pela quantidade de prováveis *motifs* de DNA para esta proteína (3 ao todo). Foi observado a formação de complexo DNA-proteína utilizando 6 µg da proteína 4390 inteira produzida na forma recombinante. A especificidade da ligação ao *motif* 5'-TACATR-3' foi comprovada por meio do ensaio de competição, no qual foi utilizado o competidor específico não marcado radioativamente utilizando uma concentração molar de 10X a concentração molar da sonda. Estes resultados corroboram as análises de CHIP-Seq realizadas no laboratório do Dr. Freitag. A partir deste dado, podemos sugerir a existência de outro ponto de controle sobre o metabolismo de glicogênio até então

nunca investigado: o papel da enzima desramificadora no controle do metabolismo do glicogênio na situação de choque térmico.

Ao analisar de forma conjunta os dados obtidos nas análises de acúmulo de glicogênio e análises dos transcritos *gsn* e *gpn* tanto a 30°C quanto em condição de choque térmico, bem como os resultados obtidos pela interação DNA-proteína *in vivo* e *in vitro* podemos pressupor que o fator de transcrição codificado pela ORF NCU04390 esteja atuando como um repressor transcricional de *gsn* sob condição de estresse térmico. Além disso, é bem provável que este fator de transcrição também atue na regulação da transcrição do gene codificador da enzima desramificadora do glicogênio, o que num primeiro momento justifica o elevado acúmulo de glicogênio sob condição de estresse térmico. Uma vez que os genes codificadores destas enzimas não estejam sendo transcritos ocorre uma deficiência tanto da síntese quanto na degradação do carboidrato, justificando o acúmulo de glicogênio na condição de estresse térmico. Análises futuras serão realizadas com o intuito de verificar o que ocorre com a transcrição do gene codificador da enzima desramificadora do glicogênio bem como com o gene codificador desta proteína hipotética com o intuito de tentar elucidar o modo de participação deste fator de transcrição no metabolismo de glicogênio.

A técnica de CHIP-Seq se mostrou uma ferramenta indispensável para a identificação de possíveis alvos no DNA genômico para fatores de transcrição. A identificação sistemática de sequências nucleotídicas alvo para fatores de transcrição no genoma aliada a análises da expressão gênica em escala genômica como, por exemplo, a técnica de RNA-Seq, fornecerá informações de suma importância sobre como proteínas de diferentes classes (fatores de transcrição, polimerases, marcas epigenéticas) são responsáveis pelo controle da regulação temporal da atividade dos genes envolvidos em uma determinada via metabólica em função a determinado estímulo.

Referências Bibliográficas

- ALLIS, C. D.; JENUWEIN, T.; REINBERG, D. Overview and concepts. In: ALLIS, C. D. **Epigenetics**. New York: Cold Spring Harbor Laboratory Press, 2007. Cap. 3, p. 63-80.
- ALMER, A.; RUDOLPH, H.; HINNEN, A.; HORZ, W. Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. **EMBO J.**, v. 5, n. 10, p. 2689-2696, Oct. 1986.
- ARNOSTI, D. N.; KULKARNI, M. M. Transcriptional enhancers: intelligent enhanceosomes or flexible billboards? **J. Cell. Biochem.**, v. 94, n. 5, p. 890-898, Apr. 2005.
- ARO, N.; PAKULA, T.; PENTTILÄ, M. Transcriptional regulation of plant cell wall degradation by filamentous fungi. **FEMS Microbiol. Rev.**, v. 29, n. 4, p. 719-739, Sept. 2005.
- BAI, G.; ZHANG, J.; WERNER, R.; NUTTAL, F. Q.; TAN, A. W.; LEE, E. Y. The primary structure of rat liver glycogen synthase deduced by cDNA cloning. Absence of phosphorylation sites 1a e 1b. **J. Biol. Chem.**, v. 265, n. 14, p. 7843-7848, May 1990.
- BASKARAN, S.; CHIKWANA, V. M.; CONTRERAS, C. J.; DAVIS, K. D.; WILSON, W. A.; DEPAOLI-ROACH, A. A.; ROACH, P. J.; HURLEY, T. D. Multiple glycogen-binding sites in eukaryotic glycogen synthase are required for high catalytic efficiency toward glycogen. **J. Biol. Chem.**, v. 286, n. 39, p. 33999-34006, Aug. 2011.
- BEADLE, G. W.; TATUM, E. L. Genetic control of biochemical reactions in *Neurospora*. **Proc. Natl. Acad. Sci. USA**, v. 27, n. 11, p. 499-506, Nov. 1941.
- BELIËN, T.; VAN CAMPENHOUT, S.; ROBBEN, J.; VOLCKAERT, G. Microbial endoxylanases: effective weapons to breach the plant cell-wall barrier or, rather, triggers of plant defense systems? **Mol Plant Microbe Interact.**, v. 19, n. 10, p. 1072-1081, Oct. 2006.
- BERNSTEIN, B. E.; LIU, C. L.; HUMPHREY, E. L.; PERLSTEIN, E. O.; SCHREIBER, S. L. Global nucleosome occupancy in yeast. **Genome Biol.**, v. 5, n. 9, p. R62/1-R62/11, Aug. 2004.
- BERTOLINI, M. C.; FREITAS, F. Z.; DE PAULA, R. M.; CUPERTINO, F. B.; GONÇALVES, R. D. Glycogen metabolism regulation in *Neurospora crassa*. In: WITZANY G. **Biocommunication of fungi**. Dordrecht: Springer Science, 2012. Cap 2, p. 39-56.
- BIENZ, M.; PELHAM, H. R. Heat shock regulatory elements function as an inducible enhancer in the *Xenopus* hsp70 gene and when linked to a heterologous promoter. **Cell**, v. 45, n. 5, p. 753-760, June 1986.
- BIENZ, M.; PELHAM, H. R. Mechanisms of the heat shock gene activation in higher eukaryotes. **Adv. Genet.**, v. 24, p. 31-72, 1987.
- BISTIS, G. N.; PERKINS, D. D.; READ, N. D. Different cell types in *Neurospora crassa*. **Fungal Genet. Newsl.**, v. 50, p. 17-19, Dec. 2003.

BORKOVICH, K. A.; ALEX, L. A.; YARDEN, O.; FREITAG, M.; TURNER, G. E.; READ, N. D.; SEILER, S.; BELL-PEDERSEN, D.; PAIETTA, J.; PLESOFSKY, N.; PLAMANN, M.; GOODRICH-TANRIKULU, M.; SCHULTE, U.; MANNHAUPT, G.; NARGANG, F. E.; RADFORD, A.; SELITRENNIKOFF, C.; GALAGAN, J. E.; DUNLAP, J. C.; LOROS, J. J.; CATCHESIDE, D.; INOUE, H.; ARAMAYO, R.; POLYMENIS, M.; SELKER, E. U.; SACHS, M. S.; MARZLUF, G. A.; PAULSEN, I.; DAVIS, R.; EBBOLE, D. J.; ZELTER, A.; KALKMAN, E. R.; O'ROURKE, R.; BOWRING, F.; YEADON, J.; ISHII, C.; SUZUKI, K.; SAKAI, W.; PRATT, R. Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. **Microbiol. Mol. Biol. Rev.**, v. 1, n. 68, p. 1-108, Mar. 2004.

BRANDEN, C.; TOOZE, J. **Introduction to protein structure**. 2nd ed. New York: Garland Publishing, 1999. 410 p.

CALERO NIETO, F.; HERA, C.; DI PIETRO, A.; OREJAS, M.; RONCERO, M. I. Regulatory elements mediating expression of xylanase genes in *Fusarium oxysporum*. **Fungal Genet Biol.**, v. 45, n. 1, p. 28-34, June 2008.

COLLINS, T.; GERDAY, C.; FELLER, G. Xylanases, xylanases families and extremophilic xylanases. **FEMS Microbiol. Rev.**, v. 29, n. 1, p. 3-23, Jan. 2005.

COLOT, H. V.; PARK, G.; TURNER, G. E.; RINGELBERG, C.; CREW, C. M.; LITVINKOVA, L.; WEISS, R. L.; BORKOVICH, K. A.; DUNLAP, J. C. A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. **Proc. Natl. Acad. Sci. U S A**, v. 103, n. 27, p. 10352-10357, June 2006.

DAVIS, R. H. **Neurospora**: contributions of a model organism. New York: Oxford University Press, 2000. 333 p.

DAVIS, R. H.; DE SERRES, F. J. Genetic and microbiological research technique for *Neurospora crassa*. In: HERBERT, T.; TABOR, C. W. **Methods in enzymology**. New York : Academic Press, 1970. v. 17. Cap. 4, p. 79-148.

DE PAULA, R. M.; DE PINHO, C. A.; TEREZI, H. F.; BERTOLINI, M. C. Molecular and biochemical characterization of *Neurospora crassa* glycogen synthase encoded by gsn cDNA. **Mol. Genet. Genom.**, v. 267, n. 2, p. 241-253, Mar. 2002.

DE PAULA, R. M.; WILSON, W. A.; ROACH, P. J.; TEREZI, H. F.; BERTOLINI, M. C. Biochemical characterization of *Neurospora crassa* glycogenin (GNN), the self-glucosylating initiator of glycogen synthesis. **FEBS Lett.**, v. 579, n. 10, p. 2208-2214, Apr. 2005a.

DE PAULA, R. M.; WILSON, W. A.; TEREZI, H. F.; ROACH, P. J.; BERTOLINI, M. C. GNN is a self-glucosylating protein involved in the initiation step of glycogen biosynthesis in *Neurospora crassa*. **Arch. Biochem. Biophys.**, v. 435, n. 1, p. 112-124, Mar. 2005b.

DE VRIES, R. P.; VISSER, J. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. **Microbiol. Mol. Biol. Rev.**, v.65, n. 4, p. 497-522, Dec. 2001.

DE VRIES, R. P.; POULSEN, C. H.; MADRID, S.; VISSER, J. aguA, the gene encoding an extracellular alpha-glucuronidase from *Aspergillus tubingensis*, is specifically induced on xylose and not on glucuronic acid. **J. Bacteriol.**, v. 180, n. 2, p. 243-249, Jan. 1998.

DUNLAP, J. C.; BORKOVICH, K. A.; HENN, M. R.; TURNER, G. E.; SACHS, M. S.; GLASS, N. L.; McCLUSKEY, K.; PLAMANN, M.; GALAGAN, J. E.; BIRREN, B. W.; WEISS, R. L.; TOWNSEND, J. P.; LOROS, J. J.; NELSON, M. A.; LAMBREGHTS, R.; COLOT, H. V.; PARK, G.; COLLOPY, P.; RINGELBERG, C.; CREW, C.; LITVINKOVA, L.; DECAPRIO, D.; HOOD, H. M.; CURILLA, S.; SHI, M.; CEAWFORD, M.; KOERHSEN, M.; MONTGOMERY, P.; LARSON, L.; PEARSON, M.; KASUGA, T.; TIAN, C.; BASTÜRKMEN, M.; ALTAMIANO, L.; XU, J. Enabling a community to dissect an organism: overview of the *Neurospora* functional genomics project. **Adv. Genet.**, v. 57, n. 1, p. 49-96, 2007.

ESSER, K.; KUENEN, R. **Genetics of fungi**. Heidelberg: Springer, 1967. 490 p.

FEDURCO, M.; ROMIEU, A.; WILLIAMS, S.; LAWRENCE, I.; TURCATTI, G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. **Nucleic Acids Res.**, v. 34, n. 3, p. e22/1-e22/13, Feb. 2006.

FLETTERICK, R. J.; MADSEN, N. B. The structure and related functions of phosphorylase A. **A. Rev. Biochem.**, v. 49, n. 1, p. 31-61, July 1980.

FREITAS, F. Z.; BERTOLINI, M. C. Genomic organization of the *Neurospora crassa* *gsn* gene: possible involvement of the STRE and HSE elements in the modulation of transcription during heat shock. **Mol. Gen. Genomics**, v. 272, n. 5, p. 550-561, Nov. 2004.

FREITAS, F. Z.; CHAPEAUROUGE, A.; PERALES, J.; BERTOLINI, M. C. A systematic approach to identify STRE-binding proteins of the *gsn* glycogen synthase gene promoter in *Neurospora crassa*. **Proteomics**, v. 8, n. 10, p. 2052-2061, May 2008.

FREITAS, F. Z.; DE PAULA, R. M.; BARBOSA, L. C.; TEREZI, H. F.; BERTOLINI, M. C. cAMP signaling pathway controls glycogen metabolism in *Neurospora crassa* by regulating the glycogen synthase gene expression and phosphorylation. **Fungal Genet. Biol.**, v. 47, n. 1, p. 43-52, Jan. 2010.

GALAGAN, J. E.; CALVO, S. E.; BORKOVICH, K. A.; SELKER, E. U.; READ, N. D.; JAFFE, D.; FITZHUGH, W.; MA, L. J.; SMIRNOV, S.; PURCELL, S.; REHMAN, B.; ELKINS, T.; ENGELS, R.; WANG, S.; NIELSEN, C. B.; BUTLER, J.; ENDRIZZI, M.; QUI, D.; IANAKIEV, P.; BELL-PEDERSEN, D.; NELSON, M. A.; WERNER-WASHBURNE, M.; SELITRENNIKOFF, C. P.; KINSEY, J. A.; BRAUN, E. L.; ZELTER, A.; SCHULTE, U.; KOTHE, G. O.; JEDD, G.; MEWES, W.; STABEN, C.; MARCOTTE, E.; GREENBERG, D.; ROY, A.; FOLEY, K.; NAYLOR, J.; STANGE-THOMANN, N.; BARRETT, R.; GNERRE, S.; KAMAL, M.; KAMVYSSELIS, M.; MAUCELI, E.; BIELKE, C.; RUDD, S.; FRISHMAN, D.; KRYSTOFOVA, S.; RASMUSSEN, C.; METZENBERG, R. L.; PERKINS, D. D.; KROKEN, S.; COGONI, C.; MACINO, G.; CATCHESIDE, D.; LI, W.; PRATT, R. J.; OSMANI, S. A.; DESOUZA, C. P.; GLASS, L.; ORBACH, M. J.; BERGLUND, J. A.; VOELKER, R.; YARDEN, O.; PLAMANN, M.; SEILER, S.; DUNLAP, J.; RADFORD, A.; ARAMAYO, R.; NATVIG, D. O.; ALEX, L. A.; MANNHAUPT, G.; EBBOLE, D. J.; FREITAG, M.; PAULSEN, I.; SACHS, M. S.; LANDER, E. S.; NUSBAUM, C.; BIRREN, B. The genome sequence of the filamentous fungus *Neurospora crassa*. **Nature**, v. 422, n. 6934, p. 859-868, Apr. 2003.

GIBBONS, B. J.; ROACH, P. J.; HURLEY, T. D. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. **J. Mol. Biol.**, v. 2, n. 319, p. 463-477, May 2002.

GIELKENS, M. M. C.; DEKKERS, E.; VISSER, J.; DE GRAAFF, L. H. Two cellobiohydrolase-encoding genes from *Aspergillus niger* require D-xylose and the xylanolytic transcriptional activator XlnR for their expression. **Appl. Environ. Microbiol.**, v. 65, n. 10, p. 4340-4345, Oct. 1999.

- GIVAN, A. L. Flow cytometry: an introduction. In: HAWLEY, T. S.; ROBERT, G. H. **Methods in molecular biology**. Totowa: Humana Press, 2004. p.1-32.
- GONÇALVES, R. D.; CUPERTINO, F. B.; FREITAS, F. Z.; LUCHESSI, A. D.; BERTOLINI, M. C. A genome-wide screen for *Neurospora crassa* transcription factors regulating glycogen metabolism. **Mol. Cell. Proteomics.**, v.10, n. 11, p. M1111.007963/1- M1111.007963/13, July 2011.
- HARDING, R. W.; TURNER, R. V. Photoregulation of the carotenoid biosynthetic pathway in albino and white collar mutants of *Neurospora crassa*. **Plant Physiol.**, v. 68, n. 3, p. 745-749, Sept. 1981.
- HARDY, T. A.; ROACH, P. J. Control of yeast glycogen synthase-2 by COOH-terminal phosphorylation. **J. Biol. Chem.**, v. 268, n. 32, p. 23799-27805, Nov. 1993.
- HARRIS, R. A. Carbohydrate metabolism I: major metabolic pathways and their control. In: DEVLIN, T. M. **Textbook of biochemistry with clinical correlations**. 4th ed. New York: Wiley-Liss, 1997. p. 267-331.
- HARTREE, E. F. Determination of protein: a modification of the Lowry method that gives a linear photometric response. **Anal. Biochem.**, v. 48, n. 2, p. 422-427, Aug, 1972.
- HASPER, A. A.; DEKKERS, E.; VAN MIL, M.; VAN DE VONDERVOORT, P. J. I.; DE GRAAFF, L.H. EglC, a new endoglucanase from *Aspergillus niger* with major activity towards xyloglucan. **Appl. Environ. Microbiol.**, v. 68, n. 4, p. 1556-1560, Apr. 2002.
- HOLMES, D. S.; QUIGLEY, M. A rapid boiling method for the preparation of bacterial plasmids. **Anal. Biochem.**, v. 114, n. 1, p. 193-197, June 1981.
- HONDA, S.; SELKER, E. U. Tools for fungal proteomics: multifunctional *Neurospora* vectors for gene replacement, protein expression and protein purification. **Genetics**, v. 182, n. 1, p. 11-23, Jan, 2009.
- HYNES, M. J.; MURRAY, S. L.; DUNCAN, A.; KHEW, G. S.; DAVIES, M. A. Regulatory genes controlling fatty acid catabolism and peroxisomal functions in the filamentous fungus *Aspergillus nidulans*. **Eukaryot. Cell.**, v. 5, n. 5, p. 794-805, May 2006.
- KANG, S.; METZENBERG, R. L. Molecular analysis of nuc-1+, a gene controlling phosphorus acquisition in *Neurospora crassa*. **Mol Cell Biol.**, v. 10, n. 11, p. 5839-5848, Nov. 1990
- KIM, T. H.; REN, B. Genome-wide analysis of protein–DNA interactions. **Annu. Rev. Genomics Hum. Genet.**, v. 7, n. 1, p. 81-102, Sept. 2006.
- KULMBURG, P.; MATHIEU, M.; DOWZER, C.; KELLY, J.; FELENBOK, B. Specific binding sites in the alcR and alcA promoters of the ethanol regulon for the CREA repressor mediating carbon catabolite repression in *Aspergillus nidulans*. **Mol. Microbiol.**, v. 7, n. 6, p. 847-857, Mar. 1993.
- KUMAR, S.; RAMÓN, D. Purification and regulation of the synthesis of a β -xylosidase from *Aspergillus nidulans*. **FEMS Microbiol. Lett.**, v. 135, n. 1, p. 287-293, Nov. 1995.
- LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680-685, Aug. 1970.

- LANDSCHULZ, W. H.; JOHNSON, P. F.; McKNIGHT, S.L. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. **Science**, v. 240, n. 4860, p. 1759-1764, June 1988.
- LATCHMAN, D. S. Transcription factors: an overview. **Int J. Biochem. Cell. Biol.**, v. 29, n. 12, p. 1305-1312. Dec. 1997.
- LATCHMAN, D. S. **Gene regulation**: a eukaryotic perspective. 5th ed. New York: Taylor and Francis, 2005. 374 p.
- LATCHMAN, D. S. **Eukaryotic transcription factors**. 5th ed. New York: Elsevier, 2007. 488 p.
- LEE, C. K.; SHIBATA, Y.; RAO, B.; STRAHL, B. D.; LIEB, J. D. Evidence for nucleosome depletion at active regulatory regions genome-wide. **Nat. Genet.**, v.36, n. 8, p. 900-905, Aug. 2004.
- LEVINE, M.; TJIAN, R. Transcription regulation and animal diversity. **Nature**, v. 424, n. 6945, p. 147-151, July 2003.
- LINDEN, H.; MACINO, G. White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in *Neurospora crassa*. **EMBO J.**, v. 16, n. 1, p. 98-109, Jan. 1997.
- LUGER, K.; MÄDER, A. W.; RICHMOND, R. K.; SARGENT, D. F.; RICHMOND, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. **Nature**, v. 389, n. 6648, p. 251-260, Sept. 1997.
- LUSCOMBE, N. M.; AUSTIN, S. E.; BERMAN, H. M.; THORNTON, J. M. An overview of the structures of protein-DNA complexes. **Genome Biol.**, v. 1, n. 1, p. REVIEWS001, June. 2000.
- MAHESHWARI, R. Microconidia of *Neurospora crassa*. **Fungal Genet. Biol.**, v. 26, n. 1, p. 1-18, Apr. 1999.
- MARDIS, E. R. ChIP-seq: welcome to the new frontier. **Nat. Methods**, v. 4, n. 8, p. 613-614, Aug. 2007.
- MARDIS, E. R. Next-generation DNA sequencing methods. **Annu. Rev. Genomics. Hum. Genet.**, v. 1, p. 387-402, Sept. 2008.
- MARTIN, C.; ZHANG, Y. The diverse functions of histone lysine methylation. **Nat. Rev. Mol. Cell. Biol.**, v. 6, n. 11, p. 838-849, Nov. 2005.
- MARTINEZ-PASTOR, M.; MARCHLER, G.; SCHULLER, C.; MARCHLER, B. A.; RUIS, H.; ESTRUCH, F. The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). **EMBO J.**, v. 15, n. 9, p. 2227-2235, May 1996.
- MARUI, J.; TANAKA, A.; MIMURA, S.; DE GRAAFF, L. H.; VISSER, J.; KITAMOTO, N.; KATO, M.; KOBAYASHI, T.; TSUKAGOSHI, N. A transcriptional activator, AoXlnR, controls the expression of genes encoding xylanolytic enzymes in *Aspergillus oryzae*. **Fungal Genet. Biol.**, v. 35, n. 2, p. 157-169, Mar. 2002.

McGHEE, J. D.; FELSENFELD, G. Nucleosome structure. **Annu. Rev. Biochem.**, v. 49, n. 1, p. 1115-1156, July 1980.

MELÉNDEZ, R.; MELÉNDEZ-HEVIA, E.; CASCANTE, M. How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building. **J. Mol. Evol.**, v. 45, n. 4, p. 446-455, Oct. 1997.

MONTAGNE, C. Quatrième centurie de plantes cellulaires exotiques nouvelles. **Ann. Sci. Nat. Bot.**, v. 20, n. 2, p. 352-379, 1843.

MU, J.; CHENG, C.; ROACH, P. J. Initiation of glycogen synthesis in yeast – requirement of multiple tyrosine residues for function of the self-glucosylating Glg proteins *in vivo*. **J. Biol. Chem.**, v. 271, n. 43, p. 26554-26560, Oct. 1996.

MURRE, C.; BAIN, G.; VAN DIJK, M. A.; ENGEL, I.; FURNARI, B. A.; MASSARI, M. E.; MATTHEWS, J. R.; QUONG, M. W.; RIVERA, R. R.; STUIVER, M. H. Structure and function of helix-loop-helix proteins. **Biochim. Biophys. Acta**, v. 1218, n. 2, p. 129-135, June 1994.

NI, H. T.; LAPORTE, D. C. Response of a yeast glycogen synthase gene to stress. **Mol. Microbiol.**, v. 16, n. 6, p. 1197-1205, June 1995.

NINOMIYA, Y.; SUZUKI, K.; ISHII, C.; INOUE, H. HIGHLY efficient gene replacements in *Neurospora* strains deficient for nonhomologous end-joining. **Proc. Natl. Acad. Sci. U S A**, v. 101, n. 33, p. 12248-12253, Aug. 2006.

NOVENTA-JORDÃO, M. A.; DE LOURDES, M.; POLIZELI, T. M.; BONINI, B. M.; JORGE, J. A.; TERENCEZI, H. F. Effects of temperature shifts on the activities of *Neurospora crassa* glycogen synthase, glycogen phosphorylase and trehalose-6-phosphate synthase. **FEBS Lett.**, v. 378, n. 1, p. 32-36, Jan. 1996.

OMICIANSKI, J. G.; CLORE, G. M.; SCHAAD, O.; FELSENFELD, G.; TRAINOR, C.; APPELLA, E.; STAHL, S. J.; GRONENBORN, A. M. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. **Science**, v. 261, n. 5120, p. 438-446, July 1993.

PAIETTA, J. V.; MARZLUF, G. A. Gene disruption by transformation in *Neurospora crassa*. **Mol Cell Biol.**, v. 5, n. 7, p. 1554-1559, July 1985.

PAYEN, A. Extrait d'un rapport adresse' a' M. Le Marechal Duc de Dalmatie, Ministre de la Guerre, President du Conseil, sur une alteration extraordinaire du pain du muniton. **Ann. Chim. Phys.**, v. 9, n. 1, p. 5-21, 1843.

PELHAM, H. B. R.; BIENZ, M. A synthetic heat shock promoter element confers heat inducibility on the herpes simplex virus *tk* gene. **EMBO J.**, v. 11, n. 1, p. 1473-1477, Oct. 1992.

PÉREZ-GONZÁLEZ, J. A.; VAN PEIJ, N. N.; BEZOEN, A.; MacCABE, A. P.; RAMÓN, D.; DE GRAAFF, L. H. Molecular cloning and transcriptional regulation of the *Aspergillus nidulans* xlnD gene encoding a beta-xylosidase. **Appl. Environ. Microbiol.**, v. 64, n. 4, p. 1412-1419, Apr. 1998.

PERKINS, D. D.; DAVIS, R. H. *Neurospora* at the millennium. **Fungal Genet. Biol.**, v. 31, n. 3, p. 153-167, Dec. 2000.

- POLIZELI, M. L.; RIZZATTI, A. C.; MONTI, R.; TEREZI, H. F.; JORGE, J. A.; AMORIM, D. S. Xylanases from fungi: properties and industrial applications. **Appl. Microbiol. Biotechnol.**, v. 67, n. 5, p. 577-591, June 2005.
- POMRANING, K. R.; SMITH, K. M.; FREITAG, M. Genome-wide high throughput analysis of DNA methylation in eukaryotes. **Methods**, v. 47, n. 3, p. 142-150, Mar. 2009.
- RAJU, N. B. Meiosis and ascospore in *Neurospora*. **Eur. J. Cell Biol.**, v. 283, n. 1, p. 208-223, July 1980.
- RAUSCHER, R.; WÜRLEITNER, E.; WACENOVSKY, C.; ARO, N.; STRICKER, A. R.; ZEILINGER, S.; KUBICEK, C. P.; PENTTILÄ, M.; MACH, R. L. Transcriptional regulation of *xyn1*, encoding xylanase I, in *Hypocrea jecorina*. **Eukaryot. Cell**, v. 5, n. 3, p. 447-456, Mar. 2006.
- ROACH, P. J. Control of glycogen synthase by hierarchal protein phosphorylation. **FASEB J.**, v. 4, n. 12, p. 2961-2968, Sept. 1990.
- ROACH, P. J.; SKURAT, A. V. Self-glucosylating initiator proteins and their role in glycogen biosynthesis. In: MOLDAVE, K. **Progress in nucleic acid research**. San Diego: Academic Press, 1997. v. 57, p. 289-316.
- ROSELL-PEREZ, M.; VILLAR-PALASI, C.; LARNER, J. Studies on UDPG-glycogen transglucosylase. I. Preparation and differentiation of two activities of UDPG-glycogen transglucosylase from rat skeletal muscle. **Biochemistry**, v. 1, n. 1, p. 763-768, Sept. 1962.
- RUIJTER, G. J. G.; VANHANEN, S. A.; GIELKENS, M. M.; VAN DE VONDERVOORT, P. J.; VISSER, J. Isolation of *Aspergillus niger creA* mutants: effects on expression of arabinases and L-arabinose catabolic enzymes. **Microbiology**, v. 1, n. 143, p. 2991-2998, Sept. 1997.
- SAMBROOK, J.; RUSSELL, D. W. **Molecular cloning, a laboratory manual**. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 2001. p. 2344.
- SCHJERLING, P.; HOLMBERG, S. Comparative amino acid sequence analysis of the C6 zinc cluster family of transcriptional regulators. **Nucleic Acids Res.**, v. 24, n. 23, p. 4599-607, Dec. 1996.
- SCHULTE, U.; BECKER, I.; MEWES, H. W.; MANNHAUPT, G. Large scale analysis of sequences from *Neurospora crassa*. **J. Biotechnol.**, v. 14, n. 1, p. 3-13, Mar, 2002.
- SEKINGER, E. A.; MOQTADERI, Z.; STRUHL, K. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. **Mol. Cell.**, v. 18, n. 6, p. 735-748, June 2005.
- SHEAR, C. L.; DODGE, B. O. Life histories and heterothallism of the red bread-mold fungi of the *Monilia sitophila* group. **J. Agric. Res.**, v. 34, n. 1, p. 1019-1042, June 1927.
- SIKORSKI, R. S.; HIETER, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. **Genetics**, v. 122, n. 1, p. 19-27, May 1989.
- SOKOLOVSKY, V.; KALDENHOFF, R.; RICCI, M.; RUSSO, V. E. A. Fast and reliable mini-prep RNA extraction from *Neurospora crassa*. **Fungal Genet. Newsl.**, v. 37, p. 41-43, Feb. 1990.

