UNIVERSIDADE DE SÃO PAULO INSTITUTO DE QUÍMICA

Análise do promotor bidirecional que controla os genes citrato sintase e isocitrato liase do fungo filamentoso *Trichoderma reesei*.

ESTELA YNÉS VALENCIA MORANTE

Dissertação de Mestrado

Orientador: Prof. Dr. Hamza El-Dorry

São Paulo 30/06/2006

Ficha Catalográfica

Elaborada pela Divisão de Biblioteca e Documentação do Conjunto das Químicas da USP.

Valencia Morante, Estela Ynés

V152a Análise do promotor bidirecional que controla os genes citrato sintase e isocitrato liase do fungo filamentoso *Trichoderma rees ei* / Estela Ynés Valencia Morante. -- São Paulo, 2006. 76p.

Dissertação (mestrado) — Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica. Orientador : Dorry, Hamza Fahmi Ali El

1. Controle gênico 2. Trichoderma : Micologia I. T. II Dorry, Hamza Fahmi Ali El, orientador.

574.88 CDD

"Análise do promotor bidirecional que controla os genes citrato sintase e isocitrato liase do fungo filamentoso Trichoderma reesei"

ESTELA YNÉS VALENCIA MORANTE

Dissertação de Mestrado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Mestre em Ciências – Área: Bioquímica.

Aprovado(a) por:

Prof. Dr. HAMZA FAHMI ALI EL DORRY
(Orientador e Presidente)

Profa. Dra. SUELY LOPES GOMES
IQ – USP

Mania Cala Battem

Profa. Dra. MARIA CELIA BERTOLINI IQ – UNESP – Araraquara

> SÃO PAULO 11 DE AGOSTO 2006

In memorian Meu filho Andrés

e

Papai Mario

Que desde onde eles estão, me acompanham em todo momento.

A minha mãe Isidora, meus filhos Sergio e Ricardo, meu esposo Felipe, meus irmãos: Arecio, Mario, Julia, Baltazar e Celia que me deram força para seguir sempre adiante. Ao Professor Doutor Hamza El-Dorry, pela orientação sólida e objetiva e por todas as oportunidades a mim dadas durante o período em que estive em seu laboratório.

Aos docentes do Departamento de Bioquímica do Instituto de Química da Universidade de São Paulo, pela transmissão de conhecimentos e ensinamentos durante o curso de pós-graduação.

Aos meus colegas de laboratório Dr. Ari, Dr. Augusto, César, Dr. Eric, Erik, Fabiana, Dr. Felipe, Marluce, Wilton e Zilda pelo companheirismo com que compartimos o dia a dia e pelas amenas e constantes discussões científicas durante todo o trabalho.

Aos funcionários do Departamento de Bioquímica e do Instituto de Química da Universidade de São Paulo, pelo incentivo e apoio a mim proporcionado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro concedido durante a realização deste trabalho no Instituto de Química da Universidade de São Paulo.

ÍNDICE

Resumo	V
Abstract	vi
I. Introdução	1
1.1 Organização do genoma	2
1.2 Genes divergentes e promotor bidirecional	5
1.3 Análise de promotores bidirecionais.	9
1.4 Trichoderma	10
1.5 Citrato sintase e isocitrato liase	11
II. Objetivo	13
III. Materiais e métodos	14
3.1 Microorganismos utilizados	14
3.2 Manipulação de DNA e técnicas gerais de Biologia Molecular	14
3.3 Identificação de clone de <i>lócu</i> s genômico de citrato sintase (<i>TrCit</i>)	15
3.4 Montagem e sequenciamento de subbiblioteca de locus genômico	
de citrato sintase (<i>TrCit</i>)	15
3.5 Análise da região intergenica entre os genes isocitrato liase/citrato sintase	
(TrIcl/TrCit)	16
3.6 Construção do vetor pHygromicina(pHyg)	21
3.7 Transfecção de <i>T. reesei</i>	21
3.8 Análise do crescimento dos transformantes.	23
3.9 Análise da expressão gênica dos genes <i>TrCit e TrIcl</i>	23
3.10 Clonagem e expressão do fator TrZnFSp1	24
3.11 Ensaio de retardamento de mobilidade eletroforética	24
IV. Resultados	27
4.1 Sequênciamento e análise do gene citrato sintase de <i>T. reesei</i>	27
4.2 Sequênciamento e analise do gene isocitrato liase de <i>T. reesei</i>	30
4.3 Análise do padrão de expressão transcricional dos genes TrCit e TrIcl	36
4.4 Análise da região intergênica <i>TrIcl/TrCit</i>	36

V. Discussão	49
5.1 Genoma e rede de regulação transcricional	49
5.2 Organização gênica e metabolismo	57
VI. Conclusões	62
VII. Perspectivas	63
VIII. Referencias bibliográficas	64
IX. Curriculum vitae	70

Índice de figuras

Figura 1. Organização dos genes no genoma eucariótico	4
Figura 2. Organização e estrutura de um gene eucariótico	7
Figura 3. Classes de promotores divergentes	7
Figura 4. Construção do vetor pHyg	22
Figura 5. Montagem do contiguo número 20.	28
Figura 6. Identificação dos genes para Isocitrato liase e Citrato sintase	29
Figura 7. Sequência nucleotídica e de aminoácidos prevista para o lócus	
que contem os genes citrato sintase (TrCit) e isocitrato liase (TrIcl)	31
Figura 8. Alinhamento da sequência de aminoácidos de proteínas citrato sintase	34
Figura 9. Alinhamento da sequência de aminoácidos de proteínas isocitrato liase	35
Figura 10. Análise do padrão de expressão transcricional dos genes citrato sintase	
(TrCit) e isocitrato liase(TrIcl)	37
Figura 11. Identificação de ilhas CpG.	39
Figura 12. Análise da região intergênica entre os genes citrato sintase (TrCit) e isocitra	to
liase(TrIcl) deTrichoderma reesei	41
Figura 13. Determinação do crescimento de transformantes de <i>T. reesei</i>	42
Figura 14. Análises da região intergênica (TrIcl/TrCit) .após deleção da caixa GC	45
Figura 15. Predição de possíveis sítios de ligação de fatores de transcrição	46
Figura 16. Análise da putativa proteína <i>TrZnFSp1</i>	47
Figura 17. Análise de ensaio de mudança de mobilidade eletroforética	48
Figura 18. Relações entre os ciclos do glioxalato e do ácido cítrico	60
Índice de tabelas	
Tabela 1. Oligonucleotídeos empregados na elaboração deste trabalho	18
Tabela 2. Lista de plasmídeos empregados na elaboração deste trabalho	19

Resumo

O gene *TrCit* do fungo filamentoso *Trichoderma reesei* codifica a proteína citrato sintase, uma enzima chave do ciclo de Krebs. Análise da região 5' upstream de *TrCit* mostra que o gene está adjacente ao gene *TrIcl* (que codifica a proteína isocitrato liase, uma enzima do ciclo de glioxalato), em uma orientação cabeça-cabeça. A região promotora intergênica de 647 pb rica em G + C, apresenta uma ilha CpG, seqüência INR, caixas GC, caixas CAAT, sítios de ligação para diversos fatores de transcrição e é isenta de caixa TATA.

O gene *TrCit* de 1573 pb contém 3 éxons e 2 íntrons. Sua seqüência codificadora de 1422 pb produz uma proteína de 474 aminoácidos, com um peso molecular estimado de 52,3 kD. O gene *TrIcl* de 1880 pb contém 3 éxons e 2 íntrons. Sua seqüência codificadora de 1788 pb produz uma proteína de 596 aminoácidos, com um peso molecular estimado de 65,4 kD.

A atividade transcricional da região promotora foi analisada utilizando como repórter o gene de higromicina B fosfotransferase (*hph*). Uma região funcional necessária à transcrição de ambos os genes foi identificada na região central do promotor e contém uma caixa GC que liga o putativo fator de transcrição Sp1 de *T. reesei* (TrZnFSp1). O gene do putativo fator de transcrição "zinc-finger" TrZnFSp1 de 1500 pb contém 3 éxons e 2 íntrons. Sua seqüência codificadora de 1344 pb produz uma proteína de 448 aminoácidos, com um peso molecular estimado de 48,4 kD.

Os resultados mostram que ambos os genes são transcritos de forma divergente a partir de um promotor bidirecional que compartilha na região central uma caixa GC, necessária para a transcrição de ambos os genes.

Abstract.

The *TrCit* gene from the filamentous fungus *Trichoderma reesei* codes for the citrate synthase protein, a key enzyme in the Krebs cycle. Analysis of *TrCit* 5' upstream region showed that it is adjacent to the *TrIcl* gene that codes for isocitrate lyase protein, an enzyme involved in the glyoxylate cycle. Both genes, on a head-to-head orientation, are separated by an intergenic GC-rich and TATA-less promoter region of 647 base pairs. This bidirectional promoter has diverse *cis* regulatory elements: a CpG island, two INR sequences, GC boxes, CAAT boxes and several putative interaction sites for different transcription factors.

The *TrCit* gene, 1,573-base pair-long, has an open reading frame of 1,422 base pairs interrupted by two introns. The gene codes for a protein with an estimated molecular weight of 52.3 kD. The *TrIcl* gene, 1,880-base pair-long, contains 3 exons and 2 introns and a putative coding sequence of 1,788 base pairs. The estimated molecular weight of TrICL is 65.4 kD.

The transcriptional activity of the intergenic promoter region was analyzed using hygromicin B phosphotransferase (*hph*) as a reporter gene. A functional region required for the transcription of both genes was identified in the centre of this promoter. It has a GC box that interacts with a putative transcription factor Sp1 from *T. reesei* (*TrZnFSp1*). The results presented in this work show that both genes are divergently transcribed from a bidirectional promoter that shares an essential central GC box.

I. Introdução.

O dogma central da biologia molecular descreve, como um gene, o segmento de DNA que codifica uma seqüência primária de algum produto gênico final, como um polipeptídeo ou um RNA com uma função catalítica ou estrutural, e daquela que controla a expressão de tal produto. Em 1953 Watson e Crick descreveram o arranjo linear dos genes e como acontece o processo de replicação do DNA. Sendo o genoma de uma espécie seu conjunto de genes característico, ele dirige cada fase de seu desenvolvimento e determina, por exemplo, quando, como e onde as células crescerão e se dividirão.

Porém, a compreensão detalhada da complexidade genética de um organismo será obtida apenas quando cada um de seus genes tiverem sido identificados, e o(s) instante(s) e a(s) condição(ões) para a expressão de cada transcrito forem determinados. Assim, o mínimo necessário é o conhecimento das seqüências completas dos gene e das regiões que o controlam, enquanto o conhecimento da seqüência completa e da organização do genoma do organismo é o melhor cenário possível..

O seqüenciamento do genoma de numerosos seres vivos (http://www.ncbi. nlm.nih.gov/Genomes/), como o do ser humano (Lander *et al.*, 2001; Venter *et al.*, 2001) e *Saccharomyces cerevisiae* (Goffeau *et al.*, 1996), tem permitido a elaboração de mapas genéticos e físicos desses genomas. Esses mapas permitem determinar a estrutura e organização gênica, identificando e localizando genes e regiões intergênicas; estudar a distribuição de genes entre grupos taxonômicos; manipular a herança de determinadas características do indivíduo e determinar o padrão de expressão temporal e espacial de proteínas codificadas pelos genes (Lawrence *et al.*, 2003).

Análises de sequenciamento de DNA e mapeamento de sítios de início de transcrição têm demonstrado que os genomas de vários organismos apresentam genes que são transcritos de forma divergente (em orientação oposta), a partir de promotores compartilhados ou não. Na organização divergente, a transcrição dos genes pode ser coordenada, ou co-regulada, se as regiões reguladoras da transcrição forem compartilhadas (Beck *et al.*, 1988).

A partir do banco de dados "*Trichoderma reesei* EST database and mitochondrial genome" (http://trichoderma.iq.usp.br), de *Trichoderma reesei*, um fungo filamentoso e

organismo modelo para análise genômica (Herrera-Estrella *et al.*, 1990), foi inicialmente selecionado o gene da enzima citrato sintase (*TrCit*) para um estudo do controle de sua expressão gênica que envolveu a clonagem de seu *locus* gênico. Entretanto, a análise desse *locus* mostrou a orientação divergente dos genes citrato sintase (*TrCit*) e isocitrato liase (*TrIcl*), separados por uma curta região intergênica. Assim, pretendeu-se analisar está região promotora intergênica, com o intuito de determinar se ambos os genes estão sob o controle de um promotor bidirecional, identificar os prováveis elementos responsáveis pelo controle da transcrição e descrever pela primeira vez um promotor desta natureza em *T. reesei*.

1.1 Organização do genoma.

O genoma de procariotos é, na maioria dos casos, altamente organizado. Os genes estão ordenados e orientados na direção da replicação, para evitar a colisão entre as enzimas DNA e RNA polimerases, e múltiplos genes estão organizados em operons e são co-transcritos num mRNA policistrônico (Koyanagi, *et al.*, 2005).

Em contraste direto, o arranjo dos genes no genoma eucariótico não é organizado, a maioria dos genes é regulada individualmente, com um promotor na extremidade 5' e um terminador de transcrição na extremidade 3' e os transcritos são mRNAs monocistrônicos; porém, nem todos os genes são transcritos de forma monocistrônica. Vários exemplos de transcrição policistrônica em eucariotos, desde cordados a mamíferos (como o gene de imunoglobulina, de humano), têm sido reportados (Blumenthal, 1998; Blumenthal, 2004), e nestas unidades de transcrição policistrônica, um simples promotor é responsável pela transcrição de mais de um gene (Blumenthal, 1998).

Um outro aspecto importante da organização do genoma dos seres vivos é a densidade de genes. Em *S. cerevisiae*, pelo menos 6.274 genes estão arranjados em um genoma de aproximadas 13 Mb, com um tamanho médio de 2 kb por gene, sendo considerado, então, um genoma densamente organizado. No genoma humano os 30 a 40 mil genes estão arranjados num genoma de aproximadas 3 Gb, com um espaçamento calculado de aproximadas 85 kb. Assim, os genes no genoma humano estão mais dispersos em relação ao genoma de organismos mais simples (Adachi *et al.*, 2002).

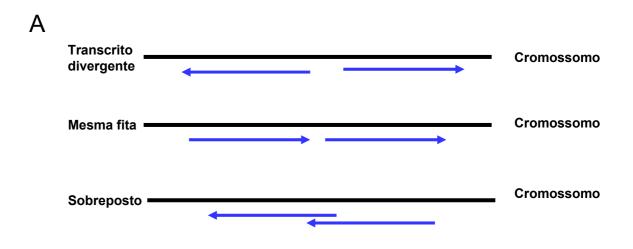
O seqüenciamento do genoma de numerosos organismos tem permitido determinar a estrutura, organização, localização e distancia intergênica, entre os genes. Assim temos:

- A) Em relação, à disposição dos genes no genoma eucariótico, Takai *et al.* (2004) analisam a disposição, independentemente da distância intergênica, dos aproximadamente 2.122 genes conhecidos e preditos nos cromossomos humanos 20, 21 e 22, e os classificam em três grupos (figura 1a):
 - 1- Genes com transcritos divergentes (~50%);
 - 2- Genes compartilhando a mesma fita (\sim 50%);
 - 3- Genes sobrepostos a outros genes;

A proporção entre genes transcritos divergentes (classe 1) e os localizados na mesma fita (classe 2), está de acordo com a suposição de que os genes nos eucariotos estão localizados aleatoriamente no genoma, sendo que para qualquer gene dado, a orientação de um gene localizado 5' na mesma direção ou na direção oposta é de ~50%.

- B) Em relação à distancia intergênica e orientação dos genes adjacentes, Koyanagi *et al.* (2005) analisaram a de 20.192 genes humanos e identificaram 15.527 pares de genes adjacentes, separados por menos de 1 kb, que segundo sua orientação foram classificados em três grupos (figura 1b):
 - 1- Cabeça cabeça ("Head-to-head") ou orientação divergente, os genes estão localizados em fitas opostas com as extremidades 5' adjacentes;
 - 2- Cauda cauda ("tail-to-tail"), os genes estão em fitas opostas com as extremidades 3' adjacentes;
 - 3- Cabeça cauda ("tail-to-head"), os dois genes estão na mesma fita, com as extremidade 5′ e 3′ adjacentes.

Tendo sido estabelecido que pares de genes arranjados numa orientação divergente, com menos de 1.000 pares de bases separando seus inícios de transcrição, são uma forma comum de organização (~10% dos genes) do genoma humano (Adachi *et al.*, 2002; Trinklein *et al.*, 2004). A presença desta forma de organização dos genes pode



В

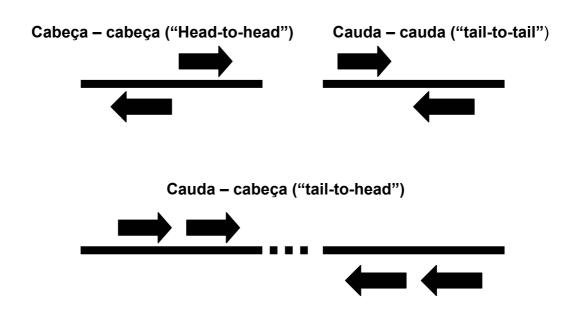


Figura 1. Organização dos genes no genoma eucariótico. A) Em relação a sua disposição (Takai *et al.*, 2004); e B) Em relação a orientação dos genes adjacentes (Koyanagi *et al.*, 2005).

sugerir a existência de um *sistema* regulador localizado entre os dois genes que poderia controlar a expressão de ambos os genes. Se ambos os genes forem controlados por um simples *sistema* regulador eles poderiam apresentar um padrão de expressão similar (Kruglyak, 2000; Koyanagi *et al.*, 2005).

1.2 Genes divergentes e promotor bidirecional.

Em geral a estrutura de um gene eucariótico (figura 2) pode ser dividida em duas partes: a região codificadora, que é transcrita pela RNA pol II e, portanto, localizada "downstream" ao sítio de início de transcrição; e a região que a antecede ("upstream" ao sítio de início de transcrição), denominada região não-codificadora, que contém os elementos reguladores, entre os quais podemos mencionar:

- 1) Promotor:contém sequências especificas de DNA, denominados elementos *cis*, que são reconhecidos por fatores de transcrição específicos e pela RNA pol II na formação do complexo de pré-iniciação. Os elementos *cis* mais comuns num promotor são a caixa TATA (por apresentar a sequência consenso TATAAA) e o elemento Iniciador (INR). Outros elementos próximos ao promotor ("promoter-proximal elements") são a caixas CAAT (consenso CCAAT) e GC (consenso GGGCGG);
- 2) Seqüências intensificadoras ("Enhancers"): elementos reguladores que são reconhecidos por proteínas ativadoras da transcrição (transativadores). Na levedura *S. cerevisiae* são denominadas seqüências ativadoras a montante ("upstream activator sequences", UASs).

Um promotor bidirecional pode regular a transcrição coordenada de um par de genes para, por exemplo, manter a relação estequiométrica de proteínas (histonas) (Beck *et al.*, 1988), ou regular a expressão de genes que participam duma mesma via metabólica (biossíntese de penicilina em *Aspergillus nidulans*) (Martín, 2000), de etapas do ciclo celular (gene humano TDAG51) (Meier-Noorden *et al.*, 2004) ou de resposta a situações de stress (genes humanos HSP60 e HSP10) (Trinklein *et al.*, 2004).

A disposição cabeça-cabeça dos genes (figura 1b) é uma forma de organização gênica presente no genoma de procariotes, vírus, eucariotes e no genoma de mitocôndria e cloroplasto (Beck *et al.*, 1988). Nesta forma de organização, a transcrição dos genes poderia ser coordenada ou co-regulada desde que as regiões reguladoras da transcrição,

presentes no promotor, sejam compartilhadas. Neste caso o sítio de ligação dos fatores de transcrição estaria localizado dentro de uma região responsável pela regulação da transcrição em ambos sentidos (Beck *et al.*, 1988).

Segundo sua disposição, os promotores que dirigem a transcrição de genes divergentes, podem ser de três classes (Beck *et al.*, 1988), (figura 3):

- 1- De costas ("back-to-back"): com uma região de DNA separando os promotores, na qual se ligam proteínas reguladoras,
- 2- Sobrepostos ("overlap"): a região de DNA entre os genes é compartilhada por ambos promotores ("promotor bidirecional"),
- 3- Lado a lado ("face to face"): na qual a região 5′ de cada gene pode, ou não, ter sobreposição.

A análise de 23.752 genes do genoma humano permitiu identificar 1.352 pares de genes divergentes com sítios de início de transcrição separados por menos de 1 kb. A análise das seqüências promotoras mostra que elas apresentam um conteúdo de dinucleotídeos G + C de aproximadamente 66%, apenas 8% delas apresentam um estrito elemento TATA e 77% estão localizados numa ilha CpG. Essas características indicam um grande contraste com aquelas de promotores de genes não divergentes, que apresentam conteúdo de dinucleotídeos GC de aproximadamente 53%; 28% apresentam um estrito elemento TATA e 38% estão localizados numa ilha CpG (Trinklein *et al.*, 2004). As ilhas CpG são regiões curtas de 1 a 2 kb de DNA que caracterizam-se por serem ricas em nucleotídeos G + C. São encontradas nas extremidades 5′ dos genes constitutivos ("housekeeping") e da maioria dos genes com padrão de expressão restrito a um tecido (Larsen *et al.*, 1992; Craig *et al.*, 1994). Entretanto, as características de região rica em G + C e de não apresentar caixa TATA ("TATA-less promoter") por si só não caracterizam um promotor compartilhado por genes transcritos divergentemente (Trinklein *et al.*, 2004).

Promotores bidirecionais que não contêm caixa TATA podem apresentar vários sítios de ligação para fatores de transcrição, como por exemplo: no promotor dos genes humanos OSGEP e APEX ocorrem sítios de ligação para o fator de transcrição UBS (USF binding site), para o fator Sp1 (specific protein 1) e caixa CCAAT, que regula a transcrição de ambos os genes (Seki *et al.*, 2002); ou os sítios de ligação para os fatores

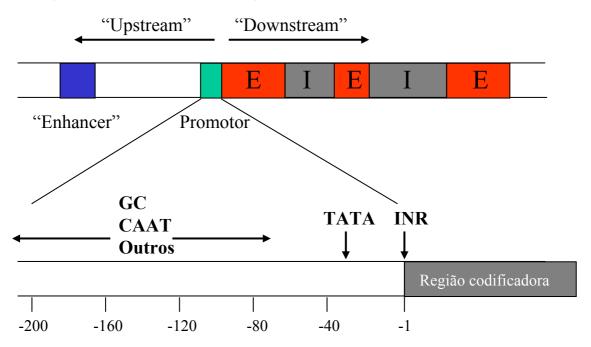


Figura 2. Organização e estrutura de um gene eucariótico. A região codificadora consiste de éxons (E) e íntrons (I). Os elementos *cis* regulatórios localizados na região não-codificadora, são o promotor, "enhancer" e outros (não mostrados). "**Downstream**" se refere a direção de transcrição e "**upstream**", a direção oposta à de transcrição. O número das bases na região do promotor se incrementa na direção da transcrição, com "+1" para a base no sítio de iniciação (INR). Não existe base com posição "0". A base na posição "upstream" à base +1 é numerada como "-1".

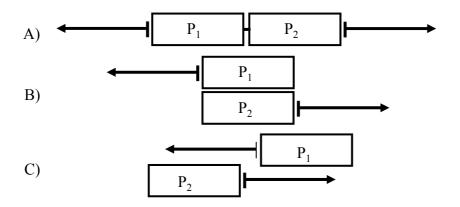


Figura 3. Classes de promotores de genes divergentes: A) De costas, com uma região de DNA entre os promotores, B) Sobrepostos, a região de DNA entre os genes é compartilhada por ambos os promotores e C) Lado a lado, em que a região 5' de cada gene pode, ou não, ter sobreposição. P_1 e P_2 são os promotores e | representa um transcrito na orientação indicada.

de transcrição GABP (GA-repeat binding protein), Sp1/3 (specific protein 1/3) e YY1 (Yin and Yang 1), presentes no promotor dos genes GABPα (GA-repeat binding protein subunit alpha) e CF6 (ATP synthase coupling factor 6), cuja interação forma um complexo que regula a atividade do promotor (Patton *et al.*, 2005).