- SØRENSEN, H. R.; MEYER, A. S.; PEDERSEN, S. Enzymatic hydrolysis of water-soluble wheat arabinoxylan. 1. Synergy between alpha-L-arabinofuranosidases, endo-1,4-beta-xylanases, and beta-xylosidase activities. **Biotechnol. Bioeng.**, v. 81, n. 6, p. 726-731, Mar. 2003.
- SPRINGER, M. L.; YANOFSKY, C. A morphological and genetic analysis of conidiophore development in *Neurospora crassa*. **Genes Devel.**, v. 3, n. 4, p. 559-571, Apr. 1989.
- STARICH, M. R.; WIKSTRÖM, M.; ARST, H. N. Jr.; CLORE, G. M.; GRONENBORN, A. M. The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. **J. Mol. Biol.**, v. 277, n. 3, p. 605-620, Apr. 1998.
- STRICKER, A. R.; MACH, R. L.; DE GRAAFF, L. H. Regulation of transcription of cellulases- and hemicellulases-encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*). **Appl. Microbiol. Biotechnol.**, v. 2, n. 78, p. 211-220, Feb. 2008.
- SUN, J.; TIAN, C.; DIAMOND, S.; GLASS, N. L. Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in *Neurospora crassa*. **Eukaryot. Cell**, v. 11, n. 4, p. 482-493, Feb. 2012.
- TAMAYO, E. N.; VILLANUEVA, A.; HASPER, A. A.; DE GRAAFF, L. H.; RAMÓN, D.; OREJAS, M. CreA mediates repression of the regulatory gene xlnR which controls the production of xylanolytic enzymes in *Aspergillus nidulans*. **Fungal Genet. Biol.**, v. 45, n. 6, p. 984-93, June 2008.
- TÉLLEZ-IÑÓN, M. T.; TERENZI, H.; TORRES, H. N. Interconvertible forms of glycogen synthetase in *Neurospora crassa*. **Biochim. Biophys. Acta**, v. 191, n. 3, 765-768, July 1969.
- THEVELEIN, J. M. Regulation of trehalose mobilization in fungi. **Microbiol. Reviews**, v. 48, p. 42-59, Mar. 1984.
- TIAN, C.; LI, J.; GLASS, N. L. Exploring the bZIP transcription factor regulatory network in *Neurospora crassa*. **Microbiology**, v. 157, n. 3, p. 747-759, Mar. 2011.
- TIMELL, T. E. Recent progress in the chemistry of wood hemicelluloses. **Wood Sci. Technol.**, v. 1, n. 1, p. 45-70, Mar. 1967.
- TODD, R. B.; ANDRIANOPOULOS, A. Evolution of a fungal regulatory gene family: the Zn(II)₂Cys₆ binuclear cluster DNA binding motif. **Fungal Genet. Biol.**, v. 21, n. 3, p. 388-405, June 1997.
- TURCATTI, G.; ROMIEU, A.; FEDURCO, M.; TAIRI, A. P. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. **Nucleic Acids Res.**, v. 36, n. 4, p. e25, Mar. 2008.
- VAN PEIJ, N. N.; VISSER, J.; DE GRAAFF, L. H. Isolation and analysis of xlnR, encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*. **Mol. Microbiol.**, v. 27, n. 1, p. 131-142, Jan. 1998.
- VAN PEIJ, N. N.; GIELKENS, M. M. C.; DE VRIES, R. P.; VISSER, J.; DE GRAAFF, L. H. The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. **Appl. Environ. Microbiol.** v. 64, n. 10, p. 3615-3619. Oct. 1998.

- VAQUERO, A.; LOYOLA, A.; REINBERG, D. The constantly changing face of chromatin. **Sci. Aging Knowledge Environ.**, v. 9, n. 14, RE4, Apr. 2003.
- VARELA, J. C.; PRAEKELT, U. M.; MEACOCK, P. A.; PLANTA, R. J.; MAGER, W. H. The *Saccharomyces cerevisiae* HSP12 gene is activated by the high-osmolarity glycerol pathway and negatively regulated by protein kinase A. **Mol. Cell. Biol.**, v. 15, n. 11, p. 6232-6245, Nov. 1995.
- VASHEE, S.; XU, H.; JOHNSTON, S. A.; KODADEK, T. How do "Zn2cys6" proteins distinguish between similar upstream activation sites? Comparison of the DNA-binding specificity of the GAL4 protein *in vitro* and *in vivo*. **J. Biol. Chem.**, v. 268, n. 33, p. 24699-24706, Nov. 1993.
- VINSON, C. R.; HAI, T.; BOYD, S. M. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. **Genes Dev.**, v. 7, n. 7, p. 1047-1058, June 1993.
- VOGEL, H. J. A convenient growth medium for *Neurospora crassa* (medium N). **Microbiol. Genet. Bulletin**, v. 13, p. 42-43, Feb. 1956.
- WANG, T. Y.; HE, F.; HU, Q. W.; ZHANG, Z. A predicted protein-protein interaction network of the filamentous fungus *Neurospora crassa*. **Mol. Biosyst.**, v. 7, n. 7, p. 2278-2285, July 2011.
- WASSERMAN, W.; FICKETT, J. Identification of regulatory regions which confer muscle-specific gene expression. **J. Mol. Biol.**, v. 278, p. 167-181, Apr. 1998.
- WASYLYK, B.; CHAMBON, P. Transcription by eukaryotic RNA polymerases A and B of chromatin assembled *in vitro*. **Eur. J. Biochem.**, v. 98, n. 2, p. 317-327, Aug. 1979.
- WESTERGAARD, M.; MITCHELL, K. *Neurospora* V.A. synthetic medium favoring sexual reproducing. **Am. J. Bot.**, v. 34, n. 10, p. 573-577, Dec. 1947.
- WILKIE, K. C. B. The hemicelluloses of grasses and cereals. **Adv. Carbohydr. Chem. Biochem.**, v. 36, n. 1, p. 215-264, Apr. 1979.
- WOLFFE, A. P.; HAYES, J. J. Chromatin disruption and modification. **Nucleic Acids Res.**, v. 27, n. 3, p. 711-720, Feb. 1999.

Anexo I

A Genome-wide Screen for *Neurospora crassa* Transcription Factors Regulating Glycogen Metabolism*

Rodrigo Duarte Gonçalves‡, Fernanda Barbosa Cupertino‡, Fernanda Zanolli Freitas‡, Augusto Ducati Luchessi§, and Maria Célia Bertolini‡¶

Transcription factors play a key role in transcription regulation as they recognize and directly bind to defined sites in promoter regions of target genes, and thus modulate differential expression. The overall process is extremely dynamic, as they have to move through the nucleus and transiently bind to chromatin in order to regulate gene transcription. To identify transcription factors that affect glycogen accumulation in *Neurospora crassa*, we performed a systematic screen of a deletion strains set generated by the *Neurospora* Knockout Project and available at the Fungal Genetics Stock Center. In a wild-type strain of *N. crassa*, glycogen content reaches a maximal level at the end of the exponential growth phase, but upon heat stress the glycogen content rapidly drops. The gene encoding glycogen synthase (*gsn*) is transcriptionally down-regulated when the mycelium is exposed to the same stress condition. We identified 17 deleted strains having glycogen accumulation profiles different from that of the wild-type strain under both normal growth and heat stress conditions. Most of the transcription factors identified were annotated as hypothetical protein, however some of them, such as the PacC, XlnR, and NIT2 proteins, were biochemically well-characterized either in *N. crassa* or in other fungi. The identification of some of the transcription factors was coincident with the presence of DNA-binding motifs specific for the transcription factors in the *gsn* 5'-flanking region, and some of these DNA-binding motifs were demonstrated to be functional by Electrophoretic Mobility Shift Assay (EMSA) experiments. Strains knocked-out in these transcription factors presented impairment in the regulation of *gsn* expression, suggesting that the transcription factors regulate glycogen accumulation by directly regulating *gsn* gene expression. Five selected mutant strains showed defects in cell cycle progression, and two transcription factors were light-regulated. The results indicate that there are connections linking different cellular processes, such as metabolism control, biological clock, and

cell cycle progression. *Molecular & Cellular Proteomics* 10: 10.1074/mcp.M111.007963, 1–13, 2011.

The fungus *Neurospora crassa* has been widely used as a model organism for the understanding of fundamental aspects of eukaryotic biology. The knowledge of its genome sequence (1) has allowed the identification of proteins required for gene regulation, such as the transcriptional regulatory proteins. An examination of the classes of transcription factors in the *N. crassa* genome reveals that the organism carries elements shared by simple and complex metazoan models (2). The availability of a set of deletion strains, each carrying a deletion in a specific ORF encoding a transcription factor, allows the screening for genes linked to a particular phenotype. Here we used this mutant strains set to identify transcription factors that either directly or indirectly regulate glycogen metabolism in *N. crassa*.

In many organisms, glycogen is a carbon and energy reserve carbohydrate with an intricate metabolism regulation that senses nutrient availability and other environmental conditions. The amount of glycogen found in a particular situation results from the balance between glycogen synthase and glycogen phosphorylase activities. These enzymes regulate, respectively, the synthesis and degradation of this compound and they are both regulated by phosphorylation. Besides reversible changes in their activities, glycogen levels are also correlated with physiological conditions. In addition, other proteins may also be involved in glycogen accumulation because protein activation resulting from different signaling pathways affects glycogen storage (3, 4).

In *N. crassa*, glycogen content reaches a maximal level at the end of the exponential growth phase. However, under stress conditions, such as heat shock, glycogen content drops rapidly (5, 6). The yeast *Saccharomyces cerevisiae* accumulates glycogen under heat shock (7), demonstrating that yeast and *N. crassa* show opposite responses concerning this environmental condition. The glycogen decrease observed in *N. crassa* might result from the regulation, at transcriptional level, of enzymes involved in the carbohydrate metabolism, as transcription of the gene encoding glycogen synthase (*gsn*) decreases under heat stress (5, 8). The *gsn* promoter has one *cis*-acting STRE DNA

From the Addresses: ‡Instituto de Química, UNESP, Departamento de Bioquímica e Tecnologia Química, 14800-900, Araraquara, SP, Brazil; §Faculdade de Ciências Aplicadas, Universidade Estadual de Campinas, Rua Pedro Zaccaria, 1300, 13484-350, Limeira, SP, Brazil
Received January 17, 2011, and in revised form, July 15, 2011

Published, MCP Papers in Press, July 18, 2011, DOI 10.1074/mcp.M111.007963

TABLE I
Oligonucleotides used in this study

Primer	Sequence ^a	Source	Name	Position (nt)
90-F	5'-CATATGTCGTCACACACAGCCAG-3'	ORF NCU00090	–	1 to 21
90-R2	5'-GGATCCTTACTTGTGAAGTGGAGCCTG-3'	ORF NCU00090	–	639 to 622
PacC-F	5'-GACCCAACAGCCCAACTT-3'	<i>gsn</i> promoter	<i>pacC</i> probe	–1918 to –1901
GSN-RP3	5'-GCAACGAATACTCCCATG-3'	<i>gsn</i> promoter	<i>pacC</i> probe	–1789 to –1806
GSN-FP4	5'-CTGATTGGGAAAGGTCAGA-3'	<i>gsn</i> promoter	<i>nit2</i> probe	–1645 to –1626
GSN-RP2	5'-CTGTTGACCTGCGTTAAC-3'	<i>gsn</i> promoter	<i>nit2</i> probe	–1269 to –1286
XLNR-F2	5'-TGAGGGTGAGAAAGTTGC-3'	<i>gsn</i> promoter	<i>xlnR</i> probe	–2173 to –2156
XLNR-R2	5'-TATTCTGCAACGGAATCC-3'	<i>gsn</i> promoter	<i>xlnR</i> probe	–2034 to –2053

^a *Nde*I and *Bam*HI restriction sites are underlined in the sequences. Positions are according to the ATG start codon of translation. Stop codon inserted in the ORF sequence is shown in bold.

motif, which is specifically bound by nuclear proteins activated under heat shock. In *S. cerevisiae*, STRE is recognized by two transcription factors, the zinc finger proteins Msn2p and Msn4p (Msn2/4p), which mediate the cellular response to multiple stresses and are components of the environmental stress response (9). We have previously combined biochemical techniques and a proteomic approach coupled to mass spectrometry in an attempt to identify *N. crassa* proteins that are activated upon heat shock and bind to the STRE motif of the *gsn* promoter (10). Only hypothetical proteins having domains that might be involved in transcription regulation were identified, and none of them had a DNA-binding domain.

To identify transcription factors regulating glycogen metabolism in the fungus *N. crassa*, we used a mutant strains set with single-gene deletions of known or putative transcription factors to search for mutant strains having glycogen accumulation profiles different from that in the wild-type strain. The mutant strains were analyzed under normal growth temperature (30 °C) and under heat shock stress (45 °C). The results described in this work showed that most of the transcription factors identified have been annotated in the *N. crassa* database (<http://www.broad.mit.edu/annotation/genome/neurospora/Home.html>) as hypothetical proteins. However, many are proteins that have been functionally characterized, either in *N. crassa* or in other fungi. This indicates that glycogen metabolism regulation in eukaryotic cells comprises a complex regulatory network involving metabolic and nutrient sensing, which under certain circumstances could lead to impairment of cellular development.

EXPERIMENTAL PROCEDURES

Neurospora crassa Strains and Growth Conditions—The *N. crassa* strain FGSC 9718 (*mus-51::bar mat a*), and a set of 147 mutant strains individually knocked-out in genes encoding transcription factors were purchased from the Fungal Genetics Stock Center (FGSC, University of Missouri, Kansas City, Missouri, <http://www.fgsc.net>) (11). The deletion strains comprise a set of mutants where each open reading frame (ORF)¹ has been disrupted from start to stop codon by the insertion of the *hph* gene (hygromycin B phosphotransferase) as a marker (12).

¹ The abbreviations used are: ORF, open reading frame; SREBP, sterol regulatory element binding protein; STRE, Stress Response Element; PMSF, phenylmethylsulfonyl fluoride.

The strains were cultivated in Vogel's minimal medium (13) supplemented with 2% sucrose. After 10 days of culture, conidia were suspended in sterile water and counted. For the heat shock experiments, conidia (10⁷/ml) were first germinated in 60 ml of Vogel's minimal medium supplemented with 2% sucrose, at 30 °C and 250 rpm during 24 h. After this time an aliquot was removed, filtered, frozen in liquid nitrogen and stored at –80 °C until use. The remaining culture was filtered and transferred into fresh Vogel's medium preheated at 45 °C. After 30 min, the mycelia were harvested by filtration, frozen in liquid nitrogen, and stored at –80 °C.

Glycogen and Protein Quantification—Mycelia pads were ground to a fine powder in a prechilled mortar in liquid nitrogen, and extracted into lysis buffer (50 mM Tris-HCl, pH 7.6, 100 mM NaF, 1 mM EDTA, 1 mM PMSF, 0.1 mM tosyl-L-lysine chloromethyl ketone, 1 mM benzamidin, and 1 μg/ml of each pepstatin and aprotinin). Cellular extracts were clarified by centrifugation at 3,000 × *g*, for 10 min at 4 °C, and the supernatants were used for glycogen and protein quantifications. Glycogen content was measured following the protocol described by Hardy and Roach (14), with slight modifications. Briefly, 100 μl of the crude extract was precipitated with 20% TCA (final concentration). The supernatant was separated after centrifugation (5,000 × *g*, 10 min, 4 °C), the glycogen was precipitated with 500 μl of 95% cold ethanol, collected by centrifugation, washed twice with 66% ethanol, dried, and digested with α-amylase (10 mg/ml) and amyloglucosidase (30 mg/ml). Free glucose was measured with a glucose oxidase kit, and the glycogen content was normalized to the total protein concentration. Total protein was quantified by the Hartree method (15), using BSA as standard.

RNA Extraction and Northern Assay—Total RNA was prepared using the LiCl method according to Sokolovsky *et al.* (16). Total RNA (15 μg) was electrophoresed on a 1.5% agarose-formaldehyde denaturing gel (17), at 65 V during 5 h, and then transferred to neutral nylon membranes (Hybond N, GE HealthCare) in 2 × SSC. The blots were probed with the full-length *gsn* cDNA (10⁶ to 10⁸ cpm), radiolabeled with [³²P]-dATP (3,000 μCi/mmol) by random priming (NEBlot kit, Biolabs) in 10 ml of ULTRAhyb hybridization solution (Ambion, Austin, TX), at 42 °C overnight. After hybridization, the blot was washed twice in 2 × SSC, 0.1% SDS for 10 min, and twice in 0.1 × SSC, 0.1% SDS for 15 min, and exposed to an x-ray film.

pacC cDNA Cloning and Production and Purification of the Recombinant Protein—For production of the truncated recombinant PACC transcription factor a 639-bp fragment of the ORF NCU00090 was amplified by PCR from a cDNA plasmid library (pYADE5-Nc) with the oligonucleotides 90-F and 90-R2 (Table I). The underlined sequences correspond to the *Nde*I and *Bam*HI sites, respectively. The amplified fragment was cloned into the *Nde*I-*Bam*HI sites of the pET28a vector leading to the plasmid construction pET-ΔPACC for the expression of a truncated protein containing the N-terminal 213 amino acids fused with His-tag. The plasmid construction was confirmed by DNA se-

quencing. For expression of the PACC recombinant protein, the plasmid construction was used to transform competent *E. coli* cells from BL21 (DE3) pLysS strain. Cells were grown at 37 °C in LB medium containing appropriate antibiotics and the recombinant protein was induced by addition of 0.4 mM final concentration of IPTG for 4 h, 37 °C. Cells were harvested by centrifugation, suspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, 20% v/v glycerol, 0.5 mM PMSF, 25 mM benzamidine and 50 mM NaF) and lysed by sonication. The supernatant was clarified by centrifugation and subjected to affinity chromatography. The recombinant protein was eluted with linear imidazole gradient (20–500 mM) and dialyzed against dialysis buffer (10 mM Tris-HCl pH 7.9, 100 mM KCl, 20% v/v glycerol, 1 mM EDTA, 0.5 mM dithiothreitol). Purified protein was analyzed by SDS-PAGE with Coomassie Brilliant Blue staining (18) and quantified (15).

Preparation of Crude Cellular Extract—Mycelium of the wild-type strain grown at 30 °C was used to prepare the cellular extract. About 10 mg of frozen samples were ground to a fine powder under liquid nitrogen in a prechilled mortar, homogenized in a 20 ml of lysis buffer (15 mM HEPES-KOH pH 7.9, 10% w/v glycerol, 500 mM KCl, 5 mM MgCl₂, 0.5 mM EDTA, 1 mM dithiothreitol, 0.5 mM PMSF, 10 µg/ml each of antipain and pepstatin A, 25 mM benzamidine, 50 mM NaF) and stirred with glass beads in 8 cycles of 30 s of agitation and 30 s on ice. Crude cellular extract was obtained after centrifugation (3,200 × *g*, 2 min, 4 °C), dialyzed against buffer D (15 mM HEPES-KOH pH 7.9, 15% w/v glycerol, 100 mM KCl, 1 mM EDTA) at 4 °C for 2 h, and cleared by centrifugation (20,000 × *g*, 20 min, 4 °C) before loading onto a HiTrap Heparin-Sepharose FF column (GE Healthcare). Total proteins were eluted by using a 0.1–1.5 M KCl linear gradient and the protein fractions were dialyzed against buffer D plus 0.5 mM PMSF, 25 mM benzamidine and 50 mM NaF, frozen in liquid N₂ and stored at –80 °C. Total protein was quantified (15).

Electrophoretic Mobility Shift Assay—DNA-protein binding reactions were carried out in 1 × binding buffer (25 mM HEPES-KOH pH 7.9, 20 mM KCl, 10% v/v glycerol, 1 mM dithiothreitol, 0.2 mM EDTA, 0.5 mM PMSF, 12.5 mM benzamidine, 5 µg/ml of each antipain and pepstatin A) containing 2–4 µg poly(dI-dC).(dI-dC) as unspecific competitor, and either 2.0 µg of PACC recombinant protein or 35 µg of crude cellular extract. The radiolabeled DNA probe (~10⁴ cpm) was added and reactions were incubated during 20 min at room temperature. Free probe was separated from DNA-protein complexes by electrophoresis on a native 4% polyacrilamide gel in 0.5 × Tris borate-EDTA (TBE) buffer (300 V, 10 mA, 10 °C). After electrophoresis, the gel was dried and autoradiographed. For competition assays, an excess of specific DNA competitor was added to the binding reactions 10 min prior the incubation with the radiolabeled probe.

DNA Probes and Competitors for EMSA—To produce the 134-bp *nit2*, 146-bp *pacC*, and 139-bp *xlnR* probes, DNA fragments containing the *nit2*, *pacC*, and *xlnR* *cis* elements from the *gsn* promoter, respectively, were amplified from the IV9A-1 plasmid (GenBank#AF417205) using the oligonucleotides described in Table I in the presence of [^α-³²P]-dATP (3,000 Ci/mmol). The DNA probes were purified on a 2% low-melting point agarose gel. Unlabeled *nit2*, *pacC* and *xlnR* probes, used as specific DNA competitors, were quantified by measuring the absorbance at 260 nm, and added to the binding reaction in 10-fold molar excess.

Flow Cytometry Analysis—Conidia were harvested by centrifugation after growing in Vogel's minimal solid medium with 2% sucrose, and a total of 5 × 10⁴ cells were suspended in 0.5 ml of PI solution (0.1% v/v Triton X-100, 0.1% sodium citrate, 0.9 U/ml RNase A, 50 µg/ml propidium iodide). The cell suspension was kept in the dark for 1.5 h at room temperature. A FACSCalibur™ flow cytometer equipped with CellQUEST™ software (Becton Dickinson, San Juan, CA) was used to analyze cell size using a Forward Scatter detector

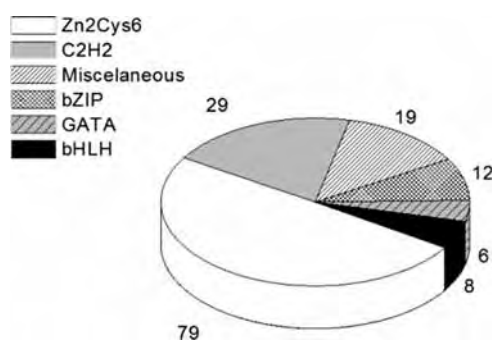


Fig. 1. Families of *N. crassa* transcription factors. Distribution in percentage of transcription factor families represented by mutant strains analyzed here.

(FSC-H), and cell complexity using a Side Scatter detector (SSC-H). Propidium iodide incorporation was measured using a fluorescence detector, with an excitation wavelength of 488 nm and an emission wavelength of 585/42 nm. Ten thousand events were evaluated per sample. The experiments were repeated at least three times.

Bioinformatic Tools—Online bioinformatic tools were used to predict the biochemical parameters of the selected transcription factors. The nucleotide sequences of the knocked-out ORFs codifying the transcription factors were identified in the fungus genome at the Broad Institute home page (<http://www.broad.mit.edu/annotation/genome/neurospora/Home.html>). The polypeptide sequences were compared against the database of sequences available at the National Center for Biotechnology Information (NCBI), using the BlastP tool (<http://www.ncbi.nlm.nih.gov/blast/>) to identify sequences with known function similar to the search sequence. For the theoretical estimates of the isoelectric point and molecular weight, the ProtParam tool at the ExPASy server was used (ca.expasy.org). The presence of protein domains was investigated using the *N. crassa* genome at the FGSC (<http://www.fgsc.net/scripts/strainsearchform.asp>), SMART (<http://smart.embl-heidelberg.de>), and Pfam 22.0 (<http://pfam.sanger.ac.uk>) web sites. The presence of classical Nuclear Localization Signals (cNLS) was determined by PSORTII (<http://psort.nibb.ac.jp/form2.html>), and the presence of *cis* DNA elements in the *gsn* promoter was determined by MatInspector (www.genomatix.de).

RESULTS

We screened a set of *N. crassa* mutant strains, each carrying a deletion in a single gene encoding a transcription factor. Fig. 1 shows the transcription factor families to which the proteins belong. Most belong to the Zn₂Cys₆ fungal binuclear cluster family, which is fungus-specific and the largest class of transcription factors in *N. crassa* (2). The second largest class of transcription factors analyzed in this work is the C₂H₂ family, found in both prokaryotic and eukaryotic organisms. A considerable number of miscellaneous factors were screened, including the CAAT-binding transcription factors, and transcription factors carrying the forkhead, homeobox, RING finger, and WD repeat domains. These factors usually play important roles in the cell cycle biology of eukaryotic organisms. A smaller number of transcription factors belonging to the bZIP, GATA, and bHLH families was also screened (2).