Alguns promotores sem caixa TATA apresentam sítio de ligação para o fator de transcrição Sp1 (seqüência consenso: GGGCGG ou caixa GC) e sua ativação envolve o recrutamento, por Sp1, do fator de transcrição TFIID (Kolell *et al.*, 2002). Sp1 é um membro da família multigene Sp, que inclui Sp2, Sp3, Sp4, Sp5, Sp6, Sp7 e Sp8. Todas as proteínas Sp são fatores de transcrição que se ligam a seqüências de DNA ricas em GC e apresentam um domínio de ligação ao DNA altamente conservado do tipo "zinc-finger" Cys₂His₂ (Lee *et al.*, 2005). Sp1 e Sp3 são fatores de transcrição ubíquos, envolvidos no controle de uma grande variedade de genes que participam em processos biológicos como apoptose, inibição do crescimento celular, diferenciação e carcinogênese (Kolell *et al.*, 2002; Lee *et al.*, 2005).

Em eucariotos, vários genes transcritos de forma divergente têm sido descritos. Em *S. cerevisiae*, os sítios de início de transcrição dos genes divergentes *GAL1* (galactoquinase) e *GAL10* (uridildifosfoglicose-4-epimerase), estão separados por aproximadamente 600 pares de bases (Johnston, *et al.*, 1984). Adachi, *et al.* (2002), estudaram 290 genes humanos envolvidos em processos celulares como replicação de DNA, regulação de ciclo celular, metabolismo de carboidratos, etc., determinando que 87 genes estudados (30%) tem um transcrito divergente, sendo separados por aproximadamente 1 kb, e que 41% dos genes do ciclo de Krebs (aconitase 1, isocitrato desidrogenase 1, isocitrato desidrogenase 3 gama, succinato desidrogenase A, succinato desidrogenase D, malato desidrogenase 1 e malato desidrogenase 2) estão arranjados de forma divergente com um outro gene.

Entre os genes humanos regulados por um promotor bidirecional podemos mencionar os genes APEX nuclease ("AP endonuclease, 3'-to- 5' exonuclease") e OSGEP ("O-sialoglycoprotein endopeptidase"), cuja região intergênica apresenta uma caixa CCAAT que regula a transcrição de ambos os genes (Seki et al., 2002), como mencionado anteriormente; e os genes mitocondriais de chaperoninas HSP60 ("Heat shock protein 60") e HSP10 ("Heat shock protein 10") (Hansen *et al.*, 2003).

Em fungos, os promotores bidirecionais são empregados no controle de vias metabólicas. Em *Aspergillus nidulans*, os genes que codificam para nitrato e nitrito redutase (*niaD e niiA*, respectivamente) são transcritos de forma divergente, sendo regulados por elementos (4 sítios de ligação para o fator de transcrição NirA e 10 sítios de ligação para o fator de transcrição de tipo GATA "zinc-finger" AreA) inseridos numa região intergênica de 1.267 pb (Muro-Pastor *et al.*, 1999). Os genes para a biossíntese de penicilina (*pcbAB*, *pcbC* e *penDE*) encontram-se agrupados na mesma forma em *Penicillium chrysogenum*, *Aspergillus nidulans* e *Penicillium nalgiovense*, sendo que os genes *pcbAB* e *pcbC* são transcritos de forma divergente a partir dum promotor bidirecional de 1.160 pb. (Martín, J., 2000).

1.3 Análise de promotores bidirecionais.

A identificação de promotores bidirecionais requer uma detalhada análise genética e molecular. Para a análise de promotores bidirecionais, é necessário o emprego de um gene repórter que codifica uma proteína, por exemplo, com atividade enzimática de fácil detecção. A análise consiste em ligar a sequência de DNA a testar, em posição "upstream" da região codificadora do gene repórter, gerando um gene quimérico no qual o putativo elemento regulador controla a expressão do gene repórter. Apesar da atividade do produto do gene repórter ser uma medida indireta da propriedade transcricional da sequência de DNA testada, assume-se que a atividade do gene repórter é diretamente proporcional a atividade transcricional (Alam *et al.*, 1990).

Um vetor ideal para a análise de promotores bidirecionais deve apresentar um par de genes em orientação divergente que codificam enzimas indicadoras apropriadas. Em procariontes uma serie de vetores foram construídos, com as enzimas repórter galK (galactoquinase), lacZ (β-galactosidase), phoA (fosfatase alcalina), etc., que permitem detectar e medir simultaneamente a atividade do promotor bidirecional (Schneider et al., 1986; Schneider et al., 1987; Beck et al., 1988). Em eucariontes, vetores contendo os genes repórter cloramfenicol acetiltransferase (CAT), luciferase, hormônio de crescimento, fosfatase alcalina, β-galactosidase, β-glucuronidase, GFP ("green fluorescent protein"), higromicina B fosfotransferase (hph) e outros têm sido construídos

(Alam *et al.*, 1990), e vários *sistema*s repórter estão disponíveis comercialmente (Promega, USA; Clontech, USA; Invitrogen, USA, etc.) (Sussman, 2001).

O gene higromicina B fosfotransferase (hph) é utilizado como marcador de resistência ao antibiótico higromicina B (Hyg), sendo empregado comumente para a transformação de fungos filamentosos. Plasmídeos levando fusões de uma variedade de promotores à região codificadora do gene hph têm sido construídos, para conferir resistência a Hyg a células eucarióticas, como aos fungos Aspergillus nidulans, A. niger (Punt et al., 1987) e Neurospora crassa (Staben et al., 1989). Em N. crassa a análise da atividade hph, controlada pela região 5´ reguladora do gene arg-2 (fusão arg-2-hph), permitiu estabelecer que o gene hph pode ser usado como um combinado marcador de resistência / gene repórter (Freitag, et al. 1995). Plasmídeos, levando o gene hph, têm sido empregados por nosso laboratório para a transformação eficiente de T. reesei (Henrique-Silva et al., 1996).

1.4 Trichoderma.

O gênero *Trichoderma* compreende espécies de fungos filamentosos que normalmente habitam o solo (Eveleigh, 1985). Juntamente com *Neurospora*, *Penicillium* e *Aspergillus*, o gênero *Trichoderma* é um dos mais estudados entre os fungos filamentosos devido ao seu grande potencial de aplicação do ponto de vista industrial e biotecnológico (Samuels, 1996). A espécie *T. reesei*, é caracteristicamente reconhecida pela produção de sistemas extracelulares de enzimas envolvidas na hidrólise de polissacarídeos (Béguin, 1990).

T. reesei apresenta hifas ramificadas (de 5 a 10 μm de diâmetro), polinucleadas, de núcleo haplóide, que formam colônias brancas de rápido crescimento, formando almofadas verdes ou amarelas de filamentos de esporulação. Os filamentos férteis, conidióforos, produzem fileiras laterais de pequenas fiálides. Os mitoesporos (conídios, geralmente de cor verde e de 3 a 5 μm de diâmetro) são produzidos sucessivamente no extremo da fiálide e agrupados em pequenas massas (Gams et al., 1998). A cariotipagem em uma linhagem selvagem QM6a e em cepas mutantes hipercelulolíticas, como QM9414, através de eletroforese de campo pulsado revela sete bandas, correspondentes a sete cromossomos (Herrera-Estrella et al., 1993).

Nosso laboratório sequenciou o genoma mitocondrial e sequências EST, identificando mais de 10% dos genes expressos em T. reesei, num total de 2 mil sequências EST únicas, estes dados foram depositados no GeneBank (NCBI, USA) (Chambergo et al., 2002) e estão disponíveis publicamente no "website" "Trichoderma reesei EST database and Mitochondrial genome" (http://trichoderma.iq.usp.br). Em 2002, o "Joint Genome Institute" (JGI - USA), financiado pelo Departamento de Energia norteamericano (DOE - USA) iniciou o sequenciamento do genoma de 33 Mb de T. reesei QM9414, uma cepa com capacidade de secretar grandes quantidades de enzimas celulases e hemicelulases, que podem ser empregadas para a conversão de biomassa, produzida por plantas, em bioprodutos de utilidade industrial, como açucares e bioetanol. "Draft genome Trichoderma reesei v. 2005 foi liberado 0 (http://gsphere.lanl.gov/trire1/trire1. home.html), que identifica aproximadamente 8.000 genes putativamente expressos neste organismo.

Durante a análise do "*Trichoderma reesei* EST database and Mitochondrial genome" foi identificado o gene citrato sintase (*TrCit*). A análise da região 5′ flanqueadora upstream de *TrCit*, mostra que o gene está adjacente ao gene *TrIcl* (que codifica a proteína isocitrato liase), sendo ambos transcritos de forma divergente a partir de uma região intergênica de 647 pb. Pretendeu-se, então, identificar a região e os elementos responsáveis pela transcrição de ambos os genes.

1.5 Citrato sintase e isocitrato liase.

No metabolismo energético, a enzima citrato sintase (EC 2.3.3.1), tem um papel chave, pois, catalisa a entrada de carbono no ciclo dos ácidos tricarboxílicos (TCA, ciclo de Krebs, ou do ácido cítrico), através da seguinte reação:

Em fungos e plantas, além da atividade citrato sintase no ciclo dos ácidos tricarboxílicos, uma atividade citrato sintase é encontrada no glioxissomo (peroxissomo), na qual a enzima catalisa a primeira etapa do ciclo do glioxilato.

O genoma de *S. cerevisiae* apresenta três genes que codificam isoformas de citrato sintase: Cit2p, localizada no peroxissomo e que participa do ciclo de glioxilato; Cit1p e Cit3p, que estão localizadas na mitocôndria, sendo Cit1p a isoforma que participa no

ciclo dos ácidos tricarboxílicos ("Saccharomyces Genome Database", http://www.yeastgenome.org/).

Nos vegetais, *E. coli*, levedura e alguns fungos, o acetato pode servir tanto como um substrato rico para obtenção de energia ou como uma fonte de fosfoenolpiruvato para a síntese de carboidratos. Nesses organismos o ciclo do glioxilato é bastante ativo. Neste ciclo, como no ciclo dos ácidos tricarboxílicos, condensa-se acetil-CoA com oxaloacetato para a formação de citrato. Entretanto, a clivagem do isocitrato não ocorre por meio da ação da isocitrato desidrogenase, mas é catalisada pela enzima isocitrato liase (EC 4.1.3.1), na seguinte reação:

Isocitrato → Succinato + Glioxilato

No genoma de *S. cerevisiae* dois genes isocitrato liase (ICL1 e ICL2) foram identificados. ICL1 codifica uma enzima isocitrato liase, localizada no peroxissomo e que participa do ciclo de glioxilato, enquanto que ICL2 codifica uma enzima 2-metilisocitrato liase (EC 4.1.3.30), localizada na mitocôndria e que participa do ciclo de 2-metilcitrato ("Saccharomyces Genome Database", http://www.yeastgenome.org/; Luttik *et al.*, 2000).

II. Objetivo.

A organização do genoma pode refletir o *sistema* de regulação transcricional dos genes, e a análise do arranjo divergente de pares de genes adjacentes é de grande interesse, porque este tipo de organização pode resultar na transcrição co-regulada dos mesmos. *T. reesei* é um fungo filamentoso de importância biotecnológica. Portanto, conhecer sua organização genômica e seus mecanismos de regulação transcricional é uma etapa importante para explorar suas potencialidades metabólicas.

O banco de dados "*Trichoderma reesei* EST database and Mitochondrial genome", construído por nosso laboratório, nos permite identificar e analisar o papel de genes e vias metabólicas presentes em *T. reesei*. Neste banco foi identificado o gene citrato sintase, denominado *TrCit* (de *Trichoderma reesei* citrato sintase). "Upstream" a *TrCit* foi identificado o gene *TrIcl* (de *Trichoderma reesei* isocitrato liase). Ambos os genes seriam transcritos de forma divergente, estando separados por uma região promotora intergênica de 647 pb (*TrIcl/TrCit*). O objetivo deste trabalho consistiu em analisar estrutural e funcionalmente a região promotora *TrIcl/TrCit* e determinar a estrutura e organização dos genes *TrCit* e *TrIcl*.

Para lograr esse objetivo, foi adotada a seguinte abordagem:

- 1- Determinar a estrutura e organização dos genes isocitrato liase (*TrIcl*) e citrato sintase (*TrCit*).
- 2- Analisar a região promotora intergênica *TrIcl/TrCit* e identificar os elementos *cis* e os fatores de transcrição putativos que seriam responsáveis pelo controle transcricional de *TrIcl* e *TrCit*.
- 3- Analisar a habilidade de elementos promotores, presentes na região intergênica, de interagir com fatores de transcrição expressos em *T. reesei*.