Transcription Factors Controlling Glycogen Accumulation—Previous results from our group have shown that wild-type *N.*

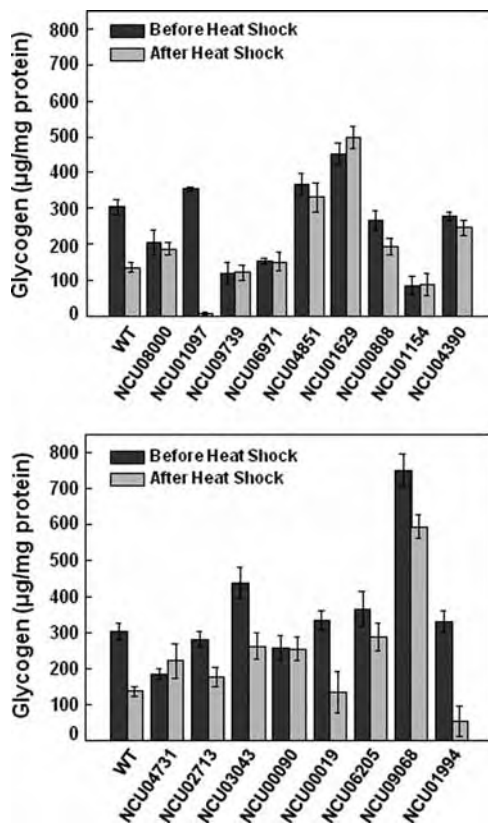


FIG. 2. Glycogen accumulation before and after heat shock in selected mutant strains. Glycogen was extracted from mycelia submitted or not to heat shock (transfer from 30 °C to 45 °C), digested with α -amylase (10 mg/ml) and amyloglucosidase (30 mg/ml), and the free glucose was enzymatically determined with a glucose oxidase kit. Results represent the average of at least three independent experiments. WT, FGSC 9718 strain.

crassa accumulates glycogen at high levels at the end of the exponential growth phase while growing under its regular temperature (30 °C), and that the glycogen content decreases when the mycelium is exposed to heat shock (45 °C) (5, 6). To identify transcription factors regulating glycogen metabolism in the fungus *N. crassa*, we used a mutant-strain set with single-gene deletions of putative or already known transcription factors to search for mutant strains having glycogen accumulation profiles different from that found in the wild-type strain. The mutant strains were analyzed under normal growth temperature (30 °C) and heat shock stress (45 °C).

Of the 147 mutant strains analyzed, 17 presented patterns of glycogen accumulation different than the wild-type strain (Fig. 2). Five strains (knocked-out in the ORFs NCU08000, NCU09739, NCU06971, NCU01154, and NCU00090) were selected because they did not show differences in their glycogen content either before or after heat shock. Note that at both temperatures, low levels of glycogen accumulated in the strains with deleted NCU09739, NCU06971, and NCU01154 ORFs. The strains with deleted NCU04851, NCU00808, NCU04390, NCU02713, NCU03043, and NCU06205 ORFs

were selected because they showed less pronounced reduction in glycogen levels after heat shock, compared with the wild-type strain. Two mutant strains (mutated in the ORFs NCU01097 and NCU01994) showed extremely large glycogen content reductions after heat shock. Interestingly, the strains mutated in the ORFs NCU01629, and NCU04731 accumulated more glycogen under the heat shock condition than under the normal growth condition. Finally, a hyper-accumulation of glycogen was observed in the NCU01629 and NCU09068 ORFs deleted strains in both temperatures. All of these mutant strains showed impaired control of glycogen accumulation, compared with the wild-type strain, suggesting that the transcription factors missing in the mutant strains might be involved in the regulation of glycogen accumulation.

The biochemical and molecular characteristics of the selected transcription factors are shown in Table II. Most belong to the Zn₂Cys₆ zinc finger family, the largest class in *N. crassa* (2). Four proteins belong to the C₂H₂ zinc finger family, and two to the GATA zinc finger, a family only found in eukaryotic organisms. The bHLH and bZIP transcription factor families both have one member among the selected proteins, and proteins having either the forkhead domain or the Tup-N-terminal and WD repeats, which belong to the miscellaneous factor family (2), were also selected in our screen. Fifteen selected transcription factors have cNLS, which are amino acid sequences that target cargo proteins into the nucleus, either monopartite or bipartite (19). Two transcription factors (ORFs NCU09739 and NCU06205) do not have amino acid sequences characteristic of cNLS. One transcription factor, annotated as hypothetical protein, might have a nonclassical NLS (ORF NCU09739), and the *rco-1* (regulation of conidiation) gene product is the *S. cerevisiae* Tup1 homolog (ORF NCU06205, described below).

Most of the selected transcription factors have not yet been characterized at the protein level, and were annotated as hypothetical proteins in the *N. crassa* database (Table II), whereas many of them have been described as being involved with growth and development in the fungus (12). Only a few proteins have functional roles that have already been extensively studied. One is the *N. crassa* NIT2 protein (ORF NCU09068), a GATA transcription factor that interacts with the Zn₂Cys₆ NIT4 protein to activate expression of nitrate and nitrite reductases (20). Another is the *S. cerevisiae* Tup1 ortholog RCO-1 protein (ORF NCU06205), functionally characterized as a regulatory protein that mediates mycelial repression of conidiation gene expression (21). The gene annotated as *pacC* (ORF NCU00090) is the *pacC/RIM101* ortholog, extensively studied in *Aspergillus nidulans* and *S. cerevisiae* as encoding a transcription factor involved in pH regulation by activating genes in alkaline conditions and repressing those genes expressed in acidic conditions (22). The *ctf-1 α* gene (ORF NCU08000) product is the well-characterized cutinase transcription factor Ctf1 α ortholog, which was described as upregulating genes encoding cutinase enzymes in other fila-

Transcription Factors and Glycogen Metabolism Regulation

TABLE II

Classification of the transcription factor family, annotation, biochemistry, and structural characteristics of the proteins selected

The identification of each strain was made according to the FGSC number. Theoretical estimate of physical and chemical characteristics and identification of cNLS were performed according to ProtParam tools (www.expasy.org/tools/protpar-ref.html) and PsortII (<http://psort.nibb.ac.jp/cgi-bin/runpsort.pl>), respectively. MW, molecular weight; pI, isoelectric point; cNLS, classical Nuclear Localization Signals; M, monopartite; B, bipartite; NI, not identified.

FGSC#	ORF	Transcription factor family	Theoretical MW/pI	cNLS	Annotation ^a	Gene	Reference
11004	NCU08000	Zn ₂ Cys ₆ zinc finger	105.43/6.78	33-PTPKRKK (M) 600-KRHRR (M)	Cutinase transcription factor 1 α	<i>ctf-1α</i>	23
11039	NCU01097	Zn ₂ Cys ₆ zinc finger	90.23/6.34	36-KRVKAVTQACHTCRRYK (B)	Hypothetical protein	NI	-
11062	NCU09739	Zn ₂ Cys ₆ zinc finger	75.09/7.85	NI	Hypothetical protein	<i>ada-7</i>	-
11067	NCU06971	Zn ₂ Cys ₆ zinc finger	103.16/6.44	87-PIRRRIS (M) 128-RKKR (M)	Transcriptional activator XlnR	<i>xlnR</i>	28
11089	NCU04851	Zn ₂ Cys ₆ zinc finger	119.36/7.87	268-PKEKRWP (M) 221-PKRRNRPAVSCIPCRGRKI (B)	Hypothetical protein	NI	-
11102	NCU01629	C ₂ H ₂ zinc finger	45.91/8.82	240-PRPKRQQ (M)	Hypothetical protein	NI	-
11123	NCU00808	Zn ₂ Cys ₆ zinc finger	66.25/5.79	185-PRIKTKK (M)	Hypothetical protein	NI	-
11126	NCU01154	GATA zinc finger	51.01/8.89	402-KRKK (M)	Hypothetical protein	<i>sub-1</i>	-
11134	NCU04390	Zn ₂ Cys ₆ zinc finger	93.67/6.6	89-PQPPRRRKKK (M) 94-RRKKKPKHERDLIDLKKY (B)	Hypothetical protein	<i>col-22</i>	-
11139	NCU04731	bHLH	117.54/6.17	276-PNSRKRK (M)	Hypothetical protein	<i>sah-2</i>	33
11348	NCU02713	C ₂ H ₂ zinc finger	31.2/8.99	107-KRPR (M)	Hypothetical protein	<i>csp-1</i>	30
11355	NCU03043	C ₂ H ₂ zinc finger	41.73/ 8.57	312-KKHK (M) 361-RRHKK (M)	Protein F1bC	NI	-
11397	NCU00090	C ₂ H ₂ zinc finger	67.3/7.19	280-PFDARKR (M) 285-KRQFDLNDFFGSVKRR (B)	pH response transcription factor <i>pacC/RIM101</i>	<i>pacC-1</i>	37
11437	NCU00019	Forkhead domain	74.61/7.86	33-PSKRKKK (M) 454- PASSRKRK (M)	FKH1 protein	NI	-
11371	NCU06205	Tup-N-Terminal & WD repeats	66.08/6.42	NI	Transcriptional repressor	<i>rco-1</i>	21
11392	NCU09068	GATA zinc finger	109.29/8.89	287-PIKARKD (M) 325-RKTSIDETSKRNPNRKR (B)	Nitrate catabolic enzyme regulatory protein	<i>nit-2</i>	39
11342	NCU01994	bZIP	36.01/5.26	140-PAQSRRK (M) 144- RRKAQNRAAQRAFRRERKE (B)	Hypothetical protein	NI	-

^a Annotation was performed according to Borkovich *et al.* (2) and the *N. crassa* database at the FGSC site (<http://www.broadinstitute.org/annotation/genome/neurospora/MultiHome.html>).

mentous fungi, such as *Fusarium solani* (23, 24) and *F. oxysporum* (25). The NCU03043 gene product revealed homology (68% identity) with the FLE1 protein of *Podospora anserina* (26) and the F1bC protein of different fungi including *A. clavatus* and *A. fumigatus* (27). Both FLE1 and F1bC are transcription factors involved in fungi development. Finally, the XlnR ortholog encoded by the ORF NCU06971 is a global transcriptional activator controlling the expression of genes encoding xylanolytic and cellulolytic enzymes, which was first isolated in *A. niger* (28).

Although most of the transcription factors identified in our screen were annotated as hypothetical proteins, some showed homology with proteins that have already been studied. The ORF NCU02713 deleted strain corresponds to the *N. crassa* *csp-1* (conidial separation-1) mutant previously isolated by Selitrennikoff *et al.* (29), which showed improper separation of conidia from hyphae. This gene was recently described to encode a light-inducible transcription factor (30). The protein codified by the ORF NCU01994 is the *Candida albicans* Fcr3 (fluconazole resistance 3) ortholog transcription factor (31). The Fcr3 protein was able to complement a *S. cerevisiae* mutant strain lacking the transcription factors Pdr1

and Pdr3, which control the expression of several genes involved in Pleiotropic Drug Resistance. The NCU00019 gene product belongs to the forkhead (or Fox, for Forkhead box) transcription factor family, which has been identified in many metazoans as playing important roles in diverse biological processes (32). The ORF NCU04731 encodes a protein having 52% identity with the *S. pombe* Sre1 transcription factor, a sterol regulatory element binding protein (SREBP) functionally conserved among different fungi. Fungal SREBPs are hypoxic transcription factors required for adaptation to a low oxygen environment (33). Finally, the ORFs NCU01097, NCU09739, NCU04851, NCU01629, NCU00808, NCU01154, and NCU04390 encode hypothetical proteins without any previously described function.

***gsn* Expression in Selected Mutant Strains** — To investigate whether the glycogen accumulation pattern of mutant strains could be related to *gsn* regulation, gene expression analysis was performed by Northern blot. Previous results from our group have shown a decrease in *gsn* transcript levels when mycelium of the wild-type *N. crassa* were transferred from 30 °C to 45 °C (5, 8). In the present study, *gsn* gene expression in the selected mutant strains was analyzed before and

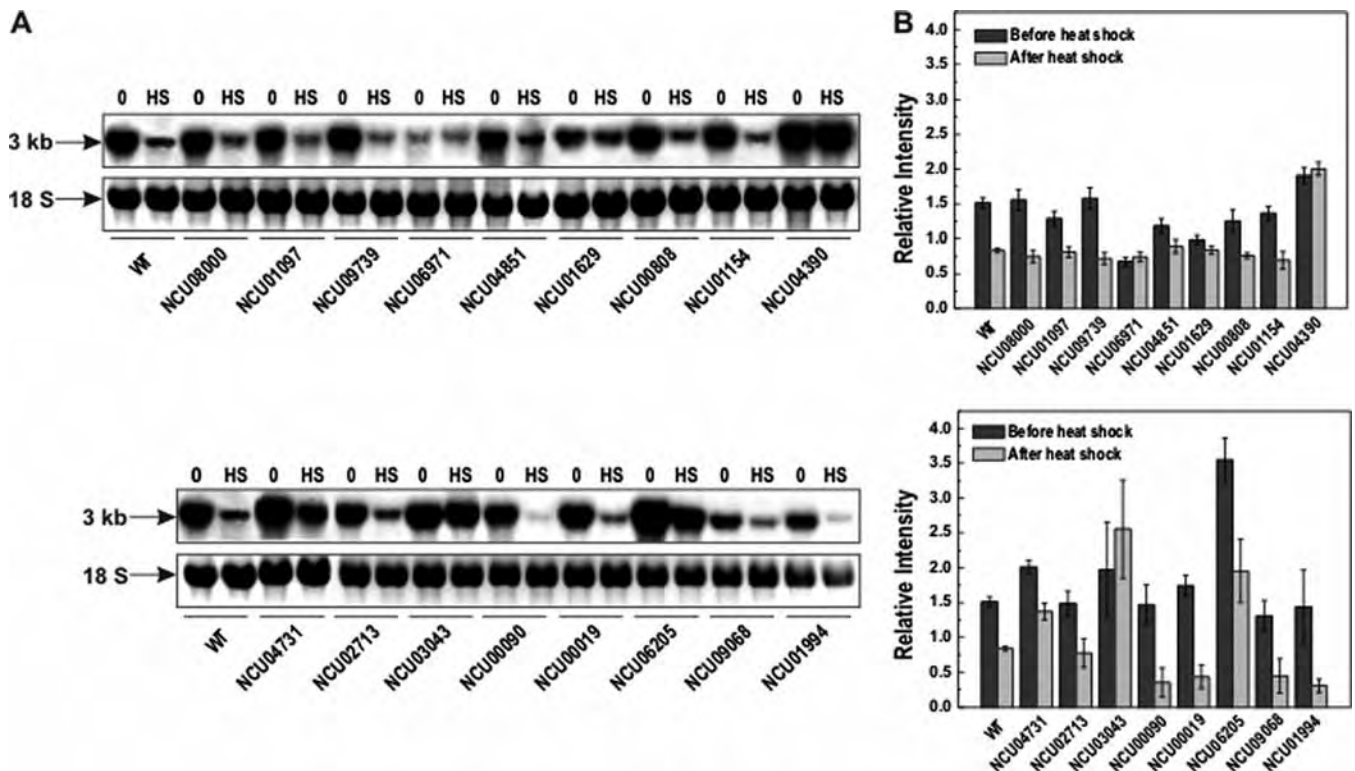


FIG. 3. *gsn* gene expression assay by Northern blot before and after heat shock in the selected mutant strains. Mycelia were cultivated at 30 °C for 24 h and then shifted to 45 °C. Samples were collected before (0) and after (HS) temperature shift and total RNA was extracted. Total RNA (15 μg) was separated by electrophoresis in a denaturing formaldehyde gel, transferred to nylon membrane and probed with the [α - 32 P] radiolabeled full-length *gsn* cDNA. A, upper panel, gel autoradiography, lower panel, the 18 S rRNA was used as loading control after ethidium bromide staining of the same gels. B, densitometric analysis of the *gsn* gene expression relative induction (ImageJ software). Results represent the average of at least three independent experiments. WT, FGSC 9718 strain.

after exposure to 45 °C. Many of the mutant strains presented a *gsn* transcription profile similar to that observed in the wild-type strain, in which a lower *gsn* expression was observed after heat shock (Fig. 3). However some mutants showed differences, either because the transcript levels were slightly reduced after heat shock (as seen, for example, for strains with deleted NCU04851, NCU01629, and NCU04731 ORFs), or because the transcript levels were strongly reduced after heat shock (as for strains with deleted NCU00019 and NCU01994 ORFs), as compared with the wild-type strain. Surprisingly, some of the mutant strains showed equal transcript levels both before and after heat shock (strains with NCU06971, NCU04390, and NCU03043 deleted ORFs), indicating loss of *gsn* gene expression regulation after heat stress. Also, high transcript levels for NCU04390 ORF deleted strain (both before and after heat shock) and NCU06205 ORF deleted strain (before heat shock) were detected.

An attempt to correlate the *gsn* expression profile with the glycogen accumulated under both environmental conditions analyzed in this work indicated that in several mutant strains the amount of glycogen correlated with *gsn* transcript levels (Fig. 3). For example, the NCU06971 ORF deleted strain showed equal amounts of glycogen and the same transcript

levels before and after heat shock. In addition, the amount of glycogen and the *gsn* transcript levels before heat shock were lower than that found in the wild-type strain. A good correlation was also observed for the strains mutated in the ORFs NCU04851, NCU04390, and NCU0199. However, the hyper-accumulation of glycogen observed in the ORF NCU09068 mutant strain may not solely result from *gsn* expression, as low transcript levels were observed under both environmental conditions. Thus, the transcription factors could act either directly on gene expression or indirectly, by regulating a gene whose product affects gene expression.

A search for putative transcription factor binding sites on the *gsn* 5'-flanking region was performed using the *MatInspector* tool. One putative binding site for the CTF α , PacC, and NIT2 proteins was found at positions -2248, -1803, and -1558 bp, respectively, two putative binding sites for the Sre1 (positions -1758 and -2024) and XlnR (positions -2071 and -2252) orthologue were found, and five sites were identified for the Mig1 orthologue at positions -2305, -2026, -1592, -1487, and -317, relative to the ATG start codon (Fig. 4). The *S. cerevisiae* Mig1 transcription factor, the Cre1/CreA protein orthologue in filamentous fungi, is a major protein that drives the complex Ssn6-Tup1 by repressing a set of glucose-repressible genes (34). Interestingly, the Tup1 protein is the *N.*

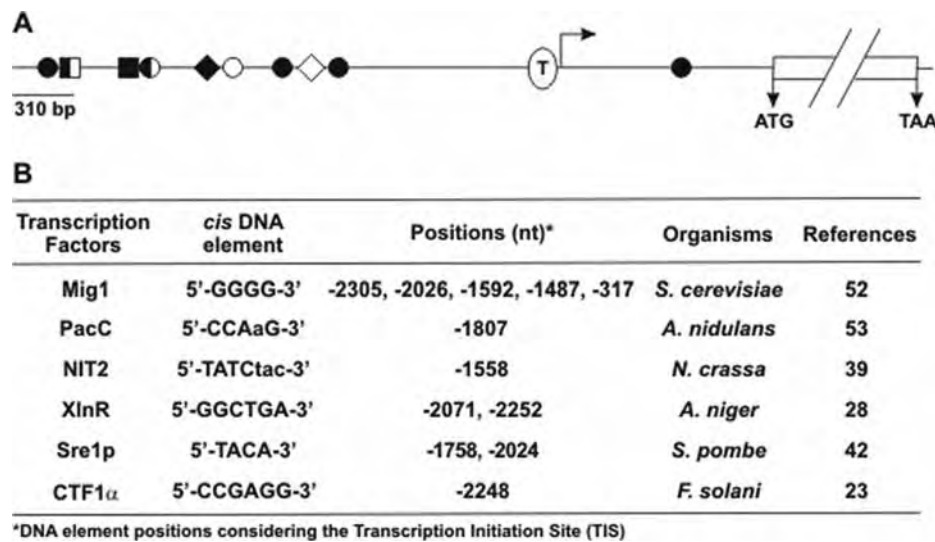


FIG. 4. **Representation of the *gsn* gene 5'-flanking region.** A, the relative position of the DNA motifs recognized by the transcription factors Mig1 (●), PacC (◆), NIT2 (◇), XlnR (■), Sre1p (○), and CTF1 α (□) are indicated. The TATA-box sequence is indicated by T. The *gsn* ORF is delimited by the ATG start codon and the TAA stop codon. The Transcription Initiation Site (TIS) is represented by an arrow. B, regulatory DNA elements found in the *gsn* gene 5'-flanking region.

crassa RCO-1 protein (NCU06205) ortholog, which was identified in the present study as a putative transcription factor involved in the regulation of glycogen metabolism.

Binding of the Transcription Factors to the *gsn* Promoter—Gel shift analysis was performed to investigate whether some of the selected transcription factors having putative *cis* motifs in the *gsn* promoter were able to recognize and bind to DNA fragments containing their respective motifs. As shown in Fig. 5, the recombinant PACC protein bound to the *pacC* probe and formed a unique and strong DNA-protein complex (Fig. 5A, lane 2), which was reduced in the presence of a 10-fold molar excess of unlabeled *pacC* probe (Fig. 5A, lane 3). This finding shows that the *pacC cis* element present in the *gsn* promoter is an active binding site for the PACC transcription factor. The NIT2 and XLNR putative binding sites were analyzed using crude cellular extract (CCE) prepared from mycelium from the wild-type strain fractionated by affinity chromatography. Two chromatographic fractions were analyzed in binding reactions with the *nit2* probe, as shown in Fig. 5B. DNA-protein complexes were observed in the two fractions (Fig. 5B, lanes 2 and 4), which were removed in the presence of 10-fold molar excess of the unlabeled probes (Fig. 5B, lanes 3 and 5). One fraction was used to analyze the *xlnR* probe and a specific complex was observed (Fig. 5C, lane 2), which was also reduced in the presence of 10-fold molar excess of the unlabeled probe (Fig. 5C, lane 3). Taken together, the results indicate that the transcription factors analyzed were able to bind to their *cis* elements present in the *gsn* promoter, suggesting they may have a role in glycogen metabolism regulation.

Flow Cytometric Analysis—Flow cytometric analysis was used to investigate the cell size, cell complexity, and level of

propidium iodide (PI) incorporation of the selected mutant strains. PI is a red fluorescent compound that binds to DNA and can be used to evaluate DNA content in individualized cells by flow cytometry. The intensity of fluorescence is proportional to the quantity of DNA available to PI intercalation. The DNA content in a cell can vary as a function of the cell cycle phase, therefore the comparative intensity of fluorescence is G0/G1 < S < G2/M. Apoptotic cells carry condensed and fragmented DNA promoting a lower fluorescence than cells at G0/G1. This analysis was performed for 10,000 cells from each mutant strain, and compared with the wild-type strain. Fig. 6 shows the results obtained from the Forward Scatter and Side Scatter detectors, which reveal cellular size and complexity, respectively. From all strains analyzed, the NCU02713, NCU03043, NCU06205, and NCU09739 ORFs deleted strains presented altered light scattering profiles, compared with the wild-type strain. The NCU02713 ORF deleted strain showed a discrete increase of cells that was smaller and slightly less complex. The NCU03043 and NCU09739 ORFs deleted strains presented expressive increase of smaller size cells and decreased complexity, whereas the NCU06205 ORF deleted strain showed a distinct increase of smaller cells together with some decrease in complexity. It is important to mention that the parameter of complexity (*i.e.* shape of the nucleus, the amount and type of cytoplasmic granules and membrane roughness) can be affected by cell size.

Fig. 7 shows the results obtained from mutants that showed changes on fluorescence profile by PI treatment. All mutants with defective morphology presented some alteration in PI incorporation, as revealed by overlapping the mutants and wild-type strain profiles (*right panels*). The areas M1 and M2

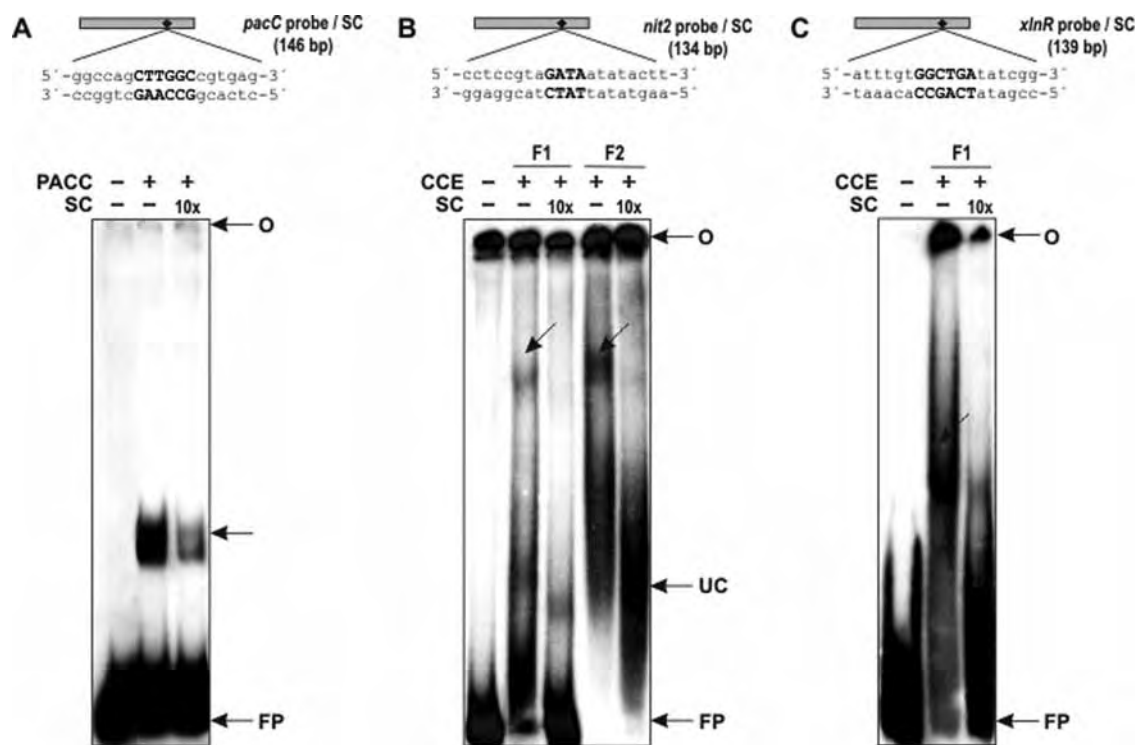


FIG. 5. Binding of proteins from *N. crassa* WT strain to fragments of the *gsn* promoter. *A*, upper panel, schematic representation of the *pacC* probe with the PACC motif (small black diamond) and part of its neighboring sequences. Lower panel, gel shift analysis of PACC motif using 2 μ g of the His- Δ PACC recombinant protein. Lane 1, *pacC* probe, no protein added. The DNA-complexes are indicated by arrows. *B* and *C*, upper panels, schematic representation of the *nit2* and *xlnR* probes, respectively, with their motifs (small black diamond) and part of their neighboring sequences. Lower panels, gel shift analysis of NIT2 and XLNR motifs using Heparin-Sepharose chromatographic fractions (F1 and F2). An amount of 35 μ g of each fraction obtained by affinity chromatography of crude cellular extract was assayed. Lane 1, *nit2* and *xlnR* probes, no protein added; Lanes 2 and 4, binding in the absence of the specific competitors; Lanes 3 and 5, binding in the presence of the specific competitors. CCE, crude cellular extract; UC, unspecific complex; O, gel origin; SC, specific competitor; FP, free probe.

represent cells at transition phases G0/G1 and G2/M of the cell cycle, respectively. The NCU02713 ORF deleted strain presented a high PI content for cells at the G2/M transition phase, indicating that this strain can carry some cellular impairment associated with the G2/M phase. The NCU03043 and NCU06205 ORFs deleted strains showed similar profiles, with a high content of cells between the M1 and M2 areas. This area represents cells at the S phase, indicating that these strains can carry impairments associated with the S phase or related to the S/G2 transition phase. The NCU09739 ORF deleted strain presented a high content of cells prior to the M1 area. Here, low PI incorporation might be a consequence of either loss of DNA content because of an irregular cell cycle, or DNA fragmentation associated to apoptosis. It is noteworthy that the mutant strains shown in Fig. 7 display visible and severe growth defects, such as low conidiation and reduction in both the extension and production of aerial hyphae, compared with the wild-type strain, indicating that the proteins are required for normal vegetative growth and development.

DISCUSSION

The release of the complete *N. crassa* genome (1) and the availability of a mutant strains collection with each strain

deleted in a single gene have allowed us to perform a screening aimed to investigate specific aspects of the fungus cell biology. In this type of screen, the phenotype analyzed corresponds to a loss of function, so that conclusions can be drawn concerning the involvement of a particular gene product in the phenotype scored as emphasized in a similar screen in *S. cerevisiae* (35). In this work, we detected alterations in glycogen accumulation in *N. crassa* strains deleted in transcription factors, to identify not yet previously described transcriptional regulators of glycogen metabolism. Transcription factors often control gene transcription through binding to specific DNA-binding sites, which can either promote (activate) or repress (inhibit) the recruitment of the transcription initiation machinery. To fully understand a gene function, it is helpful to understand the regulatory network context in which the gene participates, and that includes identification of the transcription factors involved in its regulation.

The fungus *N. crassa* has been widely used as a model organism, as it exhibits both asexual development and sexual differentiation. Following the availability of its genome sequencing, it has emerged as a suitable model organism for higher eukaryote studies because of its multicellularity and the high number of genes without orthologues available in public

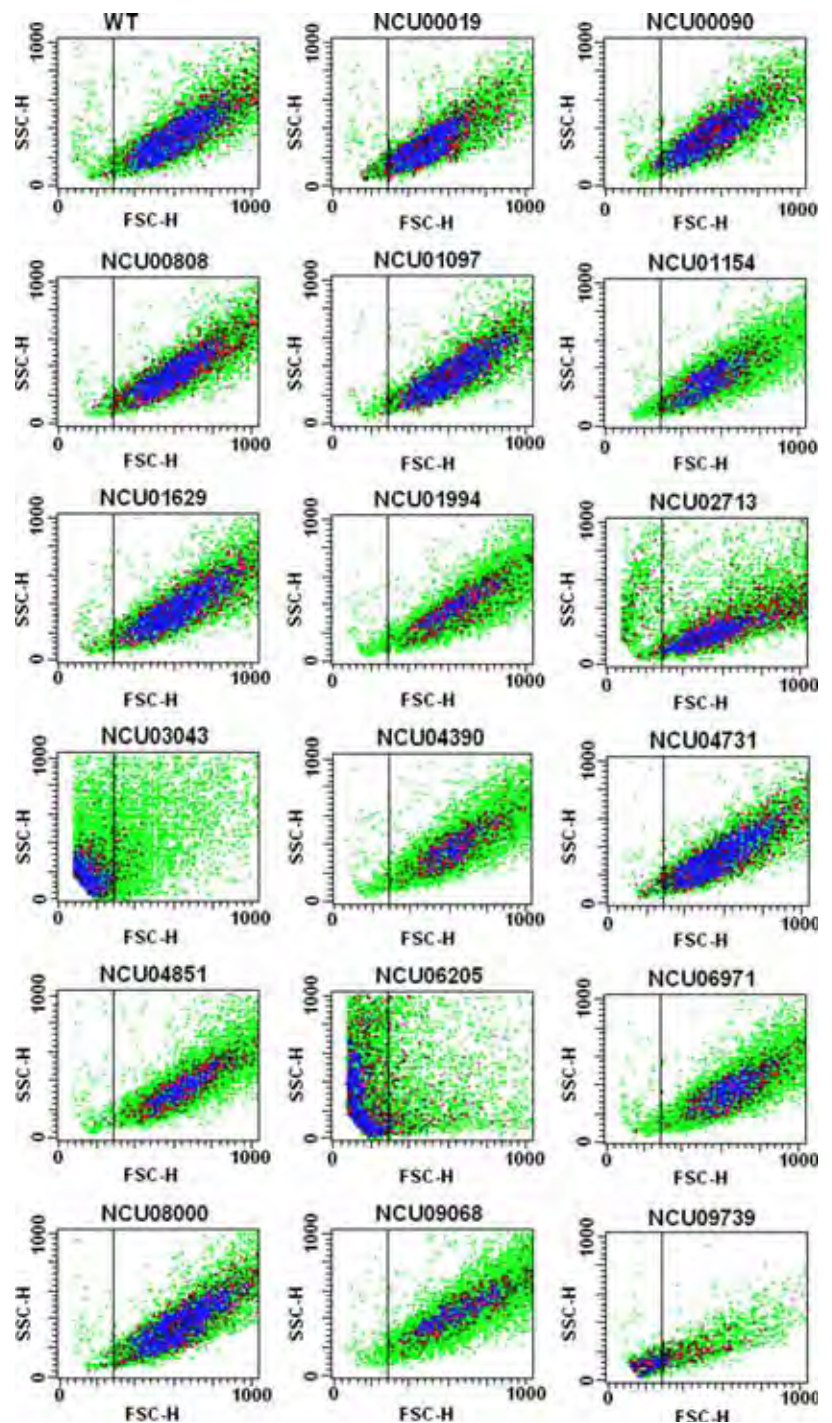


FIG. 6. **Morphological analysis of the mutant strains by flow cytometry.** Analysis of cell size (FSC-H) and cell complexity (SSC-H) was performed using 10,000 events. Results from one of at least three independent experiments are shown.

databases. It is typically a haploid organism, undergoing only a very transient diploid stage immediately prior to meiosis. In addition, the existence of gene-silencing mechanisms, such as the Repeat Induced Point Mutation, that eliminate duplicated sequences (36), makes it an advantageous organism in the type of screen performed in the present work, which is based on mutant strains having single gene deletion.

Herein we used a quantitative assay to measure glycogen accumulation in a set of strains with deleted transcription

factors. Our results demonstrate that transcription factors belonging to different families regulate glycogen metabolism during vegetative growth and also under a stress condition such as heat shock. It should be noted that the amount of glycogen that is accumulated results from the balance between glycogen synthase and glycogen phosphorylase activities. These enzymes are regulated in an opposite way; phosphorylation activates glycogen phosphorylase and inhibits glycogen synthase. In *N. crassa* glycogen synthase activity is

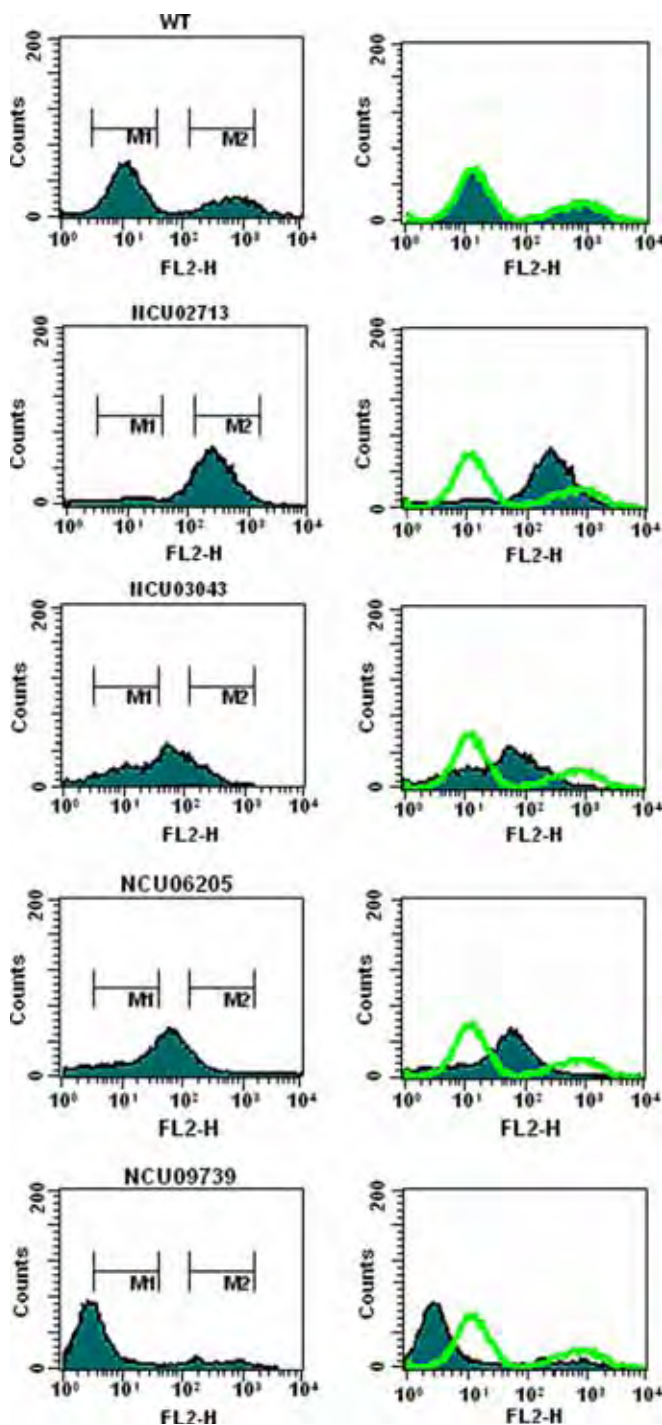


FIG. 7. Analysis of propidium iodide (PI) incorporation by the selected mutant strains using flow cytometry. PI fluorescence (FL2-H) was measured for 10,000 events. The M1 and M2 areas reveal cells at the G0/G1 and G2/M phases of the cell cycle, respectively. The area between M1 and M2 is related to cells at the S phase of the cell cycle. The sub-M1 area is related to cells with low PI incorporation. Results from one of at least three independent experiments are shown.

inhibited under heat shock (45 °C) whereas glycogen phosphorylase activity is activated under the same condition (5, 6). Many of the transcription factors identified in our screen have

been functionally characterized in different organisms, and some of them may play a role in the control of glycogen metabolism. This hypothesis is reinforced by the existence of DNA-binding sites specific for the transcription factors found in the *gsn* gene 5'-flanking region.

The transcription factor PacC responds to changes in extracellular pH by activating specific alkaline genes and repressing specific acid genes (37). Here, we demonstrated that this transcription factor binds to *gsn* promoter. The transcription factor XlnR is described as a transcriptional activator controlling the expression of genes encoding xylanolytic and cellulolytic enzymes in filamentous fungi (28). Our results indicate the latter transcription factor may also control the expression of genes encoding enzymes required for the metabolism of other carbon sources, such as the enzymes involved in glycogen metabolism, thus up-regulation of *gsn* expression results in glycogen accumulation. Interestingly, it has been shown that the XlnR-induced expression of genes encoding xylanolytic enzymes is modulated by the carbon catabolite repressor Cre1/CreA transcription factor in *Aspergillus* (38). The Cre1/CreA fungal protein is the *S. cerevisiae* Mig1 transcription factor orthologue that has five putative binding sites in the promoter *gsn*, suggesting that this transcription factor is a protein that deserves further investigation concerning the regulation of glycogen metabolism.

Another transcription factor identified in our screen is the *N. crassa* NIT2 protein (AreA in *A. nidulans*), a member of the GATA factors family, already characterized in *N. crassa* as a positive regulator of genes encoding enzymes for nitrogen source catabolism under nitrogen limiting conditions (39). Based on our results, we can suggest that this transcription factor acts as a repressor of carbon metabolism, because the knocked-out strain showed loss of glycogen accumulation regulation, despite having low *gsn* gene expression as compared with the wild-type strain. A link between carbon and nitrogen regulation was reported by Lockington *et al.* (40), who described the effect of both carbon and nitrogen sources on the amount of cellulases secreted in *A. nidulans*. Although the result was preliminary, the authors suggested the existence of a link in the regulation of the carbon and nitrogen utilization pathways in filamentous fungi.

Although most transcription factors here identified belong to the zinc finger family, at least one member of the bHLH, bZIP, forkhead, and WD repeat proteins was also identified. Interestingly, the bHLH transcription factor identified in this work (NCU04731) has, at the C-terminal region, a domain found in ER membrane-bound transcription factors called SREBP. The SREBPs can be distinguished from other bHLH proteins by the presence of a tyrosine instead of an arginine residue in their basic regions (41). The first SREBP fungal orthologue was identified in *S. pombe* (42), and more recently it was characterized in other fungal species, such as *Candida albicans* (43), *A. fumigatus* (44), and *Cryptococcus neoformans* (45). The ability to respond to sterol is conserved be-

tween mammalian and fungal SREBPs, however the fungal proteins are hypoxic transcription factors required for growth by regulating genes under low-oxygen conditions (33). The knocked-out *N. crassa* strain for this transcription factor showed impaired glycogen accumulation compared with the wild-type strain, which did not correlate with *gsn* gene expression. Considering that glycogen accumulates under anaerobic conditions, a potential involvement of this transcription factor in this metabolic process cannot be ruled out and deserves further investigation.

A very interesting transcription factor identified in our screen is the RCO-1, characterized in *N. crassa* as a regulatory protein that mediates mycelial repression of conidiation gene expression (21). It is a homologue of *S. cerevisiae* Tup1, a multidomain protein that mediates transcriptional repression of genes concerned with a variety of processes. In *S. cerevisiae*, Tup1 and Ssn6 proteins comprise a protein complex that is required for repression of several apparently unrelated genes, including glucose-repressible genes. They need to be physically associated in order to be recruited to promoters by *trans*-acting DNA-binding proteins (34). Our results indicate that this transcription factor could also control glycogen metabolism as a repressor of *gsn* gene expression in order to favor free glucose inside the cell. Recent results have shown that RCO-1 participates in photoadaptation in *N. crassa* by repressing gene transcription after a long exposure to light (46), showing that a mutation that affects the conidiation process also presents clock effects (47).

As previously mentioned, the *csp-1* gene product identified in our screen is also a light-inducible transcription factor (30), and a strain mutated in this gene developed superficially normal-looking conidia that failed to completely separate and remained tightly linked (29). The fact that both proteins (RCO-1 and CSP-1) play roles in conidiation and in circadian rhythms, and that the strains mutated in these proteins showed improper glycogen accumulation, led us to speculate on the existence of a connection between circadian clocks and glycogen metabolism, similar to what was described for trehalose in *N. crassa* (48). In the latter case, the gene encoding trehalose synthase is a clock-controlled gene (*ccg-9*), thus connecting the requirement for trehalose in clock regulation. In *Drosophila*, Zheng and Sehgal (49) demonstrated that the AKT and TOR-S6K pathways, which are the major regulators of nutrient metabolism, cell growth, and senescence, affect the brain circadian clock that drives behavioral rhythms. Another recent example of a link between metabolism and circadian rhythms was described by Doi *et al.* (50), providing direct evidence of the action of the circadian clock in the regulation of mammalian glycogen metabolism. The authors demonstrated that the CLOCK transcription factor regulates the circadian rhythms of hepatic glycogen synthesis through transcriptional activation of *Gys2* (glycogen synthase 2), which is the rate-limiting enzyme of glycogenesis in the liver.

The fact that deletion of some of the transcription factors affects cell cycle progression, as demonstrated by flow cytometry assay, is especially interesting. New cell cycle transcription factors have been discovered, which constitute an important tool for studies concerning the regulation of cell cycle transcription. Using a systematic screen to reveal new *S. cerevisiae* cell cycle transcription factors, White *et al.* (51) identified a series of transcription factors having functional roles in different biological processes, including glucose and lipid metabolism. Although the results obtained in our screen did not directly show a role for the transcription factors we identified, they constitute a valuable group of candidate proteins acting as regulators in glycogen metabolism control. Our results open new opportunities for investigating key issues concerning glycogen metabolism regulation, such as how glycogen metabolism could be connected to cell-cycle regulation, biological clock, and other aspects of cellular metabolism. A better understanding of such connections will bring insights into the importance of the energy balance in biological processes.

Acknowledgments—We thank Dr. Rui Curi from Instituto de Ciências Biomédicas, USP, São Paulo, Brazil, for the flow cytometry assays.

* This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) by grants to M. C. Bertolini and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). R. D. G. and F. B. C. are graduate fellows supported by CAPES and FAPESP, respectively, F. Z. F. is a post-doctoral fellow from FAPESP, and M. C. B. is a CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) fellow.

¶ To whom correspondence should be addressed: Instituto de Química, UNESP, R. Professor Francisco Degni, s/n, 14800-900, Araraquara, SP, Brazil. Tel.: 55-16-3301-9675; Fax: 55-16-3301-9692; E-mail: mcbertol@iq.unesp.br.

|| Both authors contributed equally to this work.

REFERENCES

- Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U. Read, N. D., Jaffe, D., FitzHugh, W., Ma, L. J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C. B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M. A., Werner-Washburne, M., Selitrennikoff, C. P., Kinsey, J. A., Braun, E. L., Zelter, A., Schulte, U., Kothe, G. O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor, J., Stange-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamvysseis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzberg, R. L., Perkins, D. D., Kroken, S., Cogoni, C., Macino, G., Catchside, D., Li, W., Pratt, R. J., Osmani, S. A., DeSouza, C. P., Glass, L., Orbach, M. J., Berglund, J. A., Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D. O., Alex, L. A., Mannhaupt, G., Ebbole, D. J., Freitag, M., Paulsen, I., Sachs, M. S., Lander, E. S., Nusbaum, C., and Birren, B. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**, 859–868
- Borkovich, K. A., Alex, L. A., Yarden, O., Freitag, M., Turner, G. E., Read, N. D., Seiler, S., Bell-Pedersen, D., Paietta, J., Plesofsky, N., Plamann, M., Goodrich-Tanrikulu, M., Schulte, U., Mannhaupt, G., Nargang, F. E., Radford, A., Selitrennikoff, C., Galagan, J. E., Dunlap, J. C., Loros, J. J., Catchside, D., Inoue, H., Aramayo, R., Polymenis, M., Selker, E. U., Sachs, M. S., Marzluf, G. A., Paulsen, I., Davis, R., Ebbole, D. J., Zelter, A., Kalkman, E. R., O'Rourke, R., Bowring, F., Yeadon, J., Ishii, C., Suzuki, K., Sakai, W., and Pratt, R. (2004) Lessons from the genome

- sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiol. Mol. Biol. Rev.* **68**, 1–108
3. Hardy, T. A., Huang, D., and Roach, P. J. (1994) Interactions between cAMP-dependent and SNF1 protein kinases in the control of glycogen accumulation in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **269**, 27907–27913
 4. Freitas, F. Z., de Paula, R. M., Barbosa, L. C. B., Terenzi, H. F., and Bertolini, M. C. (2010) cAMP signaling pathway controls glycogen metabolism in *Neurospora crassa* by regulating the glycogen synthase gene expression and phosphorylation. *Fungal Genet. Biol.* **47**, 43–52
 5. de Paula, R., de Pinho, C. A., Terenzi, H. F., and Bertolini, M. C. (2002) Molecular and biochemical characterization of the *Neurospora crassa* glycogen synthase encoded by the *gsn* cDNA. *Mol. Genet. Genomics* **267**, 241–253
 6. Noventa-Jordão, M. A., de Lourdes, M., Polizeli, T. M., Bonini, B. M., Jorge, J. A., and Terenzi, H. F. (1996) Effects of temperature shifts on the activities of *Neurospora crassa* glycogen synthase, glycogen phosphorylase and trehalose-6-phosphate synthase. *FEBS Lett.* **378**, 32–36
 7. Parrou, J. L., Teste, M. A., and François, J. (1997) Effects of various types of stress on the metabolism of reserve carbohydrates in *Saccharomyces cerevisiae*: genetic evidence for a stress-induced recycling of glycogen and trehalose. *Microbiology* **143**, 1891–1900
 8. Freitas, F. Z., and Bertolini, M. C. (2004) Genomic organization of the *Neurospora crassa gsn* gene. Possible involvement of the STRE and HSE elements in the modulation of gene transcription during heat shock. *Mol. Genet. Genomics* **272**, 550–561
 9. Martinez-Pastor, M., Marchler, G., Schüller, C., Marchler-Bauer, A., Ruis, H., and Estruch, F. (1996) The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO J.* **15**, 2227–2235
 10. Freitas, F. Z., Chapeaurouge, A., Perales, J., and Bertolini, M. C. (2008) A systematic approach to identify STRE-binding proteins of the *gsn* glycogen synthase gene promoter in *Neurospora crassa*. *Proteomics* **8**, 2052–2061
 11. McCluskey, K. (2003) The Fungal Genetics Stock Center: from molds to molecules. *Adv. Appl. Microbiol.* **52**, 245–262
 12. Colot, H. V., Park, G., Turner, G. E., Ringelberg, C., Crew, C. M., Litvinkova, L., Weiss, R. L., Borkovich, K. A., and Dunlap, J. C. (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10352–10357
 13. Vogel, H. J. (1956) A convenient growth medium for *Neurospora crassa* (medium N). *Microbiol. Genet. Bull.* **13**, 42–43
 14. Hardy, T. A., and Roach, P. J. (1993) Control of yeast glycogen synthase-2 by COOH-terminal phosphorylation. *J. Biol. Chem.* **268**, 23799–23805
 15. Hartree, E. F. (1972) Determination of protein: A modification of the Lowry method that gives a linear photometric response. *Anal. Biochem.* **48**, 422–427
 16. Sokolovsky, V., Kaldenhoff, R., Ricci, M., and Russo, V. E. A. (1995) Fast and reliable mini-prep RNA extraction from *Neurospora crassa*. *Fungal Genet. Newslett.* **37**, 41–43
 17. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning. A Laboratory Manual* (3rd ed.) Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY
 18. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
 19. Lange, A., Mills, R. E., Lange, C. J., Stewart, M., Devine, S. E., and Corbett, A. H. (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J. Biol. Chem.* **282**, 5101–5105
 20. Feng, B., and Marzluf, G. A. (1998) Interaction between major nitrogen regulatory protein NIT2 and pathway-specific regulatory factor NIT4 is required for their synergistic activation of gene expression in *Neurospora crassa*. *Mol. Cell. Biol.* **18**, 3983–3990
 21. Yamashiro, C. T., Ebbole, D. J., Lee, B. U., Brown, R. E., Bourland, C., Madi, L., and Yanofsky, C. (1996) Characterization of *rco-1* of *Neurospora crassa*, a pleiotropic gene affecting growth and development that encodes a homolog of Tup1 of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 6218–6228
 22. Peñalva, M. A., and Arst, H. N., Jr. (2002) Regulation of gene expression by ambient pH in filamentous fungi and yeasts. *Microbiol. Mol. Biol. Rev.* **66**, 426–446, Table of Contents
 23. Li, D., and Kolattukudy, P. E. (1997) Cloning of cutinase transcription factor 1, a transactivating protein containing Cys₆Zn₂ binuclear cluster DNA-binding motif. *J. Biol. Chem.* **272**, 12462–12467
 24. Li, D., Sirakova, T., Rogers, L., Ettinger, W. F., and Kolattukudy, P. E. (2002) Regulation of constitutively expressed and induced cutinase genes by different zinc finger transcription factors in *Fusarium solani* f. sp. *pisi* (*Nectria haematococca*). *J. Biol. Chem.* **277**, 7905–7912
 25. Rocha, A. L., Di Pietro, A., Ruiz-Roldán, C., and Roncero, M. I. (2008) Ctf1, a transcriptional activator of cutinase and lipase genes in *Fusarium oxysporum* is dispensable for virulence. *Mol. Plant Pathol.* **9**, 293–304
 26. Coppin, E. (2002) The *fle1* gene encoding a C₂H₂ zinc finger protein co-ordinates male and female sexual differentiation in *Podospora anserina*. *Mol. Microbiol.* **43**, 1255–1268
 27. Adams, T. H., Wieser, J. K., and Yu, J. H. (1998) *Asexual sporulation in Aspergillus nidulans*. *Microbiol. Mol. Biol. Rev.* **62**, 35–54
 28. van Peij, N. N., Gielkens, M. M., de Vries, R. P., Visser, J., and de Graaff, L. H. (1998) The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. *Appl. Environ. Microbiol.* **64**, 3615–3619
 29. Sellitrennikoff, C. P., Nelson, R. E., and Siegel, R. W. (1974) Phase-specific genes for macroconidiation in *Neurospora crassa*. *Genetics* **78**, 679–690
 30. Lambreghts, R., Shi, M., Belden, W. J., Decaprio, D., Park, D., Henn, M. R., Galagan, J. E., Bastürkmen, M., Birren, B. W., Sachs, M. S., Dunlap, J. C., and Loros, J. J. (2009) A high-density single nucleotide polymorphism map for *Neurospora crassa*. *Genetics* **181**, 767–781
 31. Yang, X., Talibi, D., Weber, S., Poisson, G., and Raymond, M. (2001) Functional isolation of the *Candida albicans* FCR3 gene encoding a bZip transcription factor homologous to *Saccharomyces cerevisiae* Yap3p. *Yeast* **18**, 1217–1225
 32. Wang, M., Wang, Q., Zhao, H., Zhang, X., and Pan, Y. (2009) Evolutionary selection pressure of forkhead domain and functional divergence. *Gene* **432**, 19–25
 33. Bien, C. M., and Espenshade, P. J. (2010) Sterol regulatory element binding proteins in fungi: hypoxic transcription factors linked to pathogenesis. *Eukaryot. Cell* **9**, 352–359
 34. Smith, R. L., and Johnson, A. D. (2000) Turning genes off by Ssn6-Tup1: a conserved system of transcriptional repression in eukaryotes. *Trends Biochem. Sci.* **25**, 325–330
 35. Wilson, W. A., Wang, Z., and Roach, P. J. (2002) Systematic identification of the genes affecting glycogen storage in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **1**, 232–242
 36. Dunlap, J. C., Borkovich, K. A., Henn, M. R., Turner, G. E., Sachs, M. S., Glass, N. L., McCluskey, K., Plamann, M., Galagan, J. E., Birren, B. W., Weiss, R. L., Townsend, J. P., Loros, J. J., Nelson, M. A., Lambreghts, R., Colot, H. V., Park, G., Collopy, P., Ringelberg, C., Crew, C., Litvinkova, L., DeCaprio, D., Hood, H. M., Curilla, S., Shi, M., Crawford, M., Koerhsen, M., Montgomery, P., Larson, L., Pearson, M., Kasuga, T., Tian, C., Baştürkmen, M., Altamirano, L., and Xu, J. (2007) Enabling a community to dissect an organism: overview of the *Neurospora* functional genomics project. *Adv. Genet.* **57**, 49–96
 37. Tilburn, J., Sarkar, S., Widdick, D. A., Espeso, E. A., Orejas, M., Mungroo, J., Peñalva, M. A., and Arst, H. N., Jr. (1995) The *Aspergillus* PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH. *EMBO J.* **14**, 779–790
 38. Tamayo, E. N., Villanueva, A., Hasper, A. A., de Graaff, L. H., Ramón, D., and Orejas, M. (2008) CreA mediates repression of the regulatory gene *xlnR* which controls the production of xylanolytic enzymes in *Aspergillus nidulans*. *Fungal Genet. Biol.* **45**, 984–993
 39. Fu, Y. H., and Marzluf, G. A. (1990) *nit-2*, the major nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain. *Mol. Cell Biol.* **10**, 1056–1065
 40. Lockington, R. A., Rodbourn, L., Barnett, S., Carter, C. J., and Kelly, J. M. (2002) Regulation by carbon and nitrogen sources of a family of cellulases in *Aspergillus nidulans*. *Fungal Genet. Biol.* **37**, 190–196
 41. Párraga, A., Bellsollell, L., Ferré-D'Amaré, A. R., and Burley, S. K. (1998) Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure* **15**, 661–672
 42. Hughes, A. L., Todd, B. L., and Espenshade, P. J. (2005) SREBP pathway responds to sterols and functions as an oxygen sensor in fission yeast. *Cell* **120**, 831–842
 43. Lane, S., Zhou, S., Pan, T., Dai, Q., and Liu, H. (2001) The basic helix-loop-