III. Material e Métodos.

3.1 Microorganismos utilizados.

A cepa *T. reesei*, QM 9414 foi obtida da "American Type Culture Collection" (ATCC), número 26921.

As bactérias utilizadas foram:

- E coli DL538 (Stratagene): para infecção do bacteriófago Lambda Dash;
- E. coli Top10 (Invitrogen): para propagação de plasmídeos;.
- E. coli DH10B (Stratagene): para construção de biblioteca genômica.

A preparação de células competentes foi realizada conforme o protocolo proposto pelo fabricante (Stratagene, USA).

No laboratório o fungo é mantido em placas contendo meio PDA ("Potato Dextrose Agar"). Para testes de crescimento, esporos foram crescidos em meio Agar peptona (suplementando com 5mM de glicose).

Culturas líquidas do fungo foram realizadas em meio completo (uréia a 0,3%, sulfato de amônio a 0,14%, sulfato de magnésio a 0,03%, tween 80 a 0,2%, proteose/peptona a 0,2%, cloreto de cálcio a 0,03%) suplementado com metais a 0,1% (FeSO₄ 0,5%, MnSO₄ 0,16%, ZnSO₄ 0,16%, CoCl₂ 0,2%), em tampão fosfato de potássio 0,1 M pH 6,0 (El-Gogary *et al.*, 1989), mantido a 30 °C sob agitação constante (300 rpm). O micélio coletado foi filtrado e congelado imediatamente em nitrogênio líquido.

3.2 Manipulação de DNA e técnicas gerais de Biologia Molecular.

Manipulação de DNA com enzimas de restrição, modificação de extremidades, ligação do DNA, clonagem, transferência e hibridização (Southern e Northern), varredura de biblioteca, marcação com material radioativo, eletroforese de DNA, RNA e proteína, isolamento de DNA, preparação de DNA plasmidial, preparação de géis, identificação de colônias recombinantes com utilização de IPTG/X-GAL, ou por hibridização, e outras técnicas utilizadas em biologia molecular foram efetuadas em geral, segundo o descrito por Sambrook *et al.* (1989) e/ou Ausbel *et al* (1992), ou de acordo com os protocolos fornecidos pelos fabricantes dos produtos utilizados.

Transformação de bactérias foi realizada por eletroporação como indicado pelo fabricante do aparelho ("Gene Pulser", BioRad Laboratories).

DNA e RNA de fungo foi preparado como descrito por El-Gogary *et al* (1989). Para amplificações de DNA ou RNA utilizamos kit Taq DNA polimerase e RT-PCR respectivamente (Invitrogen), seguindo instruções do fornecedor e adaptando temperaturas e ciclos de acordo com o T_m dos oligonucleotídeos iniciadores.

Kits para sequenciamento de DNA (GE), marcação de DNA ("Random primers") (Invitrogen), isótopos e membranas Hybond-N⁺ (GE Biosciences) e filmes radiográficos X-OMAT XK-1 (Kodak) foram empregados seguindo as especificações do fabricante.

3.3 Identificação de clone do locus genômico de citrato sintase (TrCit).

Membranas Hybond-N⁺ foram colocadas sobre placas de lise obtidas após a infecção de *E. coli* DL538, com fagos de uma biblioteca genômica de *T. reesei* (Matheucci *et al.*, 1995) e tratadas conforme instruções do fabricante.

A varredura da biblioteca foi feita pela hibridização das membranas, com sondas radioativas obtidas a partir do cDNA do gene citrato sintase (*TrCit*) do clone TrEST-A0857, identificado no "*Trichoderma reesei* EST database and Mitochondrial genome" (http://trichoderma.iq.usp.br) (Chambergo *et al.*, 2002), conforme Sambrook *et al.*,1989. A sonda foi marcada com [α-³²P]-dATP, utilizando o kit "Random Primer". A hibridização ocorreu durante 14 horas em tampão contendo 0,25 M Na₂HPO₄, pH 7,2 / SDS 7% e a sonda radioativa, sob agitação a 65°C.

Após a hibridização foram realizadas 3 lavagens: A primeira com uma solução contendo 2x SSC e 0,1% SDS por 30 minutos; a segunda com uma solução de 1x SSC e 0,1% SDS por 10 minutos; e a última lavagem com 0,1x SSC e 0,1% SDS por 5 minutos. A membrana foi exposta a filme radiográfico por 6 horas antes da revelação do mesmo.

3.4 Montagem e sequenciamento de subbiblioteca do *locus* genômico de citrato sintase (*TrCit*).

Após identificação do clone contendo o lócus genômico de citrato sintase (*TrCit*), o DNA de fago foi preparado como descrito por Sambrook *et al.* (1989). Após sua fragmentação por sonicação, recuperamos, em gel de agarose 1%, os fragmentos entre 0,5

- 2 kb, subseqüentemente tratados com enzima Klenow (Invitrogen), para a obtenção de extremidades cegas e fosforiladas com enzima T4 Kinase(Invitrogen) que pudessem ser ligadas ao vetor pUC18 *Sma*I/BAP (GE Biosciences). Realizadas a ligação e a transformação de *E. coli* DH10B (Stratagene), as colônias selecionadas, utilizando IPTG/X-GAL, foram incubadas em 11 placas de 96 poços, contendo em cada poço 100 μL de meio LB / ampicilina (150 μg/mL), por 18 horas a 37°C. Ao final do período de crescimento foram adicionados 20 μL de glicerol 50% a cada poço e as placas foram estocadas a 70 °C.

O inserto de cada clone da biblioteca foi amplificado por PCR empregando os primers M13 reverso e M13(-20) do vetor pUC18 *Sma*I/BAP. Seqüenciamento, de 1 μL dos amplificados, foi obtido a partir de ambas as extremidades empregando-se os primers do vetor e o "ET terminator sequencing kit", de acordo com a instrução do fabricante (GE Biosciences). O seqüenciamento foi realizado por eletroforese capilar no seqüenciador automático MegaBACE 1000 DNA Analysis System (GE Biosciences).

Os eletroferogramas foram processados e as seqüências com um mínimo de 100 bases com qualidade Phred Q≥ 20, foram consideradas para análise de "clustering". Os arquivos de seqüências em formato fasta, singletos e contíguos, gerados pelo conjunto de programas Phred/Crossmatch /Phrap/Consed (Gordon *et al*, 1998), foram comparados com a base de dados não redundante (NR) do GenBank (NCBI, USA) empregando os serviços de algoritmos blastx, blastn e blastp.

3.5 Análise da região intergênica entre os genes isocitrato liase / citrato sintase (*TrIcl/TrCit*)

Após a análise comparativa com a base de dados do GenBank (NCBI, USA), foi identificado o contíguo 20 (9.045 bases) que continha as seqüências codificadoras dos genes putativos da isocitrato liase (*TrIcl*) e da citrato sintase (*TrCit*), em sentido divergente e separadas por uma região intergênica de 647 pb (*TrIcl/TrCit*), indicando tal organização no genoma do fungo filamentoso *T. reesei*.

Com o intuito de analisar a atividade transcricional da região intergênica *TrIcl/TrCit*, possivelmente envolvida na expressão de ambos genes, utilizamos a fusão gênica desta região com o gene repórter higromicina B fosfotransferase (*hph*). Para tanto,

fragmentos de diferentes tamanhos da região intergênica, foram amplificados por PCR com oligonucleotídeos específicos (tabela 1) e clonados em ambas as orientações (5' \rightarrow 3' e 3' \rightarrow 5') no vetor pHygromicina (pHyg), construído especificamente para este trabalho.

Foram amplificados 10 fragmentos da região intergênica TrIcl/TrCit pela técnica de PCR, a partir do DNA genômico de T. reesei. Oligonucleotídeos (tabela 1), utilizados como iniciadores da síntese foram desenhados a partir da sequência intergênica entre os genes TrCit e TrIcl, contida no contíguo 20 acima mencionado. Os fragmentos de diferentes tamanhos, da região intergênica, foram clonados no vetor pHyg, nas duas possíveis orientações (5' \rightarrow 3' e 3' \rightarrow 5') "upstream" da região codificadora do gene repórter hph, motivo pelo qual temos 20 plasmídeos construídos (tabela 2, plasmídeos de 1 a 20). Fragmentos da região intergênica, que apresentam deleção da caixa GC localizada entre -328 a -333, foram empregados na construção de 4 plasmídeos (tabela 2, plasmídeos 21, 22, 23 e 24).

Após transformação de *E. coli* TOP10 (Invitrogen), as colônias obtidas foram analisadas por PCR, para identificar clones positivos, empregando "primers" específicos para cada fragmento. A orientação dos insertos foi estabelecida através de PCR empregando um primer específico do gene *hph* em combinação com um primer específico do inserto.

A região intergênica *TrIcl/TrCit* foi analisada computacionalmente com a finalidade de determinar seu conteúdo G + C, identificar a presença de caixa TATA, caixa GC e ilhas CpG. Possíveis sítios de ligação para fatores de transcrição foram identificados empregando o servidor do programa TESS ("Transcription Element Search System", http://www.cbil.upenn.edu/tess; Overton and Schug, 1997). O programa TESS é uma ferramenta computacional que realiza a predição de sítios de ligação de fatores de transcrição em seqüências de DNA e foi empregado com os parâmetros padrão. Para determinação do conteúdo G + C e identificação de ilhas CpG foi utilizado o servidor web "CpG plot" (Rice *et al.*, 2000; http://www.ebi.ac.uk/emboss/cpgplot/).

Tabela 1. Oligonucleotídeos empregados na elaboração deste trabalho.

Oligonucleotídeo	Seqüência
1. Pet-3a F	5' -CGTATCACGAGGCCCTTTCG- 3'
2. Pet-3a R	5' -CCGGAAGGAGCTGACTGGG- 3'
3. Hyg F	5' -GAAAAGTTCGACAGCGTCTCC- 3'
4. Hyg R	5' -ACTTCTACACAGCCATCGGTC- 3'
5. A	5' -CTTGAACCGTGTGTAGCTTCTG- 3'
6. B	5' -CAGCAATGCGGAGGAGGGGCC- 3'
7. C	5' -CTGATTTGAGGGCGGTGACGG- 3'
8. C'	5' -CCGTCACCGCCCTCAAATCAG- 3'
9. D	5' -CAAGAAAGAATGAGAATGAC- 3'
10. E	5' -GGGATGCCTTTGCAAGCAG- 3'
11. F	5' -GATTTGGTGTGGGCAAAGGTG-3'
12. G	5' -TCAAATCAGCCGAGGAAAGCC-3'
13. H	5' -TGACGGATCATTCTGGGCGTG-3'
14. ZnFTrSp1F	5'-GAAGGCCTAATGTCTTCTGCTTCG-3'
15. ZnFTrSp1R	5'-CCCAAGCTTTCACTAAGCCATATGGAC-3'
16. TrCit-F	5'-ATGGCTCTCAACCTCA-3'
17. TrCit-R	5'-GCTGCTCTCCACCTG-3'
18. TrIcl-F	5'-ATGCCGCCCTCTGCC-3'
19. TrIcl-R	5'-GTCAAACTGCCCCTCCGTGT-3'
20. OligoAP2	5'GATCGAACTGACCGCCCGCGGCCCGT 3'
(dupla fita)	3'CTAGCTTGACTGGCGGGCGCCGGGCA 5'