- helix transcription factor Cph2 regulates hyphal development in *Candida albicans* partly via TEC1. *Mol. Cell. Biol.* **21**, 6418–6428
44. Willger, S. D., Puttikamonkul, S., Kim, K. H., Burritt, J. B., Grahl, N., Metzler, L. J., Barbuch, R., Bard, M., Lawrence, C. B., and Cramer, R. A., Jr. (2008) A sterol-regulatory element binding protein is required for cell polarity, hypoxia adaptation, azole drug resistance, and virulence in *Aspergillus fumigatus*. *PLoS Pathog.* **4**, e1000200
45. Bien, C. M., Chang, Y. C., Nes, W. D., Kwon-Chung, K. J., and Espenshade, P. J. (2009) Cryptococcus neoformans Site-2 protease is required for virulence and survival in the presence of azole drugs. *Mol. Microbiol.* **74**, 672–690
46. Olmedo, M., Navarro-Sampedro, L., Ruger-Herreros, C., Kim, S. R., Jeong, B. K., Lee, B. U., and Corrochano, L. M. (2010) A role in the regulation of transcription by light for RCO-1 and RCM-1, the *Neurospora* homologs of the yeast Tup1-Ssn6 repressor. *Fungal Genet. Biol.* **47**, 939–952
47. Brody, S., Oelhafen, K., Schneider, K., Perrino, S., Goetz, A., Wang, C., and English, C. (2010) Circadian rhythms in *Neurospora crassa*: Downstream effectors. *Fungal Genet. Biol.* **47**, 159–168
48. Shinohara, M. L., Correa, A., Bell-Pedersen, D., Dunlap, J. C., and Loros, J. J. (2002) *Neurospora* clock-controlled gene (*ccg-9*) encodes trehalose synthase: circadian regulation of stress responses and development. *Eukaryot. Cell* **1**, 33–43
49. Zheng, X., and Sehgal, A. (2010) AKT and TOR signaling set the pace of the circadian pacemaker. *Curr. Biol.* **20**, 1203–1208
50. Doi, R., Oishi, K., and Ishida, N. (2010) CLOCK regulates circadian rhythms of hepatic glycogen synthesis through transcriptional activation of Gys2. *J. Biol. Chem.* **285**, 22114–22121
51. White, M. A., Riles, L., and Cohen, B. A. (2009) A systematic screen for transcriptional regulators of the yeast cell cycle. *Genetics* **181**, 435–446
52. Lundin, M., Nehlin, J. O., and Ronne, H. (1994) Importance of a flanking AT-rich region in target site recognition by the CG box-binding zinc finger protein MIG1. *Mol. Cell. Biol.* **14**, 1979–1985
53. Arst, H. N., and Peñalva, M. A. (2003) pH regulation in *Aspergillus* and parallels with higher eukaryotic regulatory systems. *Trends Genet.* **19**, 224–231

Anexo II

Glycogen Metabolism Regulation 1 in *Neurospora crassa* 2 [AU1](#)

Maria Célia Bertolini, Fernanda Zanolli Freitas, Renato Magalhães de Paula, 3 [AU2](#)
Fernanda Barbosa Cupertino, and Rodrigo Duarte Gonçalves 4 [AU3](#)

Abstract Microorganisms accumulate glycogen as carbon and energy reserves 5
to face adverse environmental conditions during growth and development. The 6
processes of glycogen synthesis and degradation share similarities among different 7
microorganisms. However, the regulation of the metabolism as a whole shows 8
differences, likely due to the environmental conditions to which they individually 9
respond. This chapter aims to present some molecular mechanisms that regulate 10
glycogen metabolism in the fungus *Neurospora crassa*. The availability of its 11
genome sequence (Galagan et al., Nature 422:859–868, 2003) and a collection of 12 [AU4](#)
mutant strains, each carrying a deletion in a specific ORF, allowed investigation 13
into the role of specific proteins as regulators of glycogen metabolism to begin. 14
Here we present some biochemical and molecular mechanisms that have already 15
been described for this fungus, and additionally, we focused on more recent 16
findings including the molecular basis underlying the metabolism regulation, 17
mainly at transcriptional level. 18

1 Introduction 19

Glycogen is a polymer of glucose which is widely distributed in nature, being 20
found in microorganisms all the way through to higher eukaryotes including 21
plants and animals. The glycogen structure is characterized by glucose units 22

M.C. Bertolini (✉) • F.Z. Freitas • F.B. Cupertino • R.D. Gonçalves
Departamento de Bioquímica e Tecnologia Química, Instituto de Química, UNESP, R. Prof.
Francisco Degni, 55, 14800-900 Araraquara, São Paulo, Brazil
e-mail: mcbertol@iq.unesp.br; fzfreitas@iq.unesp.br; fernanda_cupertino@yahoo.com.br;
rdgoncalves@gmail.com

R.M. de Paula
Nalco Company, 7705 Hwy 90A, Sugar Land, TX 77478, USA
e-mail: rdepaula@nalco.com

23 linked by α -1,4 linear glycosidic bonds and α -1,6-linked glucose at the branching
24 points. One great advantage of glycogen being used as a reserve carbohydrate
25 is that this molecule has little effect on the cell's osmotic pressure. Together
26 with starch in plants, glycogen is considered the main intracellular carbon and
27 energy storage molecule. Most of the biochemical and molecular studies of
28 glycogen metabolism regulation were performed in mammalian cells and in
29 the yeast *Saccharomyces cerevisiae*. In mammalian cells, the liver and skeletal
30 muscle cells are the main depository of glycogen. Yeast cells accumulate glyco-
31 gen in the diauxic phase of growth, or in response to a limitation of carbon,
32 nitrogen, sulfur or phosphorus, and they hydrolyze it under conditions of carbon
33 starvation (Johnston and Carlson 1992; François and Parrou 2001). The fungus
34 *Neurospora crassa*, the focus of this chapter, accumulates glycogen during
35 the exponential growth phase and degrades when the growth rate decreases
36 (de Paula et al. 2002). Although *N. crassa* and *S. cerevisiae* accumulate glycogen,
37 they differ in the way they regulate their intracellular storage under environmental
38 conditions. For example, under stressful environmental conditions, such as heat
39 shock, *Neurospora* degrades glycogen while the yeast accumulates glycogen.
40 The genome sequence of *Neurospora* has been completed (Galagan et al. 2003)
41 and comparing the multiple filamentous fungi genomic sequences available
42 with that of the yeast has revealed how divergent they are at the genomic level.
43 An interesting feature of *N. crassa* is the high number of genes without identifi-
44 cable homologues to known proteins ("orphan" genes) (Galagan et al. 2005); only
45 approximately 40% of the proteins have been functionally annotated in the
46 *N. crassa* genome database ([http://www.broadinstitute.org/annotation/genome/
47 neurospora/MultiHome.html](http://www.broadinstitute.org/annotation/genome/neurospora/MultiHome.html)) (Wang et al. 2011). This information indicates
48 how dynamic the fungal genomes are compared to other microorganisms,
49 and this stimulates the investigation of specific aspects of cell biology in a
50 particular organism. Many reviews on glycogen metabolism have been published
51 focusing on the yeast *S. cerevisiae* and mammalian cells (François and Parrou
52 2001; Roach et al. 2001; Wilson et al. 2010); while this chapter describes the
53 data we have obtained with the fungus *N. crassa* and compares them to the
54 yeast system.

55 2 Glycogen Synthesis and Degradation

56 Synthesis of glycogen involves three steps, which are: initiation, elongation, and
57 branching; and requires the activities of glycogenin, glycogen synthase, and the
58 branching enzyme, respectively. Degradation of glycogen requires the activities of
59 glycogen phosphorylase and the debranching enzyme. An overall schematic repre-
60 sentation is shown in Fig. 1.

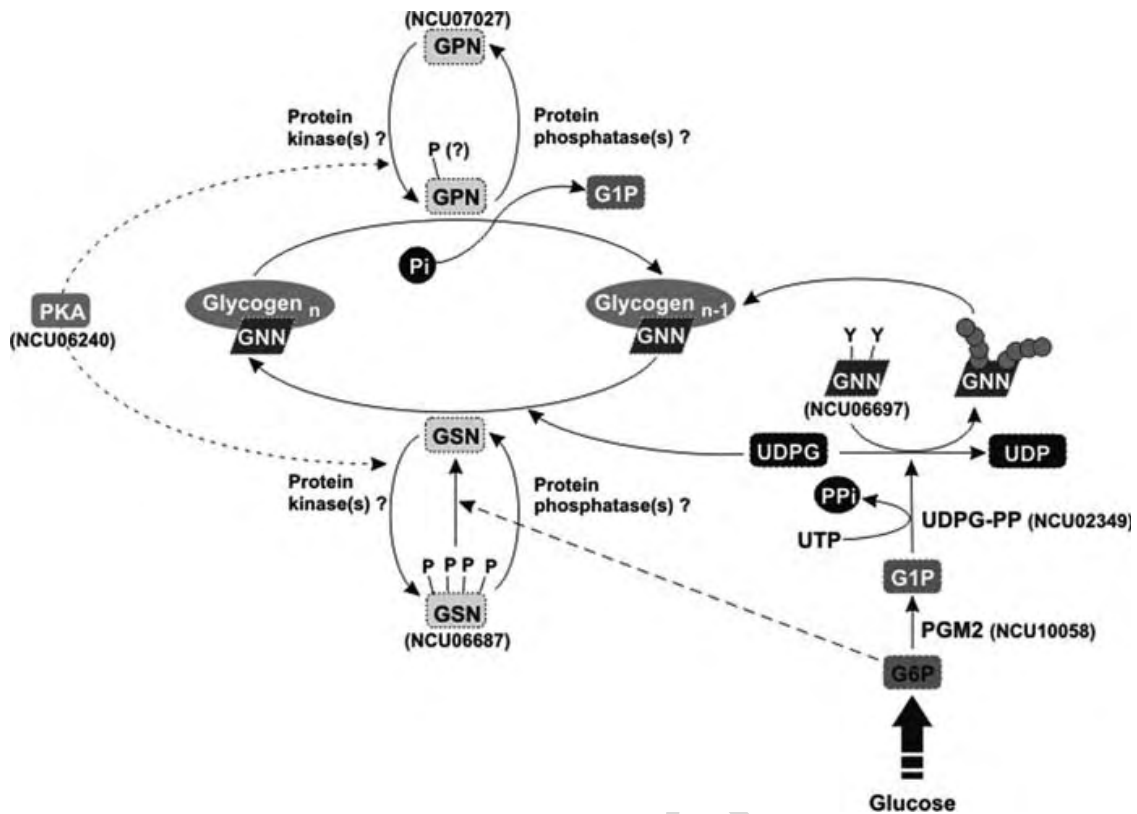


Fig. 1 Schematic representation of glycogen synthesis and degradation. The *N. crassa* orthologs are represented by their respective ORFs (<http://www.broadinstitute.org/annotation/genome/neurospora/MultiHome.html>). Glucose enters the cell and is converted to UDPG, the glucose donor for synthesis. GNN, the initiator of glycogen synthesis, is self-glucosylated, providing the non-reducing ends that ensure the glycogen elongation by the GSN enzyme through the formation of the α -1,4-glycosidic linkages, using UDPG as glucose donor. GPN enzyme catalyzes the phosphorolysis of glycogen yielding G1P and shortened glycogen as product. GSN and GPN are both controlled by reversible phosphorylation by the action of protein kinases and protein phosphatases. The four putative phosphorylation sites in GSN are shown. *G6P* glucose-6-P, *PGM2* phosphoglucomutase 2, *G1P* glucose-1-P, *UDPG-PP* uridine 5'-diphosphoglucose pyrophosphorylase, *UDPG* uridine 5'-diphosphoglucose, *GNN* *N. crassa* glycogenin, *GSN* *N. crassa* glycogen synthase, *GPN* *N. crassa* glycogen phosphorylase, *PKA* cyclic AMP-dependent protein kinase

2.1 Glycogen Initiation

61

The de novo synthesis of glycogen has been the subject of intense study over 62 decades. The question on how new glycogen molecules start to be synthesized 63 came from a first study reported by Krisman and Barengo (1975). These authors 64 identified a glucan-protein complex composed of a protein moiety attached to an α - 65 1,4 glucan chain and proposed that this protein might function as an initiator 66 molecule. Later, a protein that was covalently attached to the glycogen molecule 67 and co-purified with glycogen synthase was identified (Alonso et al. 1995). 68 The initiator protein was named later as glycogenin (Viskupic et al. 1992). One 69 of the most fascinating characteristics of glycogenin is its ability to self-glucosylate 70 using UDP-glucose (UDPG) as the glucan donor, a reaction highly stimulated by 71

72 Mn^{2+} ions (Lomako et al. 1988). A glycosidic bond is formed between the anomeric
73 C1 of the glucose moiety derived from UDPG and the hydroxyl oxygen of a
74 tyrosine side-chain of glycogenin. The glucosylated tyrosine in the rabbit muscle
75 glycogenin was identified as Tyr194 (Smythe et al. 1988) and mutations in this
76 residue resulted in a protein unable to self-glucosylate (Cao et al. 1993a, b).
77 Glycogenin then catalyzes glucosylation at C-4 of the attached glucose once again
78 (using UDPG as donor) to yield an O-linked disaccharide with the α -1,4 glycosidic
79 linkage. This is repeated until a short linear glucose polymer (8–12 unities) with
80 α -1,4 glycosidic linkages is built up on the glycogenin. Thus, glycogenin possesses
81 two distinct enzymatic actions: the former C-1-O-tyrosyl and the subsequent α -1,4
82 glycosyl activities. Additionally, glycogenin is able to transfer glucose residues to
83 a number of small acceptors in a process named trans-glucosylation (Smythe et al.
84 1990). However, the physiological significance of this process is not really under-
85 stood. The three-dimensional structure of the rabbit skeletal muscle glycogenin has
86 provided insight into understanding the catalytic action of this complex enzyme
87 (Gibbons et al. 2002). The model revealed the dimeric nature of the protein but
88 the mechanism of attachment of the first glucose residue is not yet fully understood.
89 From the structure, it was suggested that transfer of the first residue is via an
90 intermolecular reaction whereas the transfer of subsequent residues may be achieved
91 by an intramolecular reaction.

92 The yeast *S. cerevisiae* has two glycogenin isoforms, which are products of the
93 genes *glg1* and *glg2* (Cheng et al. 1995). The Glg1p isoform carries only one self-
94 glucosylating Tyr residue (Tyr²³²) while the Glg2p isoform can be modified into
95 two adjacent residues (Tyr²³⁰ and Tyr²³²) (Mu et al. 1996). Cells lacking either one
96 of these genes can accumulate glycogen at levels comparable to wild-type cells,
97 indicating that these genes are redundant in function. The absence of both genes
98 completely abolishes the glycogen accumulation in cells under normal conditions.
99 However, significant amounts of glycogen can be found in cells where glycogen
100 synthase activity has been enhanced, suggesting the existence of alternative primers
101 for the glycogen synthesis in the absence of a functional glycogenin (Torija et al.
102 2005). An auxiliary protein called GNIP (for glycogenin interacting protein)
103 has been identified in yeast cells (Skurat et al. 2002), which enhances the
104 self-glucosylation reaction upon the binding of GNIP (Zhai et al. 2004).

105 In the filamentous fungus *N. crassa*, only one isoform of glycogenin (GNN) has
106 been identified (de Paula et al. 2005a) and gene inactivation by RIP (repeat-induced
107 point mutation) completely abolished the accumulation of glycogen in this organ-
108 ism, suggesting that GNN is likely the only glycogen initiator present in *N. crassa*
109 cells. This protein is unusually long (664 amino acids) and all domains required
110 for the self- and trans-glucosylation activities are located within the 300 amino
111 acids in the N-terminal region (de Paula et al. 2005a). However, the long C-terminal
112 extension seems to be important to enhance the interaction with the glycogen
113 synthase (GSN) enzyme as determined by yeast two-hybrid assays (de Paula
114 et al. 2005b). Furthermore, expression of a truncated form of GNN containing
115 only the first 360 amino acids was enough to rescue the glycogen deficient
116 phenotype in yeast cells, indicating that the truncated protein is fully active

AU5

(de Paula et al. 2005a). GNN has two glucosylation sites – Tyr196 and Tyr198; however, each residue contributes differently to the self-glucosylation process (de Paula et al. 2005b). Tyrosine 196 is the major glucosylation site as mutation of this residue to Phe resulted in the accumulation of only 30% of total glycogen, compared to the wild-type protein. It is unclear whether both sites are glucosylated at the same time or if Tyr198 represents an alternative glucosylation site which can be modified when Tyr196 is unavailable. The regulation of glycogenin expression has been poorly understood and only a few studies on this subject have been reported. In *N. crassa*, GNN expression seems to be regulated by nutrient availability and stress (de Paula et al. 2005a).

Once glucosylated, glycogenin acts as substrate for glycogen synthase (GS), which will elongate the glucose chain. It is speculated that the length of the oligosaccharide chain in glycogenin necessary to allow the proper elongation by GS is determined by the balance between the activities of glycogen phosphorylase and glycogen synthase (Cao et al. 1993a, b). Direct interactions by the two-hybrid approach between glycogenin and glycogen synthase (Pitcher et al. 1987) confirmed the physical interactions between these two proteins (Skurat et al. 2006; de Paula et al. 2005b).

2.2 Glycogen Maturation

Primed glycogenin will serve as the substrate for the two enzymes involved in the maturation process of glycogen particles: glycogen synthase and branching enzyme, which are the proteins that catalyze the formation of α -1,4 and α -1,6 glycosidic bonds, respectively. Similar to glycogenin, GS uses UDP-glucose as the donor of the glucose residues. Once GS extends the glycosidic chain to a certain length, the branching enzyme will transfer a string of approximately seven glucoses from the nascent chain to the glucose C-6 in an adjacent chain, thus creating a ramification point. These ramifications occur, equidistantly, every 10–14 glucose residues apart. Bacterial glycogen synthase differs from its eukaryotic counterparts, due to the fact that they utilize ADP-glucose as the glucan donor (Preiss and Romeo 1994) and they lack regulation.

Overall, all glycogen synthases are conserved proteins among microbes and higher organisms and the differences are located mostly in the N- and C-termini of the protein, where the regulatory phosphorylation sites are located. *S. cerevisiae* contains two GS isoforms, Gsy1p and Gsy2p, in which the latter accounts for 80% of the total GS activity in the cells (Farkas et al. 1990, 1991). Three phosphorylation sites were found in the C-terminus of this enzyme (Hardy and Roach 1993). The kinases PKA, Snf1p, Yak1, Psk2, and Pho85 were described as involved in phosphorylation and regulation of Gsy2p, either directly or indirectly (Cameron et al. 1988; Wang et al. 2001; Huang et al. 1998, reviewed in Wilson et al. 2010).

In *N. crassa*, one GS isoform was identified (GSN), which shared much conservation with the counterparts in yeast and mammals (de Paula et al. 2002). Although

158 the mechanisms of GSN regulation have not been completely unraveled, there is
159 evidence that the PKA signaling pathway exerts a very important role in the control
160 of its activity. Strains defective in the PKA signaling pathway showed impairments
161 in accumulation of glycogen and in *gsn* expression. The in vitro studies showed
162 that the PKA pathway also influences the GSN phosphorylation status (Freitas et al.
163 2010). However, GS is not an essential enzyme; *gsn* inactivation by RIP (Repeat
164 Induced Point Mutation) does not seem to impair growth and development of the
165 mutant strain (unpublished results). Four putative phosphorylation sites were
166 identified in GSN (S632, S636, T641, and T645) based on a sequence alignment
167 of different GS enzymes, all located at the C-terminus. Systematic site-directed
168 mutagenesis of the four sites was carried out, mutating all residues to alanine.
169 In vitro incorporation of radioactive phosphate by the mutant proteins produced in
170 *E. coli* using cellular extract from a non-synthesizing GSN strain confirmed that all
171 residues can be phosphorylated. However, compared to the yeast enzyme Gsy2p,
172 GSN seems to have an additional phosphorylation site, since a truncated protein
173 that was missing the C-terminal region containing the putative phosphorylation
174 sites was still able to incorporate radioactive phosphate (unpublished results).

175 Among the phosphatases involved in the dephosphorylation and activation of
176 glycogen synthases in microorganisms, PP1A seems to be the main player in yeast
177 cells. Mutant alleles in *GLC7*, which encodes for the catalytic subunit of PP1A,
178 displayed reduced glycogen accumulation, consistent with its inability to dephos-
179 phorylate Gsy2p (Cannon et al. 1994). Four different PP1A regulatory subunits
180 (Gac1p, Pig1p, Pig2p and Gip2p) were found to be associated with Gsy2p, indicating
181 that the specificity of the reaction is dictated by proper association with these
182 subunits (Cheng et al. 1997). Although the protein phosphatases involved in
183 the dephosphorylation of GSN in *Neurospora* have not been identified yet, *GLC7*
184 and *GAC1* are conserved in this organism.

185 Only recently, the tri-dimensional structure of a eukaryotic GS was determined
186 and thus provided the first insights into the role of the activator glucose-6-phosphate
187 (Baskaran et al. 2011). Crystal structures of the basal activity state and those in
188 the glucose-6-phosphate-activated form of the yeast protein (Gsy2p) showed an
189 unusual tetramer assembling. Binding of glucose-6-phosphate induced conforma-
190 tional changes of the subunits leading to increased catalytic efficiency. Also,
191 site-directed mutagenesis demonstrated that the residues Arg583 and Arg587 are
192 necessary and sufficient for glucose-6-phosphate activation.

193 2.3 Glycogen Degradation

194 The breakdown of glycogen particles is accomplished by the action of glycogen
195 phosphorylase and debranching enzymes. Glycogen phosphorylase releases
196 glucose-1-phosphate from a terminal α -1,4 glycosidic bond, which then is con-
197 verted to glucose-6-phosphate by the action of phosphoglucomutase before entering
198 the catabolic pathway. The debranching enzyme carries out two distinct enzymatic

activities: glucosyltransferase – the transfer of three glucose residues from one 199
branch to another, and glucosidase – the breaking of α -1,6 glycosidic bonds. Similar 200
to glycogen synthase, glycogen phosphorylase is regulated by reversible phosphor- 201
ylation and allosteric effectors, mainly AMP. 202

In *N. crassa*, the two forms of glycogen phosphorylase and their dependency on 203
the allosteric modulators have been described many years ago (Télliez-Iñón and 204
Torres 1970). Moreover, the activities of glycogen synthase and glycogen phosphor- 205
ylase oppose each other during stress conditions, favoring a breakdown of glycogen 206
to allow enough energy availability for survival (Noventa-Jordão et al. 1996). 207

3 Regulation of Glycogen Metabolism 208

Regulation of glycogen metabolism has been studied for decades focusing on the 209
activities of the enzymes catalyzing its synthesis and degradation, which are 210
glycogen synthase and glycogen phosphorylase, respectively. The main control 211
point is thought to be at glycogen synthase level, which is subject to multiple forms 212
of regulation. The two enzymes are regulated by reversible covalent modification, 213
in which phosphorylation activates glycogen phosphorylase and inhibits glycogen 214
synthase (Télliez-Iñón et al. 1969; Fletterick and Madsen 1980). In addition, they 215
are also regulated by allosterism, where glucose-6-phosphate and AMP are the 216
allosteric effectors of glycogen synthase and glycogen phosphorylase, respectively. 217
Glucose-6-phosphate reverses the glycogen synthase inactivation by phosphoryla- 218
tion and AMP is the allosteric activator for the dephosphorylated glycogen phos- 219
phorylase. The ratio between the activities in the absence and in the presence of the 220
allosteric effectors is an index of the phosphorylation state for both enzymes. 221
Multiple phosphorylation sites were identified in glycogen synthases, which are 222
phosphorylated by different protein kinases, depending on the organism, whereas 223
glycogen phosphorylase is phosphorylated in a single residue, Ser14, which is 224
modified by the phosphorylase kinase protein. 225

Besides reversible changes in the regulatory enzymes activities, glycogen levels 226
are also correlated with physiological conditions through control of gene expression 227
mainly at transcriptional level of the gene encoding glycogen synthase. Environ- 228
mental conditions indirectly regulate glycogen metabolism by activating signaling 229
pathways and leading to either activation or repression of gene expression. 230

3.1 *Environmental Conditions Regulating Glycogen Accumulation in N. crassa: Regulation of the gsn Gene Expression* 231 [AU6](#)

232
233

N. crassa accumulated glycogen late in the exponential phase of the vegetative 234
growth (around 24 h) and degraded it at the beginning of the stationary phase 235

236 (de Paula et al. 2002). At this time, *gsn* expression was at its maximum. In addition,
237 glycogen levels were highly regulated on exposure of cultures to some stress situa-
238 tions, such as heat shock (transfer from 30°C to 45°C) and carbon source limitation
239 (sugar-free medium) (de Paula et al. 2002). Trehalose is another reserve carbohy-
240 drate that can be mobilized under different growth conditions. However, while
241 *N. crassa* degraded glycogen under heat shock, trehalose was accumulated
242 (de Pinho et al. 2001; Noventa-Jordão et al. 1996). Mycelia exposed to heat
243 shock and carbon starvation presented changes in the expression of the gene
244 encoding glycogen synthase (*gsn*), the glycogen accumulated and the glycogen
245 synthase activity. There was a decrease in gene expression and a concomitant fall in
246 the levels of glycogen and glycogen synthase activity. However, the *gsn* expression
247 and glycogen levels were recovered when the cultures were returned to normal
248 growth conditions (30°C and 2% sugar) (de Paula et al. 2002). These results
249 suggested that transcriptional regulation may account for the decrease in glycogen
250 synthase activity and subsequent glycogen mobilization observed under these
251 conditions. On the other hand, glycogen phosphorylase was activated under heat
252 shock showing that reversible changes in the two regulatory enzymes were observed
253 upon temperature shifting (Noventa-Jordão et al. 1996).

254 The yeast *S. cerevisiae* exhibits opposite responses with respect to the environ-
255 mental condition of heat shock. Compared to *N. crassa*, transcription of the genes
256 encoding glycogen synthase *GSY1* and *GSY2* was induced (Unnikrishnan et al.
257 2003; Ni and LaPorte 1995) and glycogen was accumulated under a heat-shock
258 condition (Parrou et al. 1997). *GSY1* and *GSY2* mRNA induction was mediated
259 by the STRE (STress Responsive Elements) motifs present in the promoter regions
260 of these genes (Ni and LaPorte 1995; Unnikrishnan et al. 2003; Enjalbert et al.
261 2004). STRE motifs are DNA regulatory elements found in the promoter regions of
262 genes responsive to different stressing conditions in *S. cerevisiae* (Martinez-Pastor
263 et al. 1996). This cis-regulatory element is trans-activated by two yeast C₂H₂ zinc-
264 finger transcription factors, which are the proteins Msn2p and Msn4p (Estruch and
265 Carlson 1993; Schmitt and McEntee 1996). These two proteins quickly translocate
266 to the nucleus after a stressful stimulus in a PKA-dependent way (Smith et al. 1998;
267 Gorner et al. 1998).

268 The decrease in glycogen content observed in *N. crassa* cells exposed to heat
269 stress may result from the down regulation of the *gsn* gene probably mediated by
270 the STRE motif within the promoter region (Freitas and Bertolini 2004). The *gsn*
271 gene has two STRE motifs, one upstream from a canonical TATA-box and another
272 in an intron within the 5'-UTR. DNA fragments containing both motifs were
273 specifically bound by nuclear proteins activated by heat shock (Freitas and Bertolini
274 2004). However, Msn2/4p orthologue proteins were not identified in the *N. crassa*
275 database suggesting the existence of a different mechanism to regulate the heat
276 shock response (Freitas et al. 2008). Whether STRE mediates *gsn* down regulation,
277 it is possible to speculate that in *N. crassa* STRE motifs are involved in gene
278 repression instead of gene activation. Although the STRE motifs in *S. cerevisiae* are
279 frequently correlated with the activation of gene expression, the role of the STRE
280 motif as a transcriptional repressor has also been reported in literature (de Groot
281 et al. 2000; Vyas et al. 2005).