Tabela 2. Lista de plasmídeos empregados na elaboração deste trabalho

Plasmídeos	Característica
1. pHygAF	Gene repórter <i>hph</i> (higromicina fosfotransferase) controlado por
1 30	fragmento de DNA de 647 pb 5' flanqueadora do gene citrato
	sintase de <i>T reesei</i> .
2. pHygFA	Gene repórter <i>hph</i> controlado por fragmento de DNA de 647 pb
1 70	5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
3. pHygAE	Gene repórter <i>hph</i> controlado por fragmento de DNA de 486 pb
1 30	5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
4. pHygEA	Gene repórter <i>hph</i> controlado por fragmento de DNA de 486 pb
1 30	5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
5. pHygAD	Gene repórter <i>hph</i> controlado por fragmento de DNA de 426pb
1 30	5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
6. pHygDA	Gene repórter <i>hph</i> controlado por fragmento de DNA de 426 pb
1 70	5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
7. pHygAC'	Gene repórter <i>hph</i> controlado por fragmento de DNA de 326 pb
1 70	5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
8. pHygC'A	Gene repórter <i>hph</i> controlado por fragmento de DNA de 326 pb
1 30	5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
9. pHygBF	Gene repórter <i>hph</i> controlado por fragmento de DNA de 442 pb
1 70	5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
10. pHygFB	Gene repórter <i>hph</i> controlado por fragmento de DNA de 442 pb
	5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
11. pHygBD	Gene repórter <i>hph</i> controlado por fragmento de DNA de 221 pb
	5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
12. pHyg DB	Gene repórter <i>hph</i> controlado por fragmento de DNA de 221pb
	pb 5' flanqueadora do gene isocitrato liase de T. reesei
13. pHygBC'	Gene repórter <i>hph</i> controlado por fragmento de DNA de 121 pb
	5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
14. pHygC'B	Gene repórter <i>hph</i> controlado por fragmento de DNA de 121 pb
	5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
15. pHygCF	Gene repórter <i>hph</i> controlado por fragmento de DNA de 342 pb
	5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
16. pHygFC	Gene repórter <i>hph</i> controlado por fragmento de DNA de 342 pb
	5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
17. pHygCE	Gene repórter <i>hph</i> controlado por fragmento de DNA de 181 pb
	5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
18. pHygEC	Gene repórter <i>hph</i> controlado por fragmento de DNA de 181 pb
	5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
19. pHygCD	Gene repórter <i>hph</i> controlado por fragmento de DNA de 121 pb
	5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
20. pHygDC	Gene repórter <i>hph</i> controlado por fragmento de DNA de 121 pb
	5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>

21. pHygdAF	Gene repórter <i>hph</i> controlado por fragmento de DNA de 641pb 5' flanqueadora do gene citrato sintase de <i>T. reesei</i>
22. pHygdFA	Gene repórter <i>hph</i> controlado por fragmento de DNA de 641 pb 5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
23. pHygdGA	Gene repórter <i>hph</i> controlado por fragmento de DNA de 314 pb 5' flanqueadora do gene isocitrato liase de <i>T. reesei</i>
24. pHygdHF	Gene repórter <i>hph</i> controlado por fragmento de DNA de 327pb 5' flanqueadora do gene de citrato sintase <i>T. reesei</i>
25. pTrZnFSp1	Plasmídeo para expressão da proteína recombinante TrZnFSp1, com cauda de 6 histidinas na extremidade carboxiterminal.
рНуд	Gene repórter <i>hph</i> sem nenhum promotor controlando sua expressão.
pCSN43	Gene <i>hph</i> controlado por promotor constitutivo de triptofano C de <i>Aspergillus nidulans</i> . Terminador de Triptofano C. observação, o plasmídeo foi utilizado como controle positivo das transformações e não foi construída por nos.

3.6 Construção do vetor pHygromicina (pHyg)

Para a análise da região intergênica *TrIcl/TrCit* foi construído um vetor contendo como marcador de resistência / gene repórter o gene *hph* de *Escherichia coli*. Cada nova cepa de *T. reesei*, criada a partir da transformação da cepa QM9414, que consideraremos como selvagem ("wild type"), com uma fusão "região intergênica-gene *hph*" específica, apresentará um fenótipo de resistência a Hyg dependente do nível de transcrição induzido pela região de controle clonada.

O vetor pHygromicina (pHyg) (figura 4) contém como repórter e marcador de resistência o gene de higromicina B fosfotransferase (*hph*). Para sua construção foi empregado o vetor pCSN43, que contém o promotor do gene triptofano C (PtrpC) de *Aspergillus nidulans* controlando a expressão do gene *hph* de *E. coli* e o terminador de transcrição do mesmo genes de *A. nidulans* (TtrpC) após a região codificadora. A clivagem de pCSN43 com a enzima de restrição *Sal*I liberou um fragmento de 2,4 kb contendo o promotor PtrpC, a região codificadora de *hph*, o terminador TtrpC e um fragmento do vetor pBR322. Uma clivagem posterior desse fragmento com *Cla*I liberou um fragmento *Cla*I – *Sal*I de 2,0 kb, contendo a região codificadora de *hph* e o terminador TtrpC. Esse último fragmento foi então clonado no vetor pET3a (Novagen, USA), previamente clivado com as enzimas de restrição *Cla*I – *Sal*I.

3.7 Transfecção de T. reesei.

A transfecção de esporos de *T. reesei* foi feita pela técnica de biobalística, de acordo com Herrera-Estrella *et al.* (1990). A técnica de biobalística consiste na aceleração de micropartículas de ouro, ou tungstênio, recobertas com o DNA exógeno e a dispersão delas sobre células alvo para que atravessem a parede celular e a membrana plasmática de forma não letal. O DNA dissocia-se das micropartículas e é integrado ao DNA genômico. O aparelho que utilizamos para a aceleração das partículas tem propulsão de gás hélio de alta pressão: 1400-1600 psi (9,65 x 10⁶ e 1,10 x 10⁷ Pa) e um sistema de vácuo de 0,1-0,2 atm (24-27 mmHg.).

Na transformação de *T. reesei* empregamos DNA dos plasmídeos construídos (tabela 2) precipitados com CaCl₂ 2,5 M e espermidina 0,1 M, e micropartículas de tungstênio M5 com diâmetro médio de 0,2 μm (BioRad). Esporos de *T. reesei* obtidos

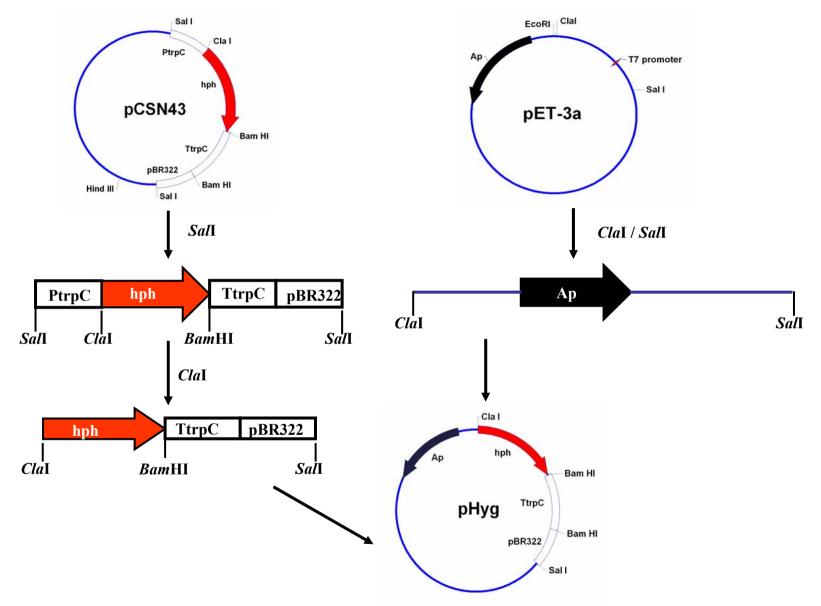


Figura 4. Construção do vetor pHygromicina (pHyg). pHyg contém o fragmento de 2 kb *Cla*I –*Sal*I do vetor pCSN43, clonado no sítio de *Cla*I e *Sal*I do vetor pET-3a. *hph*: gene de higromicina B fosfotransferase; PtrpC: promotor de triptofano C; TtrpC: terminador de triptofano C; Ap: gene de resistência a ampicilina.

após 7 - 10 dias de cultivo foram espalhados sobre placas de Agar Peptona Glicose (APG, contendo 5 mM de glicose), $\approx 10^7$ esporos por placa, que permaneceram, então, abertas em um ambiente estéril por 4 horas para inicio da germinação dos esporos.

Após o bombardeio com as micropartículas revestidas com DNA, as placas foram incubadas a 28 °C por 7 h. Adicionou-se, então, uma camada de APG suplementada com o antibiótico higromicina B (Hyg) para uma concentração final de 100 μg.mL⁻¹.

Colônias dos primeiros transformantes surgiram de 2 a 3 dias após a transformação. Na etapa de seleção, os transformantes foram transferidos, individualmente, para uma nova placa de APG suplementada com Hyg (100 µg.mL⁻¹). Após a repetição desse procedimento por 3 vezes, os transformantes estáveis foram finalmente transferidos para uma placa contendo apenas PDA (Potato Dextrose Agar) e incubados por 7 dias para que ocorresse a esporulação.

Os esporos dos transformantes selecionados foram coletados e plaqueados novamente em meio seletivo com Hyg. Nesta etapa, uma grande porcentagem dos transformantes não foi capaz de germinar na presença do antibiótico. Os que se desenvolveram foram submetidos a uma nova passagem no meio seletivo e considerados, a partir desse momento, como transformantes estáveis.

3.8 Análise do crescimento dos transformantes.

A medida do crescimento radial dos transformantes estáveis foi obtida pelo inóculo de esporos de 7 - 10 dias de cultivo (~10⁴ esporos) no centro de placas de Petri de 35 mm de diâmetro, contendo APG-5 mM de glicose / Hyg - 100 μg.mL⁻¹ seguida de incubação a 28 °C. O crescimento foi acompanhado e o diâmetro do halo de hifas determinado a cada 24 horas, durante 3 dias. A análise foi realizada por triplicata para a determinação da média do crescimento radial de cada transformante.

3.9 Análise da expressão gênica dos genes TrCit e TrIcl.

Foi realizada uma pré-cultura de *T. reesei* QM9414, com inóculo de 1,0 x 10⁷ esporos.mL⁻¹, em 100 mL de meio fungo completo (El-Gogary et al., 1989) contendo glicerol 1% como fonte de carbono, em frasco erlenmeyer de 1,0 L por 16 horas a 28°C sob agitação de 200 rpm em agitador orbital. Após centrifugação para substituição do

meio e lavagem da massa micelial em tampão fosfato 100 mM pH 6,0, a cultura foi incubada em tampão fosfato por 1 hora a 28°C sob agitação de 200 rpm. Passado este período a massa obtida por centrifugação foi suspensa em meio mínimo contendo como fonte de carbono; glicose (5mM ou 250mM), acetato de sódio (10mM ou 50mM) e etanol (5% ou 1%) e incubada por 6 horas a 28°C sob agitação de 200 rpm. A massa micelial foi coletada e conservada a -80 °C para extração de RNA, seguindo o protocolo do kit de Trizol (Invitrogen, USA) e análise por northern blot (Sambrook *et al.*, 1989). Sondas foram obtidas a partir da amplificação da seqüência codificadora dos genes *TrCit* e *TrIcl* por RT-PCR, empregando oligonucleotideos específicos (tabela 1, oligonucleotídeos 16, 17, 18 e 19).

3.10 Clonagem e expressão do fator TrZnFSp1.

A seqüência codificadora da proteína TrZnFSp1 (de *Trichoderma reesei* "zinc-finger specific protein 1") de 1.347 pb, que codifica 448 aminoácidos (peso molecular estimado de 48,4 kD), foi obtida por RT-PCR (Invitrogen), empregando oligonucleotídeos específicos (tabela 1, oligonucleotídeos 14 e 15) e foi inserida no vetor de expressão pPROEX-Hta (Invitrogen), com cauda de histidina na extremidade 5'. A proteína recombinante foi expressa em *E. coli* BL21, nas seguintes condições.O inóculo foi realizado em meio LB / Ampicilina (150 μg.mL⁻¹), a 37 °C, e acompanhado até atingir OD₆₀₀ ≈ 0,8 – 1,2. Seguiu-se a indução da expressão da proteína recombinante por IPTG (0,5 mM), por 6 horas. A massa bacteriana foi, então coletada e rompida empregando "French Press" a 16.000 psi, em tampão de lise (Tris-HCl 20 mM, pH 8,0). Após centrifugação do lisado, o sobrenadante e o sedimento foram analisados por eletroforese em gel de SDS/PAGE. O sobrenadante protéico foi conservado a -20°C em glicerol 20%.

3.11 Ensaio de retardamento de mobilidade eletroforética (EMSA - "electrophoretic mobility shifting assay").

Para determinar a interação funcional entre o putativo fator de transcrição e a seqüência específica de DNA da região intergênica, utilizamos a técnica de retardamento de mobilidade eletroforética (EMSA). A análise EMSA consiste na identificação da interação proteína-DNA pelo retardamento observado na mobilidade eletroforética num gel de poliacrilamida / bisacrilamida não desnaturante de uma sequência específica de

DNA pré-tratada com a proteína em estudo em relação à da mesma sequência livre ou pré-tratada com outras proteínas que não interagem com DNA. A análise proporciona um método simples para determinar a especificidade e afinidade de ligação de uma proteína a uma sequência de DNA.

Oligonucleotídeos de sequência específica (1.75 pmol. μ L⁻¹), (tabela 1, oligonucleotídeos 7 e 8; figura 17), foram anelados, marcados em suas extremidades 5' com γ -³²P (ATP) e incubados com a proteína recombinante TrZnFSp1. A especificidade da ligação foi testada pela pré-incubação da proteína com oligonucleotídeos competidores, ou não—competidores, não-marcados, antes da adição dos oligonucleotídeos marcados radioativamente.

Reação de interação proteína-DNA foi realizada segundo as instruções do fabricante do kit "Gel shift assay core system" (Promega). Foram montadas as seguintes reações:

1-	Controle negativo:	
	Proteína recombinante (~5 μg).	$0 \mu L$
	Buffer de ligação 5x.	2 μL
	Água bidestilada estéril	, 7 μL
	Vol. total	9 μL
2-	Controle positivo:	
	Proteína recombinante (~5 μg).	$2 \mu L$
	Buffer de ligação 5x.	2 μL
	Água bidestilada estéril	5 μL
	Vol. total	9 μL
•		
3-	Competidor específico:	
	Proteína recombinante (~5 μg).	$2 \mu L$
	Buffer de ligação 5x.	2 μL
	Oligo competidor não-marcado (1.75 pmol.µL ⁻¹).	3 μL
	Água bidestilada estéril	2 μL
	Vol. total	9 μL
4-	Competidor não específico:	
	Proteína recombinante (~5 μg).	$2 \mu L$
	Buffer de ligação 5x.	$2 \mu L$
	Oligo competidor não-específico AP2 não-	3 μL
	marcado (1.75 pmol. μ L ⁻¹).	·
	Água bidestilada estéril	$2 \mu L$
	Vol. total	9 μL

As reações foram incubadas por 10 minutos a temperatura ambiente e posteriormente foi adicionado $1\mu L$ de oligonucleotídeo específico marcado com $\gamma[^{32}P]$ -ATP. A incubação transcorreu por 20 minutos a temperatura ambiente e subsequentemente as reações foram aplicadas num gel de poliacrilamida / bisacrilamida 6% não desnaturante, para o desenvolvimento da eletroforese, seguindo as instruções do fabricante do "kit Gel shift assay core system" (Promega). Os géis foram secos e expostos a filme radiográfico (X-OMAT, Kodak) até o aparecimento dos sinais.

IV. Resultados

4.1 Sequenciamento e análise do gene citrato sintase de T. reesei (TrCit)

Utilizando-se como sonda o cDNA do gene de citrato sintase, proveniente do clone TrEST-A0857, identificado no "*Trichoderma reesei* EST Database and Mitochondrial Genome" (http://trichoderma.iq.usp.br), foi realizada a varredura de uma biblioteca genômica de *T. reesei* (El-Gogary *et al.*, 1989), o que permitiu a identificação de um fago contendo o *locus* genômico do gene. O DNA deste fago foi preparado conforme descrito em Sambrook *et al.* (1989) e, após fragmentação por sonicação, os fragmentos de peso molecular entre 0,5 e 2,0 kb foram clonados no vetor pUC18 *SmaI*/BAP para a criação de uma subbiblioteca de seqüenciamento.

Foram seqüenciados 960 clones da subbiblioteca genômica de citrato sintase, obtendo-se 1.705 seqüências a partir de ambas as extremidades dos clones. Após análise de "clustering", com o pacote phred/phrap/crossmatch/consed, 20 contíguos foram formados: o menor contendo 224 bases e o maior, 9.045 bases. Para a determinação da similaridade das seqüências obtidas, foi empregado o algoritmo de comparação BLASTX (Altschul *et al.*, 1990), através do serviço fornecido pelo Centro Nacional de Informação Biotecnológica dos E.U.A. (NCBI; www.ncbi.nlm.nih.gov). Este algoritmo traduz uma seqüência de nucleotídeos nas 6 possíveis fases de leitura e realiza uma comparação das seqüências traduzidas com a base não-redundante de seqüências de proteínas. Em nosso caso, empregamos o banco público de dados GenBank do NCBI.

Após a análise por BLASTX foram identificadas no contíguo 20, composto por 9.045 bases e formado pela sobreposição de 173 seqüências (figura 5), as seqüências codificadoras dos genes de putativas citrato sintase e isocitrato liase, denominados respectivamente, de *TrCit* e *TrIcl*. Tais seqüências codificadoras encontram-se separadas por uma região intergênica de 647 bases (figura 6), sugerindo que os dois genes de *T. reesei* poderiam ser transcritos de forma divergente.

A figura 7 mostra a sequência nucleotídica e de aminoácidos (figura 7a) e a estrutura gênica (figura 7b) previstas para o *locus* que contém os genes citrato sintase e isocitrato liase do fungo filamentoso *T. reesei*.

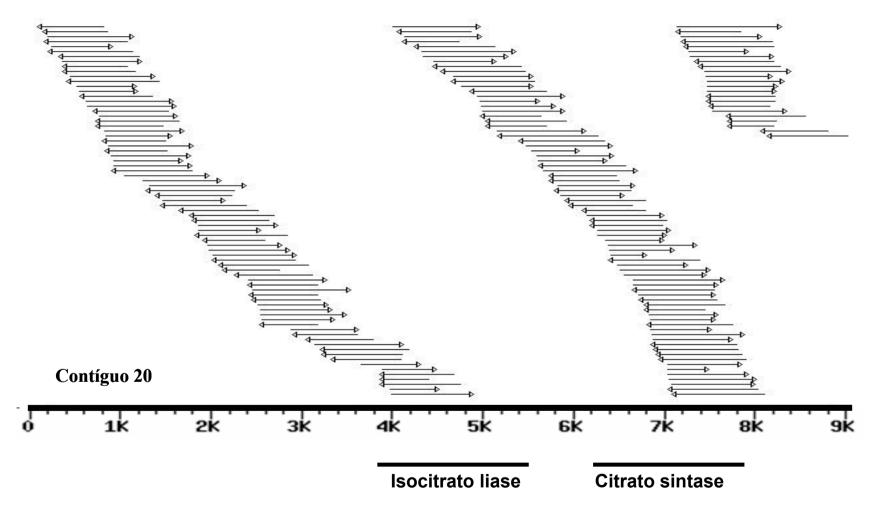


Figura 5. Montagem do contíguo número 20 da biblioteca genômica de citrato sintase. As sequências obtidas, após o sequênciamento, são analisadas pelo conjunto de programas Phred/Crossmatch /Phrap/Consed (Gordon *et al*, 1998), gerando sequências singletos e contíguos. O contíguo número 20 é formado pela sobreposição de 173 sequências.

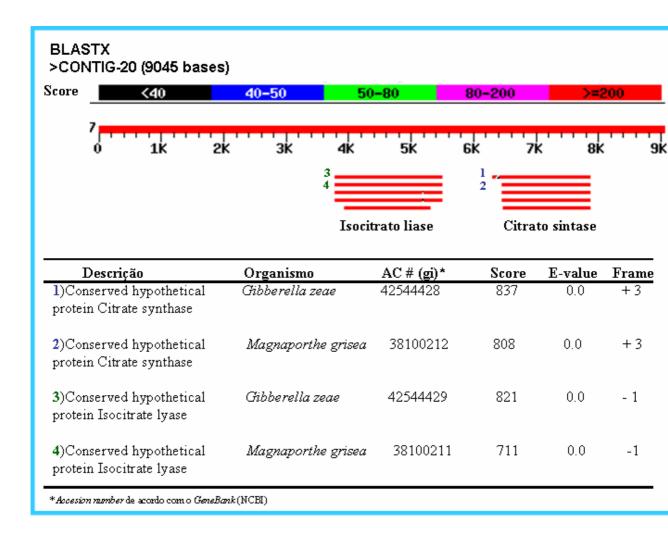


Figura 6. Identificação dos genes para Isocitrato liase e Citrato sintase de *T. reesei*. São mostrados os cinco melhores alinhamentos obtidos na análise do contíguo 20 pelo algoritmo BLASTX e descritos os dois primeiros alinhamentos para cada gene (Adaptação de resultados fornecidos pelo serviço BLASTX do NCBI).

A análise comparativa do cDNA e do DNA genômico de citrato sintase revelou a presença de um gene de 1.573 pb, contendo 3 éxons (de, respectivamente, 48, 881 e 493 pb) e 2 íntrons (de, respectivamente, 73 e 78 pb) (figura 7b); a seqüência codificadora possui 1.422 pb e produz uma proteína de 474 aminoácidos, com um peso molecular estimado de 52,3 kD.

A fim de se avaliar a similaridade da proteína TrCIT com proteínas citrato sintase de diferentes organismos, através de alinhamentos completos, utilizamos o programa FASTA, com a matriz de comparação BLOSUM 50 (http://wrpmg5c.bioch.virginia.edu /fasta_www/cgi/search_frm2.cgi). Foram obtidos valores significativos de similaridade de TrCIT com proteínas citrato sintase já caracterizadas funcionalmente. Assim temos que TrCIT apresenta similaridade de 79% com NcCit1p, de *Neurospora crassa* (XP_956898); 80% com PaCitp, de *Podospora anserina* (CAC12961); 79% com AnCitAp, de *Aspergillus nidulans* (CAB77625); 80 % com ScCit2p, de *Saccharomyces cerevisiae* (NP_009931); 80% com ScCit1p, de *Saccharomyces cerevisiae* (NP_014398) e 80% com ScCit3p, de *Saccharomyces cerevisiae* (NP_015325). A similaridade das seqüências de proteínas citrato sintase é mostrada através do alinhamento múltiplo das proteínas entre si. O alinhamento foi feito com o programa ClustalX (Thompson *et al.*, 1997), utilizando-se matrizes da série BLOSUM, e editado pelo programa GeneDoc (Nicholas *et al.*, 1997) (figura 8).

4.2 Sequenciamento e análise do gene isocitrato liase de T. reesei (TrIcl)

A putativa proteína isocitrato liase de *T. reesei* (TrICL) seria codificada por um gene de 1.880 pb, contendo 3 éxons (de, respectivamente, 116, 281 e 1391 pb) e 2 íntrons (de, respectivamente, 30 e 62 pb) (figura 7b); a seqüência codificadora de 1.788 pb produz uma proteína de 596 aminoácidos, com um peso molecular estimado de 65,4 kD.