Another DNA cis-regulatory element that might be involved in glycogen accumulation in *N. crassa* by modulating the transcription of the *gsn* gene during heat shock is the HSE (Heat Shock Elements) motif, which is recognized and bound by the HSFs (Heat Shock Factors) transcription factors (Bienz and Pelham 1986). Yeast HSFs are constitutively bound to HSE motifs maintaining the basal levels of gene transcription even under normal temperature of growth. Heat stress only increases the transcriptional activity by promoting Hsf1p trimerization (Sorger et al. 1987; Jakobsen and Pelham 1988). In multicellular organisms, the HSFs bind to the consensus HSE only after heat stress has occurred (Kingston et al. 1987). In this context, *N. crassa* is similar to the yeast since gel shift analysis using nuclear extracts and DNA fragments containing HSE motifs showed that the proteins (including HSFs) are constitutively bound to the cis HSEs before heat shock (Meyer et al. 2000; Freitas and Bertolini 2004).

Attempts to identify the transcription factor(s) that bind to the STRE motif of the *N. crassa gsn* promoter were performed by gel shift assay experiments and mass spectrometry analysis. Five putative protein candidates acting as transcriptional regulators of the *gsn* gene expression during heat stress were identified. Analyses of their polypeptide sequences revealed the presence of protein domains usually found in proteins involved with transcription regulation. Among the proteins identified, two are noteworthy (Freitas et al. 2008). One is the protein encoded by the ORF NCU03482, annotated as a RuvB-like helicase 1 and belonging to a protein family having the AAA (ATPases Associated with diverse cellular Activities) and Tip49 (TBP-Interacting Protein 49) domains. Although the function of this protein family is not totally clear, they are supposed to play an important role in nuclear events, such as chromatin remodeling and regulation of transcription, since they interact with many nuclear proteins and show DNA helicase activity (Bellosta et al. 2005; Wood et al. 2000). The three-dimensional structure of the *N. crassa* protein was determined by molecular modeling based on the RUVBL1 structure (PDB 2c9o) from *Homo sapiens* (Matias et al. 2006). The monomer contains three domains, of which two are involved in ATP binding and one is a DNA/RNA-binding domain (unpublished results).

The other protein is the one encoded by the ORF NCU06679 annotated as histone acetyltransferase type B subunit 2, and previously annotated as a chromatin assembly factor subunit orthologue to the yeast Cac3p. This protein possesses five WD-40 domains within its polypeptide chain. The WD-40 domains are short tandem repetitions with approximately 40 amino acid residues bearing the GH dipeptide having 11–24 residues from its N-terminus and the WD dipeptide at its C-terminus (Smith et al. 1999). WD-repeat proteins are a large family found in all eukaryotes and are implicated in a variety of functions ranging from signal transduction and transcription regulation to cell cycle control, autophagy and apoptosis (Neer et al. 1994). The underlying common function of all WD-repeat containing proteins is coordinating multiprotein complex assemblies. The three-dimensional structure of the *N. crassa* protein was determined by molecular modeling based on the Nurf55 structure (PDB 2xyi) from *Drosophila melanogaster* (Nowak et al. 2011), a component of different chromatin-modifying complexes. Nurf55 forms a seven-bladed β -propeller characteristic for the WD40 family of proteins (unpublished results).

328 **3.2 Transcription Factors Regulating Glycogen Accumulation** 329 **in *N. crassa***

330 All living organisms need an appropriate gene regulation to express specific parts of
331 their genomes to execute critical biological functions during development and
332 differentiation, and to respond to different environmental signals. Gene transcrip-
333 tion, a key step in gene expression, is a dynamic process that involves distinct steps.
334 One is the recruitment of protein complexes to target genes, which associate with
335 specific DNA-binding sites in response to tightly controlled physiological signaling
336 cascades. These proteins fall into several classes, with the transcription factors
337 being the class of sequence-specific DNA binding proteins that mediate gene-
338 specific transcriptional activation or repression. Thus, transcription factors consti-
339 tute critical regulatory molecules that, upon a cell's signal, rapidly and transiently
340 bind to specific cis-regulatory elements (or DNA response elements) in the chro-
341 matin, in a process associated with chromatin remodeling (Hager et al. 2009).

342 The release of the complete *N. crassa* genome (Galagan et al. 2003) and the
343 establishment of a methodology for gene knockout by gene replacement by
344 Ninomiya et al. (2004) allowed the construction of a collection of mutant strains
345 each containing single-gene deletions to start. The availability of a mutant strain
346 subset with each strain knocked-out in genes encoding known or putative transcrip-
347 tion factors allowed us to perform screening aimed to investigate alterations in
348 glycogen accumulation in these mutant strains (Gonçalves et al. 2011). The glyco-
349 gen accumulated by the mutant strains was quantified under normal growth tem-
350 perature (30°C) and under heat shock stress (45°C) and compared to the wild-type
351 strain. The analysis under heat stress was performed since *N. crassa* exhibits an
352 opposite response when compared to the yeast *S. cerevisiae* concerning glycogen
353 accumulation under this environmental condition. Many transcription factors were
354 identified (Gonçalves et al. 2011, see Fig. 2) and most of them are annotated in the
355 *N. crassa* database as hypothetical proteins. The *gsn* and *gpn* (gene encoding
356 glycogen phosphorylase) expression was analyzed in all mutant strains grown
357 under the temperatures of 30°C (vegetative growth) and 45°C (heat shock). Many
358 of the mutant strains showed differences in gene expression when compared to the
359 wild-type strain, in which both *gsn* and *gpn* expression decreased after heat shock
360 (Fig. 2). In some mutant strains the gene expression could be correlated to the
361 amount of glycogen accumulated suggesting that the transcription factors would act
362 either directly on gene expression or indirectly, by regulating a gene whose product
363 affects *gsn* and *gpn* expression.

364 Many identified transcription factors are biochemically characterized proteins,
365 either in *N. crassa* or in other fungi. Among the transcription factors identified, it is
366 noteworthy to cite the pH regulator PacC (Peñalva and Arst 2002), the NIT2
367 nitrogen metabolism regulator (Fu and Marzluf 1990), the mycelial repressor of
368 conidiation gene expression RCO-1 (Yamashiro et al. 1996), the light-inducible
369 transcription factor CSP-1 in *N. crassa* (Lambreghts et al. 2009), the transcriptional
370 activator of genes required for the use of alternative carbon sources XlnR (van Peij

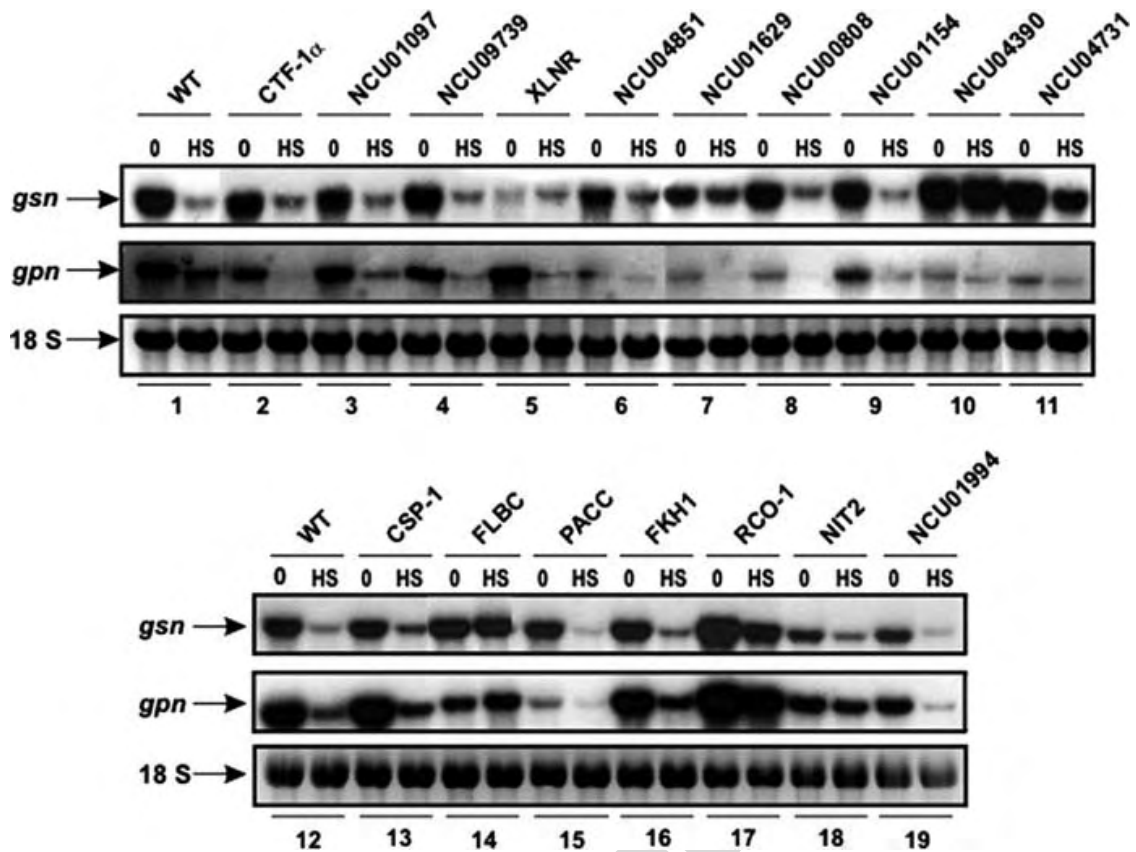
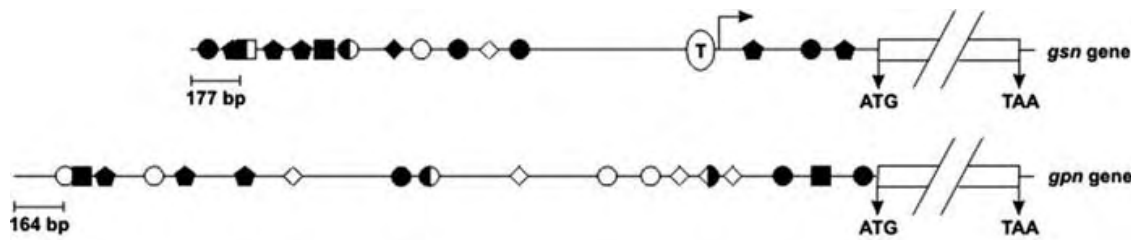


Fig. 2 Gene expression in *N. crassa* transcription-factor mutant strains. The expression of *gsn* (encoding glycogen synthase) and *gpn* (encoding glycogen phosphorylase) genes were analyzed in the transcription-factor mutant strains which exhibited a glycogen accumulation profile different from that shown by the wild-type strain. Cells from the mutant strains were cultivated for 24 h at 30°C and then subjected to a temperature shift of 45°C for 30 min. Culture samples were harvested before (0) and after the heat shock (HS) and the mycelia were used to extract total RNA. Total RNA was separated by electrophoresis in denaturing formaldehyde-agarose gels, transferred to neutral nylon membranes and the blots were probed with radio-labeled full-length *gsn* and *gpn* cDNAs. The 18 S rRNA stained with ethidium bromide was used as loading control. WT wild-type strain, *CTF-1 α* cutinase transcription factor 1 alpha mutant strain, *XLNR* transcriptional activator xlnR mutant strain, *CSP-1* conidial separation-1 mutant strain, *FLBC* protein FlbC mutant strain, *PACC* pH-response transcription factor pacC/RIM101 mutant strain, *FKH1* FKH1 protein mutant strain, *RCO-1* regulator of conidiation-1 mutant strain, *NIT2* nitrate nonutilizer-2 mutant strain. Proteins with no orthologues are represented by their respective ORFs numbers (<http://www.broadinstitute.org/annotation/genome/neurospora/MultiHome.html>)

et al. 1998), among others. Search for cis DNA motifs for the transcription factors 371
 by in silico analysis of the *gsn* and *gpn* 5'-flanking regions led to the identification 372
 of putative DNA-binding sites for some proteins (Fig. 3). This information led us to 373
 start investigating the role of some transcription factors as regulators in glycogen 374
 metabolism. One is the PacC/Rim101p, which has been extensively studied in 375
A. nidulans and *S. cerevisiae*, respectively, and plays an important role in the pH 376
 signaling pathway. This transcription factor mediates the cell adaptation to neutral- 377
 alkaline pH by activating genes that are preferentially expressed at alkaline pH 378
 and repressing those preferentially expressed at acidic pH (Tilburn et al. 1995). 379



Transcription Factors	<i>cis</i> DNA element	Positions in <i>gsn</i> promoter (nt)*	Positions in <i>gpn</i> promoter (nt)*
● Mig1	5'-GGGG-3'	-2305, -2026, -1592, -1487, -317	-1610, -1564, -618, -351, -22
◆ PacC	5'-CCAaG-3'	-1807	-
◇ NIT2	5'-TATCtac-3'	-1558	-2064, -1217, -660, -625, -594
■ XlnR	5'-GGCTGA-3'	-2071, -2252	-217, -2631
○ Sre1p	5'-TACA-3'	-1758, -2024	-2693, -2380, -1561, -825, -703
□ CTF1 α	5'-CCGAGG-3'	-2248	-
● Fkh	5'-AACA-3'	-2261, -2178, -2103, -948, -679, -214	-2548, -2327, -2115

* DNA element positions considering the ATG start codon. The Fkh DNA motif positioned at -948 is coincident with the Transcription Initiation Site (TIS) at the *gsn* promoter.

Fig. 3 Representation of the *gsn* and *gpn* genes 5'-flanking regions. The relative positions of the DNA motifs recognized by the transcription factors are indicated. The TATA-box sequence in the *gsn* gene is indicated by *T*. The ORFs are delimited by their ATG start codon and the TAA stop codon. The Transcription Initiation Site (TIS) in the *gsn* gene is represented by an *arrow*

380 In *A. nidulans*, PacC is activated by two successive proteolytic cleavage steps at the
 381 C-terminus, leading to the active protein PacC²⁷ that contains a DNA-binding
 382 domain capable of binding to the promoters of pH-regulated genes (reviewed in
 383 Arst and Peñalva 2003). The yeast protein appears to have a broader role than
 384 simply that of promoting alkaline pH-inducible responses (Lamb and Mitchell
 385 2003). In addition, it differs from *A. nidulans* protein since it requires only a single
 386 cleavage step to be activated (Li and Mitchell 1997) and while PacC acts as a
 387 transcriptional activator under alkaline pH, Rim101p exerts its role as a repressor
 388 (Lamb and Mitchell 2003). Thus, the molecular mechanisms involved in the pH
 389 response differ among organisms.

390 The existence of a DNA motif for the *A. nidulans* PacC in the promoter *gsn*
 391 prompted us to investigate whether this transcription factor regulates glycogen
 392 accumulation. We demonstrated that the pH signaling pathway regulates glycogen
 393 metabolism where the *N. crassa* PACC may play a central role. Cells of the wild-
 394 type strain grown at alkaline pH (7.8) showed reduction in *gsn* transcript levels
 395 and low intracellular glycogen accumulation. In contrast, the *pacC*^{KO} strain accu-
 396 mulated similar amounts of glycogen as the non-stressed wild-type strain and
 397 showed high *gsn* transcript levels at alkaline pH. Under the same conditions,
 398 *pacC* expression was up regulated. From these results, it was possible to charac-
 399 terize *gsn* as an acidic gene and to suggest a regulatory role for PACC in *gsn*
 400 expression. The recombinant PACC was produced in *E. coli* as a truncated protein

containing the DNA-binding domain and was able to bind in vitro to a *gsn* DNA 401
fragment containing the PacC motif. In addition, DNA-protein complexes were 402
formed with extracts from cells grown at normal and alkaline pH, and binding was 403
confirmed by CHIP analysis. An interesting result was that PACC present in 404
the extracts prepared from cells grown at normal and alkaline pH had the same 405
molecular mass, indicating that protein processing was not triggered by alkaline pH 406
in contrast to what is described for the *A. nidulans* protein. Our results for PACC 407
processing led us to suggest that PACC proteolysis may involve a mechanism that 408
is different from that described for the *A. nidulans* protein (unpublished results). 409

Many other transcription factors are being investigated at this moment for their 410
role as putative regulators of glycogen metabolism in *N. crassa*. The studies will 411
provide insights into the regulatory network that is involved in the molecular 412
mechanisms controlling glycogen metabolism in this microorganism. 413

4 Conclusions and/or Future Perspectives 414

Over the last decade, new findings in literature together with the availability of 415
numerous genome sequences have allowed advances in understanding how glyco- 416
gen metabolism is regulated in microorganisms. In addition, the establishment of 417
strategies for gene deletion has substantially accelerated the search for genes linked 418
to a particular phenotype. Regarding the fungus *N. crassa*, the majority of genes 419
were individually deleted by a consortium of laboratories resulting in the creation 420
of a deletion strain set that constitutes a powerful tool to start investigating the role 421
of specific proteins as regulators of glycogen metabolism. We have taken the 422
advantage of a transcription-factor deletion-strain collection to identify regulatory 423
proteins and interesting information is now becoming available. Most of the tran- 424
scription factors identified were annotated in the fungus database as hypothetical 425
proteins. However, some of them were biochemically well characterized either in 426
N. crassa or in other fungi, and some of them have DNA binding motifs in the *gsn* 427
and/or *gpn* 5'-flanking regions. The results suggest that glycogen metabolism could 428
be interconnected to multiple cellular processes in this fungus. 429

An interesting piece of information emerging from these studies was that some 430
transcription-factor mutant strains showing glycogen accumulation different from 431
the wild-type strain presented impairments in cell cycle progression, contributing to 432
the identification of connections between the core cell-cycle oscillator and cell- 433
cycle transcription. The fact that some transcription factors identified in our screen- 434
ing are light-inducible proteins allowed us to speculate on the existence of a 435
connection between circadian clocks and glycogen metabolism and also that circa- 436
dian rhythms and energy state might be linked. Although the identification of the 437
transcription factors did not mean that they have a direct regulatory role, they 438
constitute a valuable group of candidate proteins acting as regulators in glycogen 439
metabolism control. The results open up new opportunities for investigating key 440
questions concerning glycogen metabolism regulation, such as how glycogen 441

442 metabolism could be connected to cell-cycle regulation, the biological clock,
443 and other aspects of cellular metabolism. The understanding of such connections
444 will be valuable to understand the importance of the energy balance to biological
445 processes.

446 The work in the author's laboratory was supported by grants and fellowships
447 from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and CNPq
448 (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

449 References

- 450 Alonso MD, Lomako J, Lomako WM, Whelan WJ (1995) A new look at the biogenesis of
451 glycogen. *FASEB J* 9:1126–1137
- 452 Arst HN Jr, Peñalva MA (2003) pH regulation in *Aspergillus* and parallels with higher eukaryotic
453 regulatory systems. *Trends Genet* 19:224–231
- 454 Baskaran S, Chikwana VM, Contreras CJ, Davis KD, Wilson WA, DePaoli-Roach AA, Roach PJ,
455 Hurley TD (2011) Multiple glycogen-binding sites in eukaryotic glycogen synthase are
456 required for high catalytic efficiency toward glycogen. *J Biol Chem* 286:33999–34006
- 457 Bellosta P, Hulf T, Balla Diop S, Usseglio F, Pradel J, Aragnol D, Gallant P (2005) Myc interacts
458 genetically with Tip48/Reptin and Tip49/Pontin to control growth and proliferation during
459 *Drosophila* development. *Proc Nat Acad Sci USA* 102:11799–11804
- 460 Bienz M, Pelham HBR (1986) Heat shock regulatory elements function as an inducible enhancer
461 in the *Xenopus* hsp70 gene and when linked to a heterologous promoter. *Cell* 45:753–760
- 462 Cameron S, Levin L, Zoller M, Wigler M (1988) cAMP-independent control of sporulation,
463 glycogen metabolism and heat shock resistance in *S. cerevisiae*. *Cell* 53:555–566
- 464 Cannon JF, Pringle JR, Fiechter A, Khalil M (1994) Characterization of glycogen-deficient glc
465 mutants of *Saccharomyces cerevisiae*. *Genetics* 136:485–503
- 466 Cao Y, Mahrenholz AM, DePaoli-Roach AA, Roach PJ (1993a) Characterization of rabbit skeletal
467 muscle glycogenin. Tyrosine 194 is essential for function. *J Biol Chem* 268:14687–14693
- 468 Cao Y, Skurat AV, DePaoli-Roach AA, Roach PJ (1993b) Initiation of glycogen synthesis. Control
469 of glycogenin by glycogen phosphorylase. *J Biol Chem* 268:21717–21721
- 470 Cheng C, Mu J, Farkas I, Huang D, Goebel MG, Roach PJ (1995) Requirement of the self-
471 glucosylating initiator proteins Glg1p and Glg2p for glycogen accumulation in *Saccharomyces*
472 *cerevisiae*. *Mol Cell Biol* 15:6632–6640
- 473 Cheng C, Huang D, Roach PJ (1997) Yeast PIG genes: PIG1 encodes a putative type 1 phosphatase
474 subunit that interacts with the yeast glycogen synthase Gsy2p. *Yeast* 13:1–8
- 475 de Groot E, Bebelman JP, Mager WH, Planta RJ (2000) Very low amounts of glucose cause
476 repression of the stress responsive gene HSP12 in *Saccharomyces cerevisiae*. *Microbiology*
477 146:367–375
- 478 de Paula R, de Pinho CA, Terenzi HF, Bertolini MC (2002) Molecular and biochemical character-
479 ization of the *Neurospora crassa* glycogen synthase encoded by the gsn cDNA. *Mol Genet*
480 *Genomics* 267:241–253
- 481 de Paula RM, Wilson WA, Terenzi HF, Roach PJ, Bertolini MC (2005a) GNN is a self-
482 glucosylating protein involved in the initiation step of glycogen biosynthesis in *Neurospora*
483 *crassa*. *Arch Biochem Biophys* 435:112–124
- 484 de Paula RM, Wilson WA, Roach PJ, Terenzi HF, Bertolini MC (2005b) Biochemical characteri-
485 zation of *Neurospora crassa* glycogenin (GNN), the self-glucosylating initiator of glycogen
486 synthesis. *FEBS Lett* 10:2208–2214

Glycogen Metabolism Regulation in *Neurospora crassa*

de Pinho CA, Polizeli MLTM, Jorge JA, Terenzi HF (2001) Mobilization of trehalose in mutants of the cyclic AMP signaling pathway, cr-1 (CRISP-1) and mcb (microcycle conidiation), of <i>Neurospora crassa</i> . FEMS Microbiol Lett 199:85–89	487 488 489
Enjalbert B, Parrou JL, Teste MA, François J (2004) Combinatorial control by the protein kinases PKA, PHO85 and SNF1 of the transcriptional induction of the <i>Saccharomyces cerevisiae</i> GSY2 gene at the diauxic shift. Mol Genet Genomics 271:697–708	490 491 492
Estruch F, Carlson M (1993) Two homologous zinc finger genes identified by multicopy suppression in a SNF1 protein kinase mutant of <i>Saccharomyces cerevisiae</i> . Mol Cell Biol 13:3872–3881	493 494
Farkas I, Hardy TA, DePaoli-Roach AA, Roach PJ (1990) Isolation of the GSY1 gene encoding glycogen synthase and evidence for the existence of a second gene. J Biol Chem 265:20879–20886	495 496 497
Farkas I, Hardy TA, Goebel MG, Roach PJ (1991) Two glycogen synthase isoforms in <i>Saccharomyces cerevisiae</i> are coded by distinct genes that are differentially controlled. J Biol Chem 266:15602–15607	498 499 500
Fletterick RJ, Madsen NB (1980) The structures and related functions of phosphorylase a. Annu Rev Biochem 49:31–61	501 502
François J, Parrou JL (2001) Reserve carbohydrates metabolism in the yeast <i>Saccharomyces cerevisiae</i> . FEMS Microbiol Rev 25:125–145	503 504
Freitas FZ, Bertolini MC (2004) Genomic organization of the <i>Neurospora crassa</i> gsn gene: possible involvement of the STRE and HSE elements in the modulation of transcription during heat shock. Mol Genet Genomics 272:550–561	505 506 507
Freitas FZ, Chapeaurouge A, Perales J, Bertolini MC (2008) A systematic approach to identify STRE-binding proteins of the <i>gsn</i> glycogen synthase gene promoter in <i>Neurospora crassa</i> . Proteomics 8:2052–2061	508 509 510
Freitas FZ, de Paula RM, Barbosa LC, Terenzi HF, Bertolini MC (2010) cAMP signaling pathway controls glycogen metabolism in <i>Neurospora crassa</i> by regulating the glycogen synthase gene expression and phosphorylation. Fungal Genet Biol 47:43–52	511 512 513
Fu YH, Marzluf GA (1990) <i>nit-2</i> , the major nitrogen regulatory gene of <i>Neurospora crassa</i> , encodes a protein with a putative zinc finger DNA-binding domain. Mol Cell Biol 10:1056–1065	514 515
Galagan JE, Calvo SE, Borkovich KA, Selker EU et al (2003) The genome sequence of the filamentous fungus <i>Neurospora crassa</i> . Nature 422:859–868	516 517
Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B (2005) Genomics of the fungal kingdom: insights into eukaryotic biology. Genome Res 15:1620–1631	518 519
Gibbons BJ, Roach PJ, Hurley TD (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. J Mol Biol 319:463–477	520 521
Gonçalves RD, Cupertino FB, Freitas FZ, Luchessi AD, Bertolini MC (2011) A genome-wide screen for <i>Neurospora crassa</i> transcription factors regulating glycogen metabolism. Mol Cell Proteomics 10(11). doi:10.1074/mcp.M111.007963	522 AU7 523 524
Görner W, Durchschlag E, Martinez-Pastor MT, Estruch F, Ammerer G, Hamilton B, Ruis H, Schüller C (1998) Nuclear localization of the C ₂ H ₂ zinc finger protein Msn2p is regulated by stress and protein kinase A activity. Genes Dev 12:586–597	525 526 527
Hager GL, McNally JG, Misteli T (2009) Transcription dynamics. Mol Cell 35:741–753	528
Hardy TA, Roach PJ (1993) Control of yeast glycogen synthase-2 by COOH-terminal phosphorylation. J Biol Chem 268:23799–23805	529 530
Huang D, Moffat J, Wilson WA, Moore L, Cheng C, Roach PJ, Andrews B (1998) Cyclin partners determine Pho85 protein kinase substrate specificity in vitro and in vivo: control of glycogen biosynthesis by Pcl8 and Pcl10. Mol Cell Biol 18:3289–3299	531 532 533
Jakobsen BK, Pelham HRB (1988) Constitutive binding of yeast heat shock factor to DNA in vivo. Mol Cell Biol 8:5040–5042	534 535
Johnston M, Carlson M (1992) Regulation of carbon and phosphate utilization. In: Jones EW, Pringle JR, Broach JR (eds) The molecular and cellular biology of the yeast <i>Saccharomyces</i> . Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 193–281	536 537 538
Kingston RE, Schuetz TJ, Larin Z (1987) Heat-inducible human factor that binds to a human hsp70 promoter. Mol Cell Biol 7:1530–1534	539 540