Foram obtidos valores significativos de similaridade de TrICL com proteínas isocitrato liase já caracterizadas funcionalmente. Assim temos que TrICL apresenta similaridade de 74,5% com ScIcl1p, de *Saccharomyces cerevisiae* (NP_010987); 74,5% com CtIcl1p, de *Candida tropicalis* (P20014); 75% com AnIclp de *Aspergillus nidulans* (XP_663238); 75% com NcAcu3p, de *Neurospora crassa* (XP_323570); 83,8% com

-2531	* D F Q G E T N G D G M S K S S S G S Q I A G L I G D I Y AGTCAGTTGACGGGGGACCAATGGTAGAGAGGGTATGAGAACGACGAGCTCGACGGCGAGACCTATCGGGGGTCCTACGGCAGTTACAT	568
-2441	S A G S W K Q H T L V D C G L E K E R R Q V L E V Y A K M G CGAACGAGGCGAGGTGAAGACCACGCAGTCGTGCAGCGTGGGCTCGAGGAGAGAGCGCGGAAGACGTGGTCGAGCTGCATGCGGAAGTACGG	538
-2351	D T K F K K A L E N T I T A T S H L G A L S V L Q L T F G E CAGGCAGAACTGAAGAACCGGTCGACGACCACCACCACCGGCAGCTCACGTCCGGGCGGTCGCTGTGCTCGACGTCGACGTCCGGAG	508
-2261	K A I D W I F S K L T E D T F G H A M W N F S P S L N Y V L GAACCGTTACAGGGTCTACTTCTGAACTCGCAGAGCACGCGCACGCGGTAGGTCAACTTGCTACCGCTGTTCAACATGTGCTC	478
-2171	V K G P H A A R V T R G L K A A V E V N P D G T E V W L L D GTGGAATGGCCCCACCCGTCGGGCATGCCACCCGGGTCGAACCGCGGTGGAGCTGCAACCCCAGCGGGCAGAGCTGGGTCTCCAG	
-2081	A F P A F A L A R K T A A E M G A R F H Y Y G E R T R P V D GCGTTTCCCCCGTTTCCGGCGCGCGCACCCCCGCGCGGAGCTACGGCCGCCCCTCACCATCATCGGGAGCGCGCACGCTCCGTGCAG	
-1991	W N F Y V P T K T Y Q A A L A R R R S I G M D R D A A T A S GGTCAATTTCATCTGCCCGCAGAACCACATGACGCCCGCC	
-1901	L Y E D I K A Q A V G E K Q M H E V A A E D F T V L R S S K GTCCATGAGCAGTTAGAACCGGACCGCTGTGGGAGGAAGACGTACACGAGCTGCCGTCGAAGCAGTTTGCAGTGCTCGGCCGACGAGAA	
-1811	V W K A E F A D I E A G T A G K A E M A Q L T E A L P E I S GTGGGTGAACCGGAGCTTCCGCAGTTAGAGCCGGGACACCGAGGGAACCGGAGGTACCGGACGTCGCAAAGCCGGTCCCCGAGCTACCT	
-1721	P D A I G L I F E H D R V D I A S S L L R G S E S D T R A I TCCCAGCCGCTACGGGTCCTACTGAGCACCAGCCGTCAGCCATGCCCGTTA	
*1631	V L N E S G M V D W Q F R A A N L R N I H E G T P V L V K G CTGCTCCAAGAGGGTCGGGTACTGTAGGGTGACCTTCGCGCGCAACTCTGCCAATTACACGAGCGGGCACCCGTGGTCGTAGAACGG	
-1541	A L H G C K K G G H L Q D E F H V A A A G N E A F L K A L K GCGGTCCACCGGCGTGAAGAGTCGTTTGTCGAACCGCTCGAA	
-1451	L V A S L G G H G T D G D A I I P R L Y D I Y P T K A R Q E CTCGTGGCGGCTTTCTGGAGGCACCGGTCACAGCCGTTACTACCCAGCCTCTATCAGTTACATGCCGCAGAACCGGGCAACAAG	
-1361	P T M K R R M D W Q K R D H M S Q A K A L R Q V Q N P V T N GCCGCAGTAGAACGCCGTACAGAGCCAGAACGCGAAACCGGTCCGCGACCAGACCAAACCGCTAGACCAAGCCCTGCCACAA	
-1271	Y P Y D G F D P S V E N T S T L V S S C A W G S I Y L V E Q CATGCCCATCAGCGGTTTCAGTCCGCTTTGGAGCAACCACGACCATTCCTGCCTACTCGGGGTCGGCCTCTACATCTCCTGGAGAAC	
-1181	H P A Q Q T M Q V P D I A G M INTRONII CACTCCGCGGACGACCCAGTAGACCTGCCCCAGTTACCGCGGGTgaccaactccttctgtctaatcgctgagaagtgttcttgaacgtag	133
-1091	THLPEGKAERERILNFLKTAMVSS ttaccatacctgaatgACCATACCTCCCCGAGCGGAACCGAAGCGCGAGCGCCTAGTCCAACTTGTCGAACCACCGGTACTGTCTCGA	109
-1001	PYKFFQSGRKSAVDAASYPRRIGAFRPSAW GCCTATGAACTTCTTGACCGACGCGGAACCTCCGCTGCAGACGCCGCGACATGCCCGCAGCTTACGGCCGTTTCGCTCCGCTACGGGT	79
-911	W E E I E K V Q A E Y L A D E A P G T K Q S E P L L Q Y A D GGTGAGGAGTTAGAGGACCGGGCCGGACAGAAAACGCTGAGACCGTCGACTATGCGCAG	49
-821	A P P V A R P G S I N T R O N I T A I A R I A A S P P ACGGCCGCGTGGCGCGCCCGGCCTCgacctccctcgtcggtacgctcgtcgtctgCAACGTTACCGCGCATATCGCCGTCTCCCGCC	29
-731	M S R R L A A A S S S T L S C G R Q S A R S A I R M M CTTGAA GTAGCTAGCAGCTTCTCGCCGGCGCGCTAGTAGAACTT $Trlcl \longleftarrow$	1
-641	CCGTGTGTAGCTTCTGGGAGACGGCGGATCGAGGTTGGAATTACCGAGGCAAACAATGTGTTTCGTCACTCAAGAACCCAGTTATGTAGC GGCACACATCGAAGACCCTCTGCCGCCTAGCT <mark>CCAACC</mark> TTAATGGCTCCGTTTGTTACACAAAGCAGTGAGTTCTTGGGTCAATACATCG Inr	
-551	CAAGACACTCGTGCCTCCGGCCATGCGGCTCTTCTGCTTCCTCTTCCAGGGCGTCGACTTTATGAAGTGCTCGGCTTAATAATTCC GTTCTGTGAGCACGGAGGCCGGTACGCCGAGAGAGAGAGA	
-461	ATTCGACCTTGAATCGCCTCAGCAATGCGGAGGAGGGGCCGATAAAGGCGCGGCTGAGCCCAAATTGACCAATGGGGCCCCGCGGGGCTA TAAGCTGGAACTTAGCGGAGTCGTTACGCCTCCTCCCCGGCTATTTCCGCGCCGACTCGGGTTTAACTGGTTACCCCGGGGCGCCCCGAT	

	CCGACGGACCGAGCGGCTTTCCTCGGCTGATTTGAGGCCGGTGACGGATCATTCTGGGCCTGTACCTGGAGGAAGAGGGAGG	
-371	ggctgcctggctcgccgaaaaggagccgactaaact $ccgcc$ actgcctagtaagacccgcacatggacctccttctcctccta $cacc$ aca $caccc$ aca $cacc$ aca ca	
-281	ATGGAGGAACTGCTGGGATCAATATCTCGGGGTCTTTTTGTCATTCTCATTCTTTTTTTT	
-191	CTGATTACCGCTGCTTGCAAAGGCATCCCCGCGTTCCTCTGCCTCTTGTCGCTGTGTAAGTGAATGTCACTCCTGGAGCTTATCGGCGATGACTAATGGCGACGAACGGACACACTTCACTTACAGTGAGACCTCGAATAGCCGCTA	
-101	AGCATCATCATGCAGCTCCCATAGCTAACCGCTTCTTCTGCCCCGCCATTCAGGAGCAGCTCGCCTTGTTCATTTATCCTCACCTTTGCC TCGTAGTAGTACGTCGAGGGGTATCGATTGGCGAAGAAGAGCGGGGCGGTAAGTCCTCGTCGAGCGGAACAAGTAAATAGGAGTGGAAACGG	
-11	TrCit CACACCAAATCATGGCTCTCAACCTCACCTCGTCGGCTCGAGCCCTGCGCTCCTTCAAGGtcagtcgagtctcctcttatagccttctat GTGTGGTTTAG M A L N L T S S A R A L R S F K	16
80	cgagacccatttgacatagactgacactcctcccgtctacagCCCTACACCCGCGCCCCTCCTCGCCAACGCCGCGCGATGCTACTCA INTRONI PYTRAALLANAARCYS	32
170	ACCGCTGAGCCCGACCTCAAGACGACGCTCAAGAGCGTCATCCCTGAGAAGCGCGAGCTGCTGAAGAAGGTCAAGGCCCACGGCAGCAAG T A E P D L K T T L K S V I P E K R E L L K K V K A H G S K	62
260	GTCATTGGCGAGGTCAAGGTTGAGAACACCATTGGCGGCATGCGCGGGCTCAAGGCCATGGTCTGGGAGGGCTCCGTGCTCGACCCCAAC V I G E V K V E N T I G G M R G L K A M V W E G S V L D P N	92
350	GAGGGCATCCGCTTCCACGGCCGCACCATCAAGGACTGCCAGAAGGAGCTGCCCAAGGGCAAGACGGGCACCGAGATGCTGCCCGAGGCC E G I R F H G R T I K D C Q K E L P K G K T G T E M L P E A	122
440	ATGTTTTGGCTGCTGTTGACCGGCCAGGTGCCCTCCGTCAACCAGGTCCGCCAGTTCTCCCGCGAGCTGGCCTCCCAGACCCAGATCCCC M F W L L L T G Q V P S V N Q V R Q F S R E L A S Q T Q I P	152
530	GCCTTCGTCAACAGGATGCTCGACGATTTCCCCAAGGATCTGCACCCCATGACCCAGTTTGCCATTGCCGTCTCGGCCCTCAACTACGAG A F V N R M L D D F P K D L H P M T Q F A I A V S A L N Y E	182
620	TCCAAGTTTGCAAAGGCCTACGAGAAGGGCCTCGCCAAGGCCGACTACTGGGAGCCCACCTTTGACGACAGCATCTCGCTGCTCGCCAAG SKFAKAYEKGLAKADYWEPTFDDSISLLAK	212
710	CTGCCCACCATTGCCGCCAAGATCTACCAGAACTCTTACCGCGGCGGCGCGCCCTCCCT	242
800	TACAACTTTGCTGCCATGCTCGGCAAGGGCGCAAGGAGAACGAGGACTTCCAGGACCTCCTCCGTCTCTACCTTGCCCTCCACGGCGAC	272
890	CACGAGGGTGGCAATGTGTCTGCCCACGCCACTCACCTTGTTGGTAGTGCCCTGAGTGACCCCTTCCTGTCTTACAGCGCTGGTCTCCAG	302
980	GGTCTGGCCGGTCCTCTCACGGgtaagctgtccacttatacttacaatcatcttcctatacaaacgttgattgctaatcgttcgcgct G L A G P L H G I N T R O N II	310
1070	tatgacgatagACTTGCCGCCCAGGAAGTTCTCCGCTGGATCCTGCAGATGAAGGAGGCCATCCCCGCCAACTACACCGAGCAGGATGTC	
1160	CACGACTACCTCTGGTCCACCCTCAACTCGGGCCGCTCGTGCCCGGCTACGGACACGCCGTTCTGCGCAAGCCCGACCCTCGATTCGAG	336
1250	H D Y L W S T L N S G R V V P G Y G H A V L R K P D P R F E GCTCTCATGGACTATGCCGCTTCCCGCCCCGCGATTGCCAAGGACCCCGTCTTCCAGCTGGTTGAGAAGAACAGCCGCATCGCCCCCGAG	366
1340	A L M D Y A A S R P A I A K D P V F Q L V E K N S R I A P E GTGCTCAAGAAGCACGGCAAGAACCCCTACCCCAACGTCGACAGCAGCTCCGGCGTCCTCTTCCACCACTACGGCTTCCACGAG	396
	V L K K H G K T K N P Y P N V D S S S G V L F H H Y G F H E	426
1430	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	456
1520	GAGCGCCCCAAGAGCATCAACCTCGAGGGTATTCTGAAGCAGGTGGAGAGCAGC TAA E R P K S I N L E G I L K Q V E S S *	474



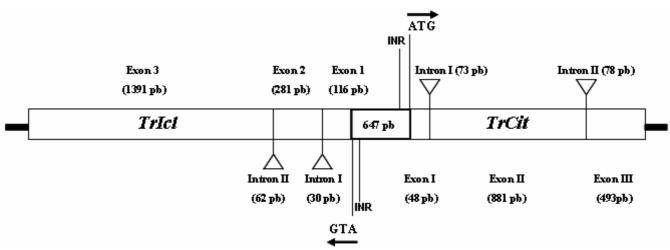


Figura 7. Sequência de nucleotídeos e de aminoácidos (A) e estrutura gênica (B) prevista para o *locus* que contém os genes citrato sintase (*TrCit*) e isocitrato liase (*TrIcI*) do fungo filamentoso *Trichoderma reesei*. Os códons de iniciação e terminação estão indicados em negrito. Os íntrons identificados na sequência codificadora das proteínas estão representados por letras minúsculas. A região intergênica é mostrada como sequência de dupla fita. A numeração à esquerda corresponde à posição dos nucleotídeos em relação à adenina do códon de iniciação da proteína citrato sintase; os aminoácidos estão numerados a direita.