- 541 Krisman CR, Barengo R (1975) A precursor of glycogen biosynthesis: alpha-1,4-glycan-protein.
542 Eur J Biochem 52:117–123
- 543 Lamb TM, Mitchell AP (2003) The transcription factor Rim101p governs ion tolerance and cell
544 differentiation by direct repression of the regulatory genes NRG1 and SMP1 in *Saccharomyces*
545 *cerevisiae*. Mol Cell Biol 23:677–686
- 546 Lambregts R, Shi M, Belden WJ, Decaprio D, Park D, Henn MR, Galagan JE, Bastürkmen M,
547 Birren BW, Sachs MS, Dunlap JC, Loros JJ (2009) A high-density single nucleotide polymor-
548 phism map for *Neurospora crassa*. Genetics 181:767–781
- 549 Li W, Mitchell AP (1997) Proteolytic activation of Rim1p, a positive regulator of yeast sporulation
550 and invasive growth. Genetics 145:63–73
- 551 Lomako J, Lomako WM, Whelan WJ (1988) A self-glucosylating protein is the primer for rabbit
552 muscle glycogen biosynthesis. FASEB J 2:3097–3103
- 553 Martinez-Pastor M, Marchler G, Schuller C, Marchler BA, Ruis H, Estruch F (1996) The
554 *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcrip-
555 tional induction through the stress response element (STRE). EMBO J 15:2227–2235
- 556 Matias PM, Gorynia S, Donner P, Carrondo MA (2006) Crystal structure of the human AAA+
557 protein RuvBL1. J Biol Chem 281:38918–38929
- 558 Meyer U, Monnerjahn C, Techel D, Rensing L (2000) Interaction of the *Neurospora crassa* heat
559 shock factor with the heat shock element during heat shock and different developmental stages.
560 FEMS Microbiol Lett 185:255–261
- 561 Mu J, Cheng C, Roach PJ (1996) Initiation of glycogen synthesis in yeast. Requirement of multiple
562 tyrosine residues for function of the self-glucosylating Glg proteins in vivo. J Biol Chem
563 271:26554–26560
- 564 Neer EJ, Schmidt CJ, Nambudripad R, Smith TF (1994) The ancient regulatory-protein family of
565 WD-repeat proteins. Nature 371:297–300
- 566 Ni HT, LaPorte DC (1995) Response of a yeast glycogen synthase gene to stress. Mol Microbiol
567 16:1197–1205
- 568 Ninomiya Y, Suzuki K, Ishii C, Inoue H (2004) Highly efficient gene replacements in *Neurospora*
569 strains deficient for nonhomologous end-joining. Proc Natl Acad Sci USA 101:12248–12253
- 570 Noventa-Jordão MA, Polizeli MLTM, Bonini BM, Jorge JA, Terenzi HF (1996) Effects of
571 temperature shifts on the activities of *Neurospora crassa* glycogen synthase, glycogen phos-
572 phorylase and trehalose-6-phosphate synthase. FEBS Lett 378:32–36
- 573 Nowak AJ, Alfieri C, Stirnimann CU, Rybin V, Baudin F, Ly-Hartig N, Lindner D, Müller CW
574 (2011) Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and
575 H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding
576 sites. J Biol Chem 286:23388–23396
- 577 Parrou JL, Teste MA, François J (1997) Effects of various types of stress on the metabolism of
578 reserve carbohydrates in *Saccharomyces cerevisiae*: genetic evidence for a stress-induced
579 recycling of glycogen and trehalose. Microbiology 143:1891–1900
- 580 Peñalva MA, Arst HN Jr (2002) Regulation of gene expression by ambient pH in filamentous fungi
581 and yeasts. Microbiol Mol Biol Rev 66:426–446
- 582 Pitcher J, Smythe C, Campbel DG, Cohen P (1987) Identification of the 38-kDa subunit of rabbit
583 skeletal muscle glycogen synthase as glycogenin. Eur J Biochem 169:497–502
- 584 Preiss J, Romeo T (1994) Molecular biology and regulatory aspects of glycogen biosynthesis in
585 bacteria. Prog Nucleic Acid Res Mol Biol 47:299–329
- 586 Roach PJ, Skurat AV, Harris RA (2001) Regulation of glycogen metabolism. In: Jefferson LS,
587 Cherrington AD (eds) Handbook of physiology. The endocrine pancreas and regulation of
588 metabolism, vol II. Oxford University Press, New York, pp 609–647
- 589 Schmitt AP, McEntee K (1996) Msn2p, a zinc finger DNA-binding protein, is the transcriptional
590 activator of the multistress response in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA
591 93:5777–5782
- 592 Skurat AV, Dietrich AD, Zhai L, Roach PJ (2002) GNIP, a novel protein that binds and activates
593 glycogenin, the self-glucosylating initiator of glycogen biosynthesis. J Biol Chem
594 277:19331–19338

Skurat AV, Dietrich AD, Roach PJ (2006) Interaction between glycogenin and glycogen synthase. Arch Biochem Biophys 456:93–97	595 596
Smith A, Ward MP, Garrett S (1998) Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. EMBO J 17:3556–3564	597 598 599
Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common architecture for diverse functions. Trends Biochem Sci 24:181–185	600 AU9 601
Smythe C, Caudwell FB, Ferguson M, Cohen P (1988) Isolation and structural analysis of a peptide containing the novel tyrosyl-glucose linkage in glycogenin. EMBO J 7:2681–2686	602 603
Smythe C, Watt P, Cohen P (1990) Further studies on the role of glycogenin in glycogen biosynthesis. Eur J Biochem 189:199–204	604 605
Sorger PK, Lewis MJ, Pelham HR (1987) Heat shock factor is regulated differently in yeast and HeLa cells. Nature 329:81–84	606 607
Télliez-Iñón MT, Torres HN (1970) Interconvertible forms of glycogen phosphorylase in <i>Neurospora crassa</i> . Proc Natl Acad Sci USA 66:459–463	608 609
Télliez-Iñón MT, Terenzi H, Torres HN (1969) Interconvertible forms of glycogen synthetase in <i>Neurospora crassa</i> . Biochim Biophys Acta 191:765–768	610 611
Tilburn J, Sarkar S, Widdick DA, Espeso EA, Orejas M, Mungroo J, Peñalva MA, Arst HN Jr (1995) The <i>Aspergillus</i> PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH. EMBO J 14:779–790	612 613 614
Torija MJ, Novo M, Lemassu A, Wilson WA, Roach PJ, François J, Parrou JL (2005) Glycogen synthesis in the absence of glycogenin in the yeast <i>Saccharomyces cerevisiae</i> . FEBS Lett 18:3999–4004	615 616 617
Unnikrishnan I, Miller ST, Meinke M, LaPorte DC (2003) Multiple positive and negative elements involved in the regulation of expression of GSY1 in <i>Saccharomyces cerevisiae</i> . J Biol Chem 278:26450–26457	618 619 620
van Peij NN, Gielkens MM, de Vries RP, Visser J, de Graaff LH (1998) The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in <i>Aspergillus niger</i> . Appl Environ Microbiol 64:3615–3619	621 622 623
Viskupic E, Cao Y, Zhang W, Cheng C, DePaoli-Roach AA, Roach PJ (1992) Rabbit skeletal muscle glycogenin. Molecular cloning and production of fully functional protein in <i>Escherichia coli</i> . J Biol Chem 267:25759–25763	624 625 626
Vyas VK, Berkey CD, Miyao T, Carlson M (2005) Repressors Nrg1 and Nrg2 regulate a set of stress-responsive genes in <i>Saccharomyces cerevisiae</i> . Eukaryot Cell 4:1882–1891	627 628
Wang Z, Wilson WA, Fujino MA, Roach PJ (2001) Antagonistic control of autophagy and glycogen accumulation by Snf1p, the yeast homolog of AMP-activated protein kinase and the cyclin-dependent kinase Pho85p. Mol Cell Biol 21:5742–5752	629 630 631
Wang TY, He F, Hu QW, Zhang Z (2011) A predicted protein-protein interaction network of the filamentous fungus <i>Neurospora crassa</i> . Mol Biosyst 7:2278–2285	632 633
Wilson WA, Roach PJ, Montero M, Baroja-Fernández E, Muñoz FJ, Eydallin G, Viale AM, Pozueta-Romero J (2010) Regulation of glycogen metabolism in yeast and bacteria. FEMS Microbiol Rev 34:952–985	634 635 636
Wood MA, McMahon SB, Cole MD (2000) An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. Mol Cell 5:321–330	637 638
Yamashiro CT, Ebbole DJ, Lee BU, Brown RE, Bourland C, Madi L, Yanofsky C (1996) Characterization of <i>rco-1</i> of <i>Neurospora crassa</i> , a pleiotropic gene affecting growth and development that encodes a homolog of Tup1 of <i>Saccharomyces cerevisiae</i> . Mol Cell Biol 16:6218–6228	639 640 641 642
Zhai L, Dietrich A, Skurat AV, Roach PJ (2004) Structure-function analysis of GNIP, the glycogenin-interacting protein. Arch Biochem Biophys 421:236–242	643 644

Anexo III

**Biophysical characterization of the recombinant importin- α from
*Neurospora crassa***

Agnes A. S. Takeda¹, Fernanda Z. Freitas², Angelo J. Magro¹, Natalia E. Bernardes^a, Carlos A. H. Fernandes^a, Rodrigo D. Gonçalves², Maria Célia Bertolini^{2,#}, Marcos R. de M. Fontes^{1,*}

¹*Departamento de Física e Biofísica, Instituto de Biociências, Universidade Estadual Paulista, Botucatu, SP, 18618-970, Brazil* ²*Departamento de Bioquímica e Tecnologia Química, Instituto de Química, Universidade Estadual Paulista, Araraquara, SP, Brazil*

#Address correspondence to this author at the Departamento de Bioquímica e Tecnologia Química, Instituto de Química, Universidade Estadual Paulista, Araraquara, SP, Brazil; Phone: 55-16-33019675; Fax 55-16-33019692; E-mail: mcbertol@iq.unesp.br.

*Address correspondence to this author at the Departamento de Física e Biofísica, Instituto de Biociências, UNESP, C. P. 510, CEP 18618-000, Botucatu-SP, Brazil; Phone: +55.14.38156254; Fax: +55.14.38153744; E-mail: fontes@ibb.unesp.br

Abstract

Neurospora crassa has been widely used as a model organism and contributed to the development of biochemistry and molecular biology by allowing the identification of many metabolic pathways and mechanisms responsible for gene regulation. Nuclear proteins are synthesized in the cytoplasm and need to be translocated to the nucleus to exert their functions which the importin- α receptor has a key role for the classical nuclear import pathway. In an attempt to get structural information of the nuclear transport process in *N. crassa*, we present herein the cloning, expression, purification and structural studies with N-terminally truncated IMP α from *N. crassa* (IMP α -Nc). Circular dichroism analysis revealed that the IMP α -Nc obtained is correctly folded and presents a high structural conservation compared to other importins- α . Dynamic light scattering, analytical size-exclusion chromatography experiments and molecular dynamics simulations indicated that the IMP α -Nc unbound to any ligand may present low stability in solution. The IMP α -Nc theoretical model displayed high similarity of its inner concave surface, which binds the cargo proteins containing the nuclear localization sequences, among IMP α from different species. However, the presence of non-conserved amino acids relatively close to the NLS binding region may influence the binding specificity of IMP α -Nc to cargo proteins.

Keywords: biophysical characterization; classical nuclear import pathway; heterologous expression; homology modeling; importin- α ; *Neurospora crassa*.

1. INTRODUCTION

The filamentous fungus *Neurospora crassa* is a well-studied organism, and has been widely used as a model organism for fundamental aspects of eukaryotic biology. These studies contributed for the development of biochemistry and molecular biology, allowing the identification of several metabolic pathways and mechanisms responsible for gene regulation. Proteins presenting activities in the nucleus such as transcription factors are synthesized in the cytoplasm and must to be translocated to the nuclear membrane through the nuclear pore complex (NPC) associated with protein transporters. One pathway involved in this process is the classical nuclear transport pathway, which depends on importin- α /importin- β heterodimer.

Importin- α (IMP α) recognizes the cargo proteins that contain a nuclear localization sequence (NLS) and, importin- β (IMP β) is responsible for translocation of the importin/cargo protein complex through the NPC. Classical NLSs (cNLSs) contain one or two clusters of positively charged amino acids, and are therefore often divided in monopartite (containing a single cluster of basic amino acids), and bipartite cNLSs (containing two clusters of basic amino acids) [1]. Structural studies have shown that both classes are recognized by the receptor IMP α . This nuclear protein receptor has two NLS binding sites formed by conserved residues in its armadillo repeat-domain (ARM), the major and minor NLS-binding sites. The N- and C-terminal clusters of a bipartite NLS interact with the minor and major NLS-binding sites while a monopartite NLS interacts only with the major binding site [2,3,4]. The structure of full-length IMP α showed that the major NLS-binding site is occupied by residues 44-54 from its N-terminal region (IMP β binding domain) that resembles an NLS [5]; IMP α is therefore autoinhibited in the absence of IMP β . Then, N-terminally truncated IMP α has been used as a model for IMP α /IMP β heterodimer since the truncated may simulate an

IMP α /IMP β heterodimer in several crystal structures [2,3,4]. Crystal structures of IMP α from *Saccharomyces cerevisiae* [6], *Mus musculus* [5] and *Homo sapiens* [7] have been reported in the literature. These structures present similar fold; include the 8 ARM repeat-domain, having sequence identities in the range of 41-61%. The structural differences among IMP α from these organisms, although small, should have fundamental importance for the specificity in the transport of cargo proteins to the cellular nuclei. It is noteworthy that there are different IMP α isoforms in the same organism, which are involved in the transport of different proteins, highlighting the importance of specificity in the nuclear transport.

In order to get structural and functional information of the nuclear transport process in *N. crassa*, we cloned the cDNA encoding the N-terminally truncated importin- α from *N. crassa* (IMP α -Nc), produced the recombinant protein in *Escherichia coli*, performed biophysical experiments (analytical size-exclusion chromatography, circular dichroism and dynamic light scattering) and bioinformatics studies. The results obtained in this work may add important elements to understand the nuclear transport process in the model microorganism *N. crassa*.

2. MATERIAL AND METHODS

*2.1. Cloning of the cDNA encoding the IMP α from *N. crassa**

Conidia (2×10^8 cells/mL) from the *N. crassa* strain FGSC 9718 were cultivated in VM liquid medium [8] supplemented with sucrose 2%, at 30 °C, 250 rpm during 24 h. The mycelium was harvested and total RNA was prepared [9]. To synthesize the cDNA first-strand, 600 ng of total RNA were treated with RQ1 RNase-Free DNase (Promega) and reverse transcribed using the oligo (dT)₂₀ reverse primer (Invitrogen) and the commercial kit

Superscript[®] III *First-strand Synthesis for RT-PCR* (Invitrogen), according to the manufacturer's protocols. The 1,425 bp cDNA fragment encoding the IMP α was amplified by PCR using 2.0 μ L of total cDNA, 1U Phusion[™] DNA polymerase (Finnzymes), 10 mM dNTPs, 1x Phusion[™] GC buffer, 1.5 μ L of DMSO, and 100 pmoles/ μ L of each specific oligonucleotides IMP α Δ 74-F (5'-CATATGACCGAGTCTCAGTTGAGCGAG-3', *Nde*I site underlined) and IMP α -R (5'-GAATTCTTACATGTCCATCGACTCGGTG-3', *Eco*RI site underlined). The oligonucleotides were designed based on the NCU01249 ORF sequence in the genome database at the FGSC site (<http://www.broadinstitute.org/annotation/genome/neurospora/MultiHome.html>). The oligonucleotide IMP α Δ 74-F was designed to amplify a cDNA encoding a truncated protein, in which the 74 N-terminal amino acid residues were missing. The amplified fragment was cloned into the *Eco*RV site of the pMOS-Blue vector (GE Healthcare) leading to the pMOS-Nc1249 plasmid construction. An *Nde*I-*Eco*RI DNA cassette was transferred from the pMOS-Nc1249 construction to pET28a vector (Novagen) to generate the pET28-Nc1249 construction. Both plasmid constructions were confirmed by DNA sequencing.

2.2. Production and purification of recombinant importin- α from *N. crassa* and *M. musculus*

The recombinant IMP α -Nc protein (75-549) was produced in LB medium as a Hexa-His-fusion protein using the *Escherichia coli* host strain Rosetta[™] (DE3) pLysS (Novagen). Briefly, the pET28-Nc1249 transformants were cultured at 37 °C and 250 rpm until an OD_{600nm} of about 0.6. The recombinant protein was induced by addition of 0.1 mM IPTG final concentration, during 6 h at 28 °C and 200 rpm. After that, cells were harvested by centrifugation, suspended in lysis buffer (50 mM Hepes, pH 7.0, 150 mM imidazole, 500 mM NaCl, 5% glycerol, 0.5% Nonidet-P40, 10 mM benzamidine, 1 mM PMSF, 5 mM EDTA) and lysed by sonication (5 cycles of 30 sec sonication followed by 30 sec on ice). The cellular

extract was clarified by centrifugation (20,000 $\times g$, 20 min at 4 °C), the supernatant was filtered in a 0.22 μm pore membrane (Millipore) and applied onto a nickel-affinity column in an Akta Prime Purification System (GE Healthcare). The recombinant protein was eluted in a 0.15-1.5 M imidazole linear gradient, and fractions containing the purified protein were pooled and dialyzed against lyses buffer. The protein was quantified by the Hartree [10] method using BSA as standard.

Hexa-His-tagged truncated *Mus musculus* importin- α isoform 2 (IMP α -Mm), corresponding to amino acids 70-529 was expressed in *E. coli* and purified by nickel affinity chromatography, as described previously [11]. The protein was eluted using a 0.15-1.5 M linear gradient of imidazol followed by dialysis and the IMP α -Mm sample was stored in buffer 20 mM Tris-HCl pH 8.0, 100 mM NaCl and 10 mM DTT at -20°C.

2.3. Circular dichroism spectroscopy

Circular dichroism (CD) experiments were performed on a Jasco J-815 (JASCO Inc., Tokyo, Japan) Spectropolarimeter, equipped with a peltier temperature control PTC-423-S. The IMP α -Nc and IMP α -Mm samples were prepared at 370 $\mu\text{g}\cdot\text{ml}^{-1}$ in buffer 10 mM Tris-HCl, pH 8.0; 20 mM NaCl. All CD measurements were taken in a spectral range of 190–260 nm at 20 °C, using an optical path-length 0.05 mm with scanning speed 100 nm/min; a band width of 2 nm and response time of 1 s. The final spectra is resultant of 20 spectra that were accumulated, averaged and corrected from the baseline for buffer solution contribution and normalized to residual molar ellipticity [θ].

The thermal denaturation analysis was obtained by monitoring the circular ellipticity changes at a fixed wavelength of 222 nm while the sample was heated from 10 to 90°C. Data points were acquired by a ramp rate of 1°C/min and an equilibration time of 5 seconds after each temperature adjustment using an optical path-length 2.0 mm. The denaturation curve

was normalized to apparent fraction folded, according to the equation $[\theta] = [(\varepsilon_F - \varepsilon_U) \alpha] + \varepsilon_U$, where $[\theta]$ is the ellipticity at any temperature, ε_F is the ellipticity when the protein is fully folded, and ε_U is the ellipticity when the protein is totally unfolded, as previously described [12]. The melting temperature (T_M) was calculated by Denatured Protein Analysis program from Spectra Manager™ II software.

2.4. Dynamic light scattering

The dynamic light scattering (DLS) measurements were performed with native N-terminally truncated IMP α -Nc at concentration of 3.5 mg.ml⁻¹ using a DynaPro TITAN instrument (Wyatt Technology). The sample was filtered through a 0.22 μ m pore membrane (Millipore) prior to the measurement. The data were collected at a fixed angle of 90°, temperature of 4°C and measured one hundred times. All results were analyzed with the Dynamics v.6.10 software.

2.5. Analytical size-exclusion chromatography

Analytical size-exclusion chromatography was performed using a Superdex-200 10/30 column, equilibrated with a buffer containing 20 mM Tris HCl pH 8.0 and 100 mM NaCl, attached to an ÄKTA purifier system (GE Healthcare™). Two samples of the purified IMP α -Nc ($V = 0.1$ ml; [5.7 mg.ml⁻¹]) were injected at a flow rate of 0.5ml/min: (i) A sample of native IMP α -Nc; (ii) A sample of IMP α -Nc mixed to FEN1 NLS peptide (³⁵⁰SSAKRKEPEPKGSTK³⁶⁹) at a proportion of 1:8, incubated for 20 minutes before chromatography. The molecular weight standards were obtained from a high molecular weight gel filtration calibration kit (Sigma-Aldrich™) containing the following components: Blue dextran (2.000 kDa), Beta-amylase (200 kDa), Alcohol dehydrogenase (150 kDa), Albumin from bovine serum (66 kDa), Carbonic anhydrase (29 kDa) and Cytochrome C

(12.4 kDa). The retention volumes for each standard and samples were measured and used to calculate the partition coefficients (K_{av}), which were defined as $K_{av} = (V_r - V_o)/(V_c - V_o)$, where V_r = retention volume, V_o = void volume (calculated based on the retention time of the blue dextran standard), and V_c = geometric bead volume for the column. The coefficient K_{av} obtained for each standard was plotted against the log of the molecular weight to generate a standard curve. The collected fractions were evaluated by 12% SDS-PAGE electrophoresis and western blotting assay using a Monoclonal Anti-polyHistidine antibody produced in mouse (1:1000) (Sigma-Aldrich).

2.6. Molecular Modeling and Dynamics

2.6.1. IMP α -Nc modeling

The full-length sequence of IMP α -Nc was submitted to HHpred server [13] (<http://toolkit.tuebingen.mpg.de/hhpred>). According to the alignment data obtained in this server, the crystallographic model of complex formed by exportin (CSE1P), importin- α (Kap60p) from *S. cerevisiae* and RanGTP (PDB ID 1wa5_chain B) [14] was selected as the more adequate template for the initial modeling of the theoretical IMP α -Nc structural models. The program Modeller 8v2 [15] and the template Kap60p, truncated at N- and C-termini (residues Glu81 and Asp510, respectively) were used to generate two sets of ten IMP α -Nc structural models: (i) models with two segments corresponding to the region of the IBB domain (segment 1: ²⁴ELRRRR²⁹; segment 2: ⁴²EENLAKRRGI⁵¹) bound to the IMP α -Nc major and minor binding sites (IMP α -Nc-IBB), and (ii) models devoid of these segments (IMP α -Nc- Δ IBB). Variable target function method (VTFM) with conjugate gradients (CG) [16] and molecular dynamics (MD) with simulated annealing (SA) [16] were used in order to refine the models. The best IMP α -Nc-IBB and IMP α -Nc- Δ IBB were selected according to stereochemical and energetic parameters calculated, respectively, with RAMPAGE

(<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) [17] and ProSA-web servers (<https://prosa.services.came.sbg.ac.at/prosa.php>) [18].

2.6.2. Molecular dynamics simulations

The best IMP α -Nc-IBB and IMP α -Nc- Δ IBB theoretical models calculated with program Modeller 8v2 were submitted to MD simulations using the program GROMACS (Groningen Machine for Chemical Simulation) v.4.5.3 [19]. All simulations were executed in presence of explicit water molecules [20] using an Ubuntu 9.04 Linux operational system and eight threads of a dual processor and quad-core Intel Xeon E5520 CPU (2.27 GHz) with 24 Gb of RAM. GROMOS 96 53a6 force field [21] was chosen to perform the MD simulations and the protonation states of the charged groups were set to pH 7.0. The minimum distance between any atom of the models and the box wall was 1.0 nm. An energy minimization (EM) using a steepest descent algorithm was performed to generate the starting configuration of the systems. After this step, 200 ps of MD simulation with position restraints applied to the protein (PRMD) were executed in order to relax the systems gently. Then, 20 ns of unrestrained MD simulation were calculated to evaluate the stability of the structures. All MD simulations were carried out in a periodic truncated cubic box under constant temperature (298 K) and pressure (1.0 bar), which were hold by coupling to an isotropic pressure and external heat bath [22]. Overall stereochemical and fold quality of the theoretical IMP α structural models obtained after MD simulations were checked with the servers RAMPAGE [17] and ProSA-web [18].

2.6.3. Statistical analysis

The Brown-Forsythe test was used to analyze the variance deviation of the backbone r.m.s.d. of IMP α -Nc-IBB and IMP α -Nc- Δ IBB models after 10 ns of MD simulation calculated by the program GROMACS v.4.5.3 [19].

3. RESULTS AND DISCUSSION

3.1. Production and purification of the recombinant IMP α -Nc

Recombinant IMP α -Nc was produced as a truncated protein missing the 74 N-terminal amino acid residues and fused to a His tag. The truncated protein lacks the amino acid sequence ¹MADRYIPEHRRTQFKAKSAFKPDELRRRREEQQVEIRKAKREEN LAKRRGIGAGDSRPGASLGAAPDSDDENPP⁷⁴, which corresponds to the auto-inhibitory domain. The protein was overexpressed in *E. coli* in a soluble form with a high yield. Single-step purification by immobilized metal affinity chromatography was enough to produce pure protein for all biophysical experiments (Fig. 1).

3.2. Circular dichroism indicates high structural similarity between Importins- α

The circular dichroism spectrum of the recombinant IMP α -Nc exhibits negative peaks at 208 (-15644.7 [θ]) and 222 nm (-15,438.8 [θ]), which are typical of α -helix rich proteins, confirming the correct fold of the protein obtained (Fig. 2A). This spectrum was compared to that acquired from IMP α -Mm (-15374.1 and -15520.9 for 208 and 222 nm, respectively), for which the structure was elucidated by X-ray crystallography [5]. Both proteins displayed very similar CD spectra confirming that the IMP α -Nc and IMP α -Mm have similar folding. The secondary structure content prediction from the spectrum obtained of IMP α -Nc using CONTINN algorithm [23] resulted in 52.7% α -helix, 16.4% of loops and 24.1% of unordered

elements. This result is similar to the values obtained for IMP α -Mm structure, 51.1% α -helix, 16.5% of loops and 25.8% of unordered elements.

To examine the thermal stability of IMP α -Nc, a thermal analysis was performed at 222 nm in the range of 10 – 90°C. Results indicate that the calculated melting point (T_m) for the truncated protein is 39.2°C, which is plausible considering the normal growth temperature for *N. crassa* (30°C) (Fig 2B). Studies of heat-shock stress (45°C) with *N. crassa* have shown a reduction in glycogen synthase transcript levels (*gsn*) and glycogen levels upon raising the temperature [24]. Proteins involved in the regulation of this metabolic pathway may depend of IMP α -Nc to be transported to the nucleus to develop their function; however for temperatures higher than 40°C, IMP α -Nc is unfolded, impairing the nuclear translocation associated with it.

3.3. Molecular modeling and molecular dynamics simulations

The initial modeling of the IMP α -Nc-IBB and IMP α -Nc- Δ IBB was based on the template Importin- α (Kap60p) from *S. cerevisiae* (PDB code 1wa5_chain B). This structure presents a large part of IBB domain modeled in both major and minor NLS binding sites [14]. For the model selection was used the alignment generated by the HHPred server [13] (score = 674.3), the identity (60%) and similarity (98.9%) between the template and IMP α -Nc amino acid sequences [13]. The best IMP α -Nc-IBB and IMP α -Nc- Δ IBB models generated by the program Modeller 8v2 [15] showed a high overall stereochemical quality, with 100% of their amino acid residues distributed in the favorable and allowed regions of their respective Ramachandran plots [14]. Furthermore, the adequate folding of the models was confirmed by the Z-scores: -12.38 and -11.64 for IMP α -Nc-IBB and IMP α -Nc- Δ IBB models, respectively [15]. These models were submitted to molecular dynamics (MD) simulations in order to get potential insights into IMP α features, especially regarding the influence of ligands on the

structural behavior of the protein in solution (Fig. 3A). After 20 ns-MD simulations, it was verified that the good stereochemical quality of the IMP α -Nc-IBB and IMP α -Nc- Δ IBB models is kept, since, respectively, 97.9% and 96.7% of their amino acid residues occupied the favorable and allowed Ramachandran plot regions. Moreover, the very similar Z-scores calculated after the MD simulations (-11.10 for IMP α -Nc-IBB and -11.91 for IMP α -Nc- Δ IBB) indicate the conservation of an appropriate folding in both models.

Despite the stabilization of both models after MD simulations, it can be observed by the Fig. 3A that IMP α -Nc- Δ IBB model present higher variation of the average backbone r.m.s.d. oscillation in comparison with IMP α -Nc-IBB model. The average backbone r.m.s.d. values calculated for IMP α -Nc-IBB and IMP α -Nc- Δ IBB (0.031 and 0.045 nm, respectively) during the last 10 ns of the MD simulations (Fig. 3A) were significantly different, as pointed out by the Brown-Forsythe test (p-value = $1,4 \cdot 10^{-20}$). Therefore, it is possible to confirm the higher degree of flexibility presented by the IMP α -Nc- Δ IBB model during the progress of the simulation.

3.4. The theoretical structure of the IMP α -Nc and the conservation of key residues

The final theoretical model of N-terminally truncated IMP α -Nc (Fig. 3B) is composed by the NLS binding domain and two segments of IBB domain. The structure conserves the elongated shape built from ten armadillo (ARM) repeat each containing three α -helices (H1, H2 and H3) connected by loops as occurs for all other IMP α structures solved to date [1,5,6,7]. In the IMP α -Nc model, the IBB domain is bound to both major and minor NLS-binding sites, which correspond to ARM repeats 2-4 (in red) and 6-8 (in blue), respectively. Additionally, the ARM repeat 5 (in green) is occupied by IBB residues of the region analogous to linker region of bipartite NLS sequences (Fig. 3B).

The Fig. 4 shows a partial alignment of the sequences IMP α -Nc, IMP α -Mm and IMP α from *S. cerevisiae* (IMP α -Sc) corresponding to the NLS binding domain. This alignment shows the identical amino acids highlighted in magenta, whereas in cyan are represented the non-conserved amino acids. These residues are also highlighted in the IMP α -Nc surface model demonstrating that its inner concave surface (defined by the H3 α -helix of ARM repeat motif), is a highly conserved region among the IMP α from different organisms. Similar results were obtained in a recently review published by Marfori and colleagues [25]. These results confirm the conserved tryptophan, asparagines, aspartic and glutamic acids of IMP α -Nc, comprising cores to accommodate the positively charged classic NLS. By contrast, there is a high concentration of non-conserved amino acids (in cyan) relatively closed to the IBB segments (represented as grey sticks in Fig. 4). Some of these regions consist in loops connecting α -helices; however others contain organized secondary structures and may influence the binding of IMP α -Nc to specific cargo proteins.

3.5. *N-terminally truncated IMP α -Nc requires a ligand at the NLS binding region to be stable*

Several crystal structures of IMP α have been solved by X-ray crystallography either in the full-length version, N-terminally truncated and N-terminally truncated bound to several peptides and ligands [2,3,5-7,26-29]. In all of these cases there was a ligand or its own auto-inhibitory sequence (in the case of full-length IMP α) bound at the major or minor binding sites of IMP α . The N-terminally truncated yeast IMP α (IMP α -Sc) is the unique exception, however in this case, the crystal structure presented as homodimer [5], forming an extensive dimeric interface which includes the linker binding region of bipartite NLSs. Co-crystallization experiments of IMP α -Sc and nucleoplasmin only were possible after of the mutation Tyr393Asp, since this is a critical residue to its dimerization process [5]. Attempts to crystallize IMP α -Mm without any ligand by us [30] or by other scientists [31] were

unsuccessful, and the reason for that, may be because IMP α -Mm does not have tyrosine residue in the position equivalent to 393 as occur for IMP α -Sc, which may aid its dimerization. These experimental results indicate that it is impossible to crystallize N-terminally truncated IMP α without a ligand (or even other IMP α monomer) in the NLS binding site region. In order to get insights about this finding we performed analytical size-exclusion chromatography experiments with apo IMP α -Nc and IMP α -Nc complexed to FEN1 bipartite NLS peptide [32]. Size-exclusion chromatography indicated the arrangement of two populations of apo IMP α -Nc (peaks 1 and 2) in solution (Fig. 5). In attempt to estimate the molecular weight of these observed peaks, a calibration curve was calculated depending on the retention volumes of each standard previously applied to the chromatography column. The peak 1 is too broad to infer its molecular weight, once it is comprised by aggregates of IMP α -Nc, whose molecular weights are higher than the recommended separation range of Superdex-200 10/30 column. The peak 2 also corresponds to apo IMP α -Nc as shown by SDS electrophoresis gel and western blot analysis (Fig. 5) and presented an estimated molecular weight of 83.7 kDa. This apparent molecular weight is higher than the expected for N-terminally truncated IMP α -Nc (52 kDa); but this can be explained by the elongated conformation of importins that may lead to reduction of its retention time at column during the chromatography. Dynamic light scattering (DLS) measurements corroborates these results, since that the DLS measures performed at 4°C indicated a mean hydrodynamic radius (R_H) of 4.1 nm with a polydispersity of 20.4 % and an estimated MW of 90 kDa.

By contrast, size-exclusion chromatography of IMP α -Nc complexed to FEN1 NLS peptide resulted in a single peak (peak 3) that corresponds to IMP α -Nc as shown by SDS electrophoresis gel and western blot analysis (Fig. 5). This peak presented an estimated molecular weight of 81.1 kDa, which is slightly smaller than the molecular weight presented by IMP α -Nc at peak 2. This result suggests that the interaction between the IMP α -Nc and the

NLS peptide may lead to a protein conformation change and could explain this small difference of retention times between IMP α -Nc samples during the chromatography.

In addition to these data, the analytical size-exclusion chromatography experiments, combined to SDS electrophoresis were also capable to detect aggregation of IMP α -Nc in absence of NLS peptide even under denaturing conditions. The fractions correspondent to peak 2 were pooled, concentrated and evaluated by SDS electrophoresis experiments. The apo IMP α -Nc sample (peak 2) presented a predominant band of approximately 55kDa and other band around 130 kDa (Fig. 5). In contrast, the sample from peak 3, which corresponds to IMP α -Nc complexed to FEN1 NLS peptide, showed a single band at the same level as the predominant apo IMP α -Nc band (55 kDa) (Fig. 5). The bands observed in the SDS electrophoresis gel were confirmed as IMP α -Nc by western blot analysis showing that even after the gel filtration the IMP α -Nc can form dimers (Fig. 5). Moreover, the presence of the single band from peak 3 indicates that the aggregation of truncated IMP α -Nc was abolished (or significantly reduced) by interaction between IMP α -Nc and FEN1 NLS peptide.

Then, taking into account all the results obtained by us in the MD simulations, size-exclusion chromatography and DLS experiments with the IMP α -Nc and the crystallography studies with several IMP α , we can hypothesize that the IMP α -Nc has lower stability when is unbound than when is bound to a ligand at its NLS binding region. Furthermore, due to the high reactive residues of the minor and major binding sites of IMP α -Nc [1,5,6,7], the N-terminally truncated IMP α -Nc may bind to any flexible segments containing clusters of positively charged amino acids which give stability to the IMP α -Nc. These results are in agreement with the high affinity constants of IMP α to proteins containing NLS basic clusters when IMP α is bound to IMP β [26], since N-terminally truncated IMP α may simulate an IMP α /IMP β heterodimer. These authors also demonstrated using several techniques that

IMP α is monomeric in the auto-inhibited state (when bound to the IBB domain) and stoichiometry of its association with importin- β is 1:1.

4. CONCLUSIONS

Importin- α from *N. crassa* was cloned, expressed, purified and, studied using biophysical and bioinformatics tools. For the first time, biophysical and bioinformatics studies were performed for importin- α for other organisms than mammalian or yeast. Dynamic light scattering, analytical size-exclusion chromatography experiments and molecular dynamics simulations showed that N-terminally truncated IMP α -Nc presents low stability when it is not bound to a ligand, suggesting thus an explanation to its partial aggregation (or dimerization) in solution. Additionally, the results showed high similarity of inner concave surface, which binds the cargo proteins containing the NLS sequences among the IMP α from different organisms. However, the presence of non-conserved amino acids relatively close to the NLS binding region may be related to the binding specificity of IMP α -Nc to cargo proteins. Additional experiments, including X-ray crystallography, site-directed mutagenesis, and cellular localization may add important elements to understand the fundamental role of IMP α in the protein transport.

ABBREVIATIONS

IMP α = importin- α ;

IMP β = importin- β ;

IBB = importin- β binding domain;

IMP α -Mm = importin- α from *Mus musculus*;

IMP α -Nc = importin- α from *Neurospora crassa*;

IMP α -Nc-IBB = importin- α from *N. crassa* bound to IBB segments;

IMP α -Nc- Δ IBB = N-terminally truncated importin- α from *N. crassa*;

IMP α -Sc = importin- α from *Saccharomyces cerevisiae*

ACKNOWLEDGMENTS

This work was financially supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil) providing research grants and fellowships.

REFERENCES

- [1] Dingwall, C.; Laskey, R.A. Nuclear Targeting Sequences - a Consensus, *Trends in Biochemical Sciences*, **1991**, *16*(12), 478-481.
- [2] Conti, E.; Kuriyan, J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha, *Structure with Folding & Design*, **2000**, *8*(3), 329-338.
- [3] Fontes, M.R.; Teh, T.; Kobe, B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha, *Journal of Molecular Biology*, **2000**, *297*(5), 1183-1194.
- [4] Rexach, M.; Blobel, G. Protein Import into Nuclei - Association and Dissociation Reactions Involving Transport Substrate, Transport Factors, and Nucleoporins, *Cell*, **1995**, *83*(5), 683-692.
- [5] Conti, E.; Uy, M.; Leighton, L.; Blobel, G.; Kuriyan, J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha, *Cell*, **1998**, *94*(2), 193-204.
- [6] Kobe, B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha, *Nature Structural Biology*, **1999**, *6*(4), 388-397.
- [7] Dias, S.M.; Wilson, K.F.; Rojas, K.S.; Ambrosio, A.L.; Cerione, R.A. The molecular basis for the regulation of the cap-binding complex by the importins, *Nat. Struct. Mol. Biol.*, **2009**, *16*(9), 930-937.
- [8] Vogel, H. A convenient growth medium for *Neurospora* (medium N), *Microbial Genetics Bulletin*, **1956**, *13*, 2.

- [9] Sokolovsky, R.R. V.K.M.; Russo, V. E. A. Fast and reliable mini-prep RNA extraction from *Neurospora crassa*., *Fungal Genetics Newsletter*, **1995**, 37,3.
- [10] Hartree, E.F. Determination of protein: a modification of the Lowry method that gives a linear photometric response, *Anal. Biochem.*, **1972**, 48(2), 422-427.
- [11] Fontes, M.R.; Teh, T.; Jans, D. Brinkworth, R.I.; Kobe, B. Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha, *J. Biol. Chem.*, **2003**, 278(30), 27981-27987.
- [12] Greenfield, N.J. Analysis of circular dichroism data, *Methods Enzymol*, **2004**, 383, 282-317.
- [13] Soding, J.; Biegert, A.; Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction, *Nucleic Acids Res.*, **2005**, 33, W244-248.
- [14] Matsuura, Y.; Stewart, M. Structural basis for the assembly of a nuclear export complex, *Nature*, **2004**, 432(7019), 872-877.
- [15] Marti-Renom, M.A.; Stuart, A.C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct.*, **2000**, 29, 291-325.
- [16] Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints, *Journal of Molecular Biology*, **1993**, 234(3), 779-815.
- [17] Lovell, S.C.; Davis, I.W.; Arendall, W.B.,3rd; de Bakker, P.I.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Calpha geometry: phi,psi and Cbeta deviation, *Proteins*, **2003**, 50(3), 437-450.
- [18] Wiederstein, M.; Sippl, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res.*, **2007**, 35, W407-410.
- [19] Berendsen, H.J.C.; Vandespoel, D.; Vandrunen, R. Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation, *Computer Physics Communications*, **1995**, 91(1-3), 43-56.
- [20] Berendsen, J.P.M.P. H.J.C.; van Gunsteren, W. F.; Hermans, J. *Interaction models for water in relation to protein hydration*, in: Pullman B. (Ed.), Intermolecular Forces, Publishing Company, Dordrecht, **1981**, pp. 331-342.
- [21] Oostenbrink, C.; Soares, T.A.; van der Vegt, N.F.; van Gunsteren, W.F. Validation of the 53A6 GROMOS force field, *Eur. Biophys. J.*, **2005**, 34(4), 273-284.
- [22] Berendsen, J.P.M.P. H.J.C.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath, *Journal of Chemical Physics*, **1984**, 81(8), 7.
- [23] Sreerama, N.; Woody, R.W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, *Anal. Biochem.*, **2000**, 287(2), 252-260.

- [24] Freitas, F.Z.; Bertolini, M.C. Genomic organization of the *Neurospora crassa* *gsn* gene: possible involvement of the STRE and HSE elements in the modulation of transcription during heat shock, *Mol. Genet. Genomics*, **2004**, 272(5), 550-561.
- [25] Marfori, M.; Mynott, A.; Ellis, J.J.; Mehdi, A.M.; Saunders, N.F.; Curmi, P.M.; Forwood, J.K.; Boden, M.; Kobe, B. Molecular basis for specificity of nuclear import and prediction of nuclear localization, *Biochim. Biophys. Acta.*, **2011**, 1813(9), 1562-1577.
- [26] Catimel, B.; Teh, T.; Fontes, M.R.; Jennings, I.G.; Jans, D.A.; Howlett, G.J.; Nice, E.C.; Kobe, B. Biophysical characterization of interactions involving importin-alpha during nuclear import, *J. Biol. Chem.*, **2001**, 276(36), 34189-34198.
- [27] Cutress, M.L.; Whitaker, H.C.; Mills, I.G.; Stewart, M.; Neal, D.E. Structural basis for the nuclear import of the human androgen receptor, *J. Cell. Sci.*, **2008**, 121(Pt7), 957-968.
- [28] Takeda, A.A.; de Barros, A.C.; Chang, C.W.; Kobe, B.; Fontes, M.R. Structural basis of importin-alpha-mediated nuclear transport for Ku70 and Ku80, *J. Mol. Biol.*, **2011**, 412(2), 226-234.
- [29] Lott, K.; Bhardwaj, A.; Sims, P.J.; Cingolani, G.. A minimal nuclear localization signal (NLS) in human phospholipid scramblase 4 that binds only the minor NLS-binding site of importin alpha1, *J. Biol. Chem.*, **2011**, 286(32), 28160-28169.
- [30] Takeda, A.A. *Structural studies with the importin*. Doctorate Thesis. Universidade Estadual Paulista Júlio De Mesquita Filho- UNESP, Botucatu- Brasil, **2009**.
- [31] Mynott, A.V. *Structural Investigations of CLIC Proteins and Importin- α Recognition of Nuclear Localisation Signals*. Doctorate Thesis. University of New South Wales, Sidney- Australia, **2009**.
- [32] Barros, A. C.; Takeda, A.A.S.; Chang, C.; Kobe, B. and Fontes, M.R.M. Structural basis of nuclear import of Flap endonuclease 1 (FEN1), *Acta Crystallogr. D Biol. Crystallogr.*, **2012**, doi:10.1107/S0907444912010281.

FIGURE LEGENDS

Fig. (1). Production in *E. coli* and purification of the recombinant IMP α from *N. crassa*. The recombinant protein production in *E. coli* host strain RosettaTM was performed as described in Materials and Methods. Protein extract containing the recombinant His6-IMP α - Δ IBB was applied onto a HisTrapTM HP column and eluted in a 0.5 M linear imidazole gradient. Fractions containing the purified protein were pooled, dialyzed, and used in further experiments. Lanes 1, flow through (FT); 2 and 3, purified IMP α - Δ IBB after dialysis. The numbers on the left side indicate the protein molecular weight in kDa.

Fig. (2). (A) Far-UV circular dichroism spectra of recombinants IMP- α from *N. crassa* (in grey) and *Mus musculus* (in black). (B) Thermal denaturation curve of IMP α from *N. crassa* obtained by circular dichroism spectra at a wavelength of 222 nm while heating from 10° to 90°C.

Fig. (3). (A) Average backbone r.m.s.d. during the 20 ns-MD simulations of the theoretical models: N-terminally truncated importin- α from *N. crassa* (IMP α -Nc Δ IBB) and same protein bound to IBB segments (IMP α -Nc-IBB). Despite of the stabilization after 10 ns observed in both simulations, the model IMP α -Nc- Δ IBB presented a higher oscillation, as pointed out by the Brown-Forsythe test. (B) Ribbons diagram of importin- α theoretical model from *N. crassa* bound to IBB segments (IMP α -Nc-IBB). The IBB segments are shown in a ball-and-stick representation.

Fig. (4). (A) Surface representation of Importin- α theoretical model from *N. crassa*. The IBB segments are shown in grey as stick representation. The conserved amino acids among importin- α from *N. crassa*, *S. cerevisiae* and *Mus musculus* are shown in dark grey and non-conserved ones are shown in light grey. (B) Partial alignment of importin- α from *N. crassa* (Nc), *S. cerevisiae* (Sc) and *M. musculus* (Mm). The conserved amino acids are highlighted in dark grey and the non-conserved in light grey. Some fundamental residues to NLS binding are written in white.

Fig. (5). (A) Analytical size-exclusion chromatography of IMP α -Nc in absence and presence of FEN1 NLS peptide in a Superdex-200 10/30 column equilibrated with a buffer containing

20 mM Tris HCl pH 8.0 and 100 mM NaCl. (B) Calibration curve determined using standard proteins shows the estimated molecular weight of the peaks 2 and 3. The peak 1 is broad to infer the molecular weight, once it is comprised by aggregates of IMP α -Nc whose molecular weights are higher than the recommended separation range of Superdex-200 10/30 column. The protein standards were obtained from a high molecular weight gel filtration calibration kit (Sigma-Aldrich) containing: blue dextran (2.000 kDa), beta-amylase (200 kDa), alcohol dehydrogenase (150 kDa), albumin from bovine serum (66 kDa), carbonic anhydrase (29 kDa) and cytochrome C (12.4 kDa). (C) Coomassie blue stained 12% SDS-PAGE gel (MM: Page RulerTM Plus Prestained Protein Ladder from Thermo Scientific) and western blot analyses of the peaks 2 and 3 using IMP α -Mm as positive control. In all lanes were applied 15 μ g of pure protein.

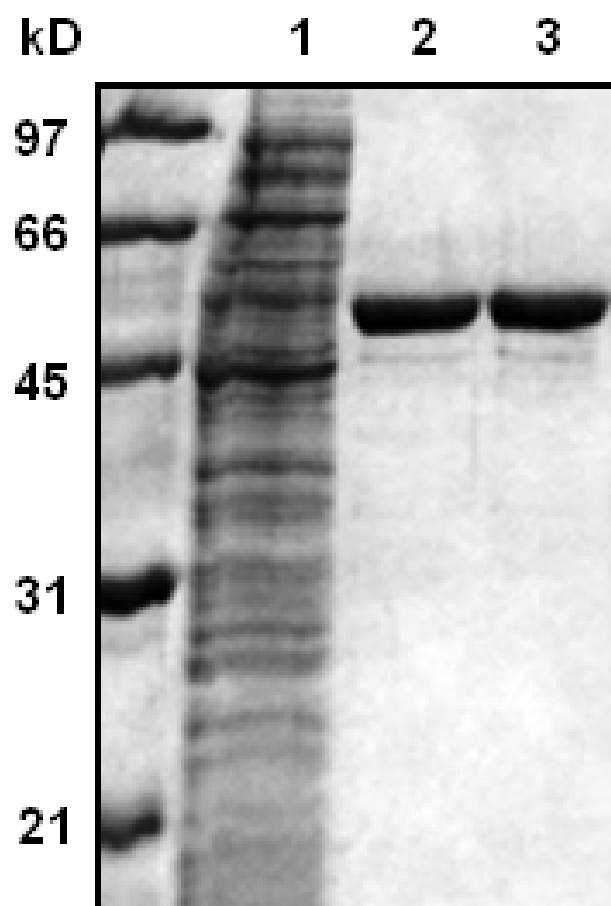


Fig. 1

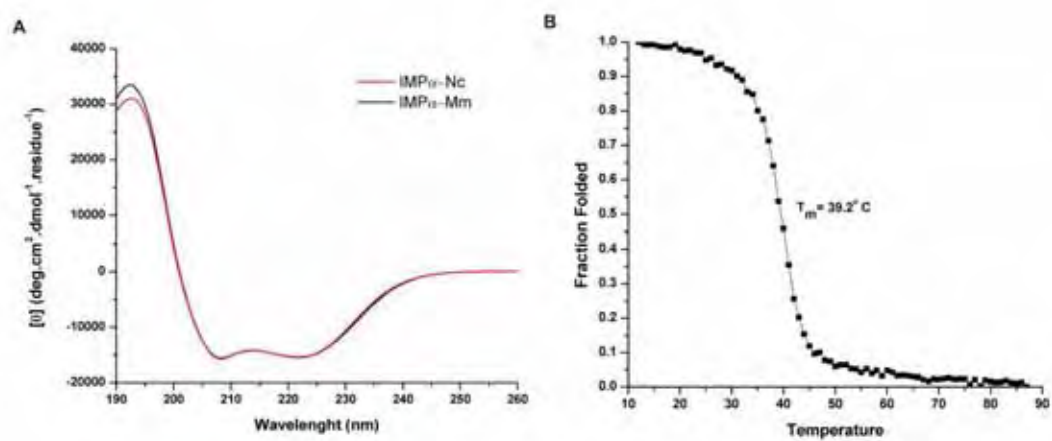


Fig. 2

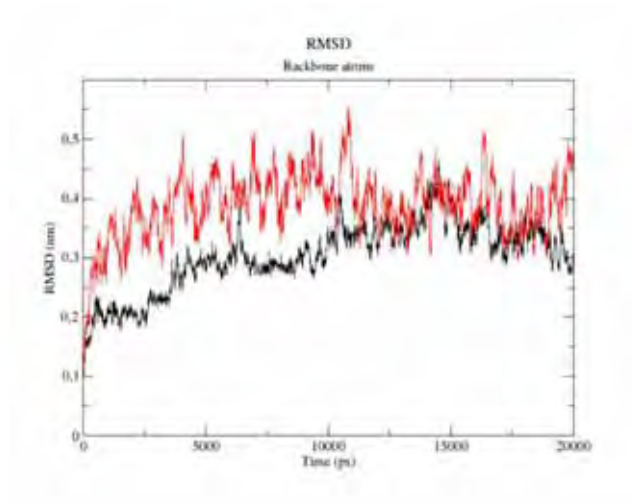


Fig. 3 A

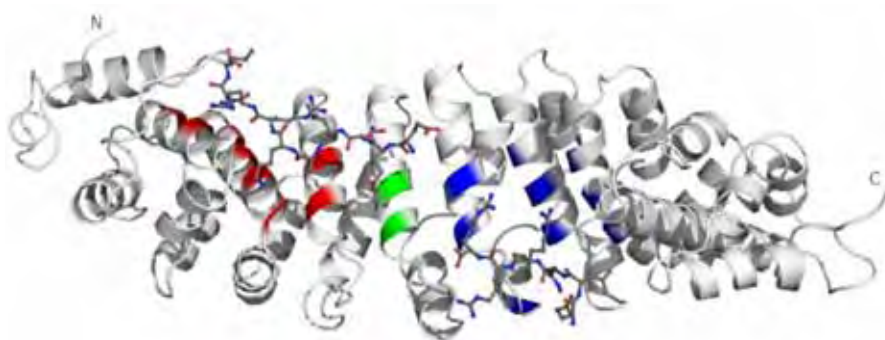


Fig. 3 B

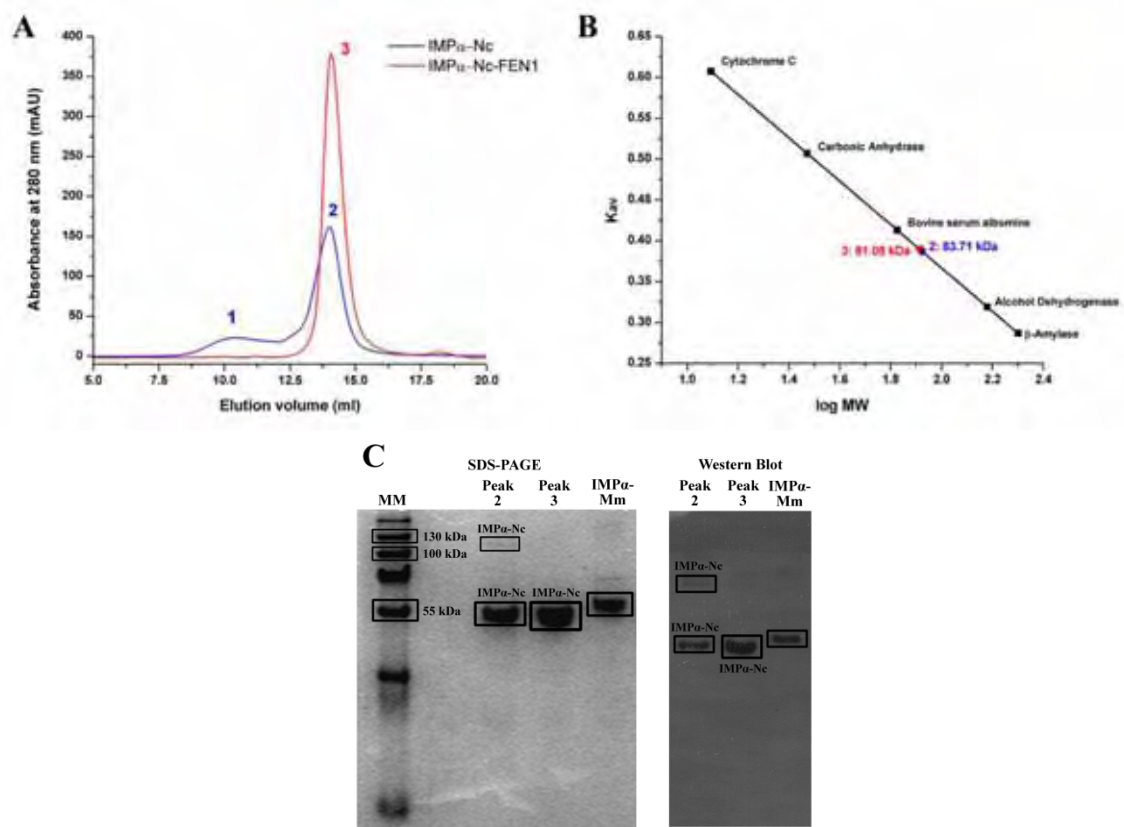


Fig. 5