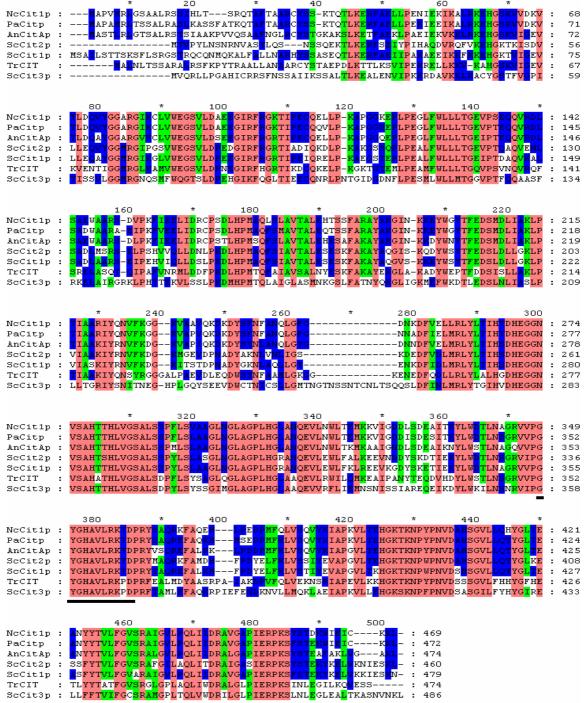


Figura 8. Alinhamento da seqüência de aminoácidos de proteínas citrato sintase. TrCIT, de *T. reesei*; NcCit1p, de *Neurospora crassa* (XP956898); PaCitp, de *Podospora anserina* (CAC12961); AnCitAp, de *Aspergillus nidulans* (CAB77625); ScCit2p, de *Saccharomyces cerevisiae* (NP009931); ScCit1p, de *Saccharomyces cerevisiae* (NP014398); ScCit3p, de *Saccharomyces cerevisiae* (NP015325). Os aminoácidos que são idênticos ou conservados são marcados num fundo colorido. Números na coluna direita, indicam o último aminoácido em cada linha da respectiva seqüência de aminoácidos. A assinatura conservada está sublinhada (Prosite, PDOC00422).

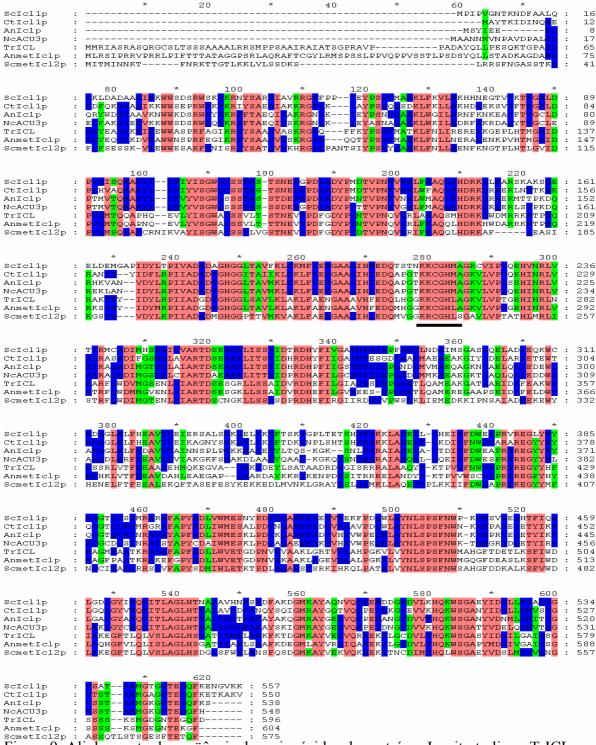


Figura 9. Alinhamento da seqüência de aminoácidos de proteínas Isocitrato liase. TrICL, de *T. reesei*; ScIcl1p, de *Saccharomyces cerevisiae* (NP010987); CtIcl1p, de *Candida tropicalis* (P20014); AnIclp de *Aspergillus nidulans* (XP663338); NcAcu3p, de *Neurospora crassa* (XP323570); AnmetIclp, de *Aspergillus nidulans* (CAI65406); ScmetIcl2p, de *S. cerevisiae* (NP015331). Os aminoácidos que são idênticos ou conservados são marcados num fundo colorido. Números na coluna direita indicam o último aminoácido em cada linha da respectiva seqüência de aminoácidos. A assinatura conservada está sublinhada (Prosite, PDOC00145).

AnmetIclp, de *Aspergillus nidulans* (CAI65406) e 80% com ScmetIcl2p, de *S. cerevisiae* (NP_015331). A similaridade das seqüências de proteínas isocitrato liase é mostrada através do alinhamento múltiplo das proteínas entre si. O alinhamento foi feito com o programa ClustalX (Thompson *et al.*, 1997), utilizando-se matrizes da série BLOSUM, e editado pelo programa GeneDoc (Nicholas *et al.*, 1997) (figura 9).

4.3 Análise do padrão de expressão transcricional dos genes TrCit e TrIcl

O padrão de transcrição dos genes *TrCit* e *TrIcI* foram analisados, por northern blot, na cepa "selvagem" QM9414 de *T. reesei* crescida em culturas de meio mínimo, contendo diferentes fontes de carbono: 250 mM ou 5 mM glicose, 50 mM ou 10 mM acetato de sódio, ou etanol 5% ou 1%. A figura 10 mostra que o gene *TrCit* é parcialmente reprimido em alta concentração de glicose, e induzido em baixa concentração (colunas 1 e 2). Este perfil de expressão indica o envolvimento do gene *TrCit* no ciclo do ácido cítrico (Chambergo *et al.*, 2002). O gene *TrIcI* apresenta um padrão de expressão constitutivo, sendo parcialmente induzido quando o fungo é crescido em meio contendo acetato ou etanol (colunas 3 e 5, respectivamente). Este perfil de expressão indica o envolvimento do gene *TrIcI* no ciclo do glioxilato (Haurie *et al.*, 2001).

É de interesse notar que, numa análise comparativa do perfil de expressão de *TrCit* e *TrIcl*, como mostrado na figura 10, podemos observar que, nas várias condições analisadas, o gene *TrCit* é mais expresso que o gene *TrIcl*.

4.4 Análise da região intergênica TrIcl/TrCit

A região intergênica de 647 pb entre os genes de isocitrato liase e de citrato sintase (*TrIcl/TrCit*), como mostrado na figura 7a, não apresenta caixa TATA e apresenta dois sítios INR putativos, localizados em -167 na fita senso e em -606 na fita anti-senso. Sua composição de bases nitrogenadas é de 19% A, 27% T, 28% C e 26% G, com um conteúdo G + C de 54 %, sendo de tal forma considerada uma seqüência rica em GC (Clark *et al.*, 2002).

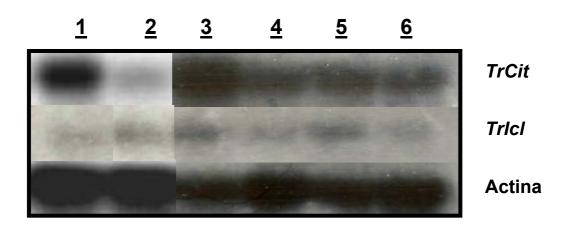


Figura 10. Análise do padrão de expressão transcricional dos genes *TrCit* e *TrIcI* de *Trichoderma reesei*. Northern blot de 10 μg de mRNA isolado de massa micelial do fungo crescido em meio mínimo, contendo diferentes fontes de carbono, como descrito em materiais e métodos. Os números indicam a condição de crescimento, como segue: 1) Glicose 5 mM, 2) Glicose 250 mM, 3) Acetato de sódio 50 mM, 4) Acetato de sódio 10 mM, 5) Etanol 5%, 6) Etanol 1%. O gene de actina foi empregado como controle interno.

O dinucleotídeo CpG é definido como uma citosina seguida de uma guanina, onde o "p" refere-se ao grupo fosfato que liga as duas bases. Os programas computacionais determinam o conteúdo CpG analisando janelas, ou regiões, de DNA de tamanho

determinado. Numa certa janela de análise de uma seqüência de DNA, o número observado de CpG é definido como o número de vezes que uma base "C" encontra-se seguida imediatamente de uma base "G". O número aleatório esperado de CpG, numa dada janela, é definido como o número de dinucleotídeos CpG esperados a partir do conteúdo de "G" e "C" desta janela. Assim, a freqüência esperada de CpG numa janela é calculada como o número de "Cs" multiplicado pelo número de "Gs" na janela, dividido pelo total de bases

CpG esperado = (N° Cs x N° Gs) / N° total de bases da janela O tamanho padrão da janela de análise é de 100 bases.

Após os valores de G + C e a freqüência CpG serem calculados para uma determinada região de DNA, a janela de análise desliza-se um nucleotídeo a frente na seqüência e uma nova janela é analisada. Um mínimo de 10 janelas deve ser analisado antes de se determinar a presença de uma ilha CpG. Assim, uma ilha CpG é definida como a região de DNA de 200 bases na qual um mínimo de 10 janelas de 100 bases apresentam um conteúdo G + C maior de 50%, e um valor CpG observado / CpG esperado maior de 0,6 (Larsen *et al.*, 1992).

Realizou-se, então, uma análise da região intergênica *TrIcl/TrCit* para identificação de ilhas CpG por meio do servidor web "CpG plot" (Rice, *et al.*, 2000; http://www.ebi.ac.uk/emboss/cpgplot/), que culminou na determinação de uma ilha CpG dentro dessa região (figura 11).

Para a análise funcional do possível promotor bidirecional na região intergênica *TrIcl/TrCit* foi construído o plasmídeo pHygromicina (pHyg) (figura 4), que contém como repórter e marcador de resistência o gene de higromicina B fosfotransferase (*hph*). Esta enzima inativa o antibiótico higromicina B, por fosforilação (Freitag *et al.*, 1995). Por meio da clonagem duma seqüência de DNA do promotor a ser testada em posição "upstream" à região codificadora do gene repórter, pode-se avaliar a habilidade de cada uma das construções em gerar transformantes capazes de germinar na presença do antibiótico.

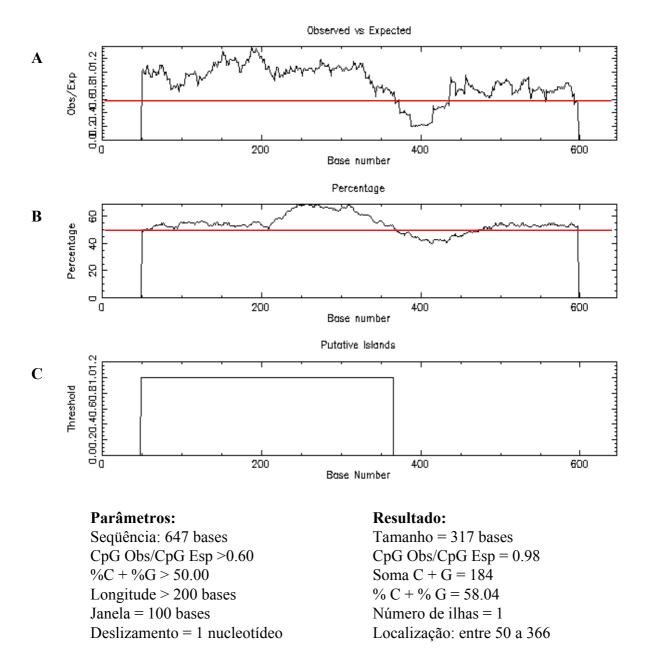


Figura 11. Identificação de ilhas CpG na região intergênica de 647 bases *TrIcl/TrCit*, utilizando o programa CpGPlot (http://www.ebi.ac.uk/emboss/cpgplot/#). É mostrado o resultado da análise, utilizando os parâmetros acima indicados. A) Freqüência CpG observado / CpG esperado, a linha vermelha indica o menor valor a ser considerado, B) Porcentagem C+G, a linha vermelha indica o menor valor a ser considerado, e C) Janela de localização da ilha CpG, na região intergênica.

Para cada fragmento do promotor estudado foram realizadas 3 transformações. Em cada transformação, 4 placas de APG (Agar Peptona Glicose) contendo esporos de *T. reesei* foram bombardeadas. Os transformantes obtidos foram avaliados em sua capacidade de crescimento e desenvolvimento em meio APG suplementado com higromicina B.

A figura 12 mostra as posições dos oligonucleotídeos utilizados para a clonagem dos fragmentos do promotor (figura 12a), e os resultados obtidos após a transformação da cepa QM9414 de *T. reesei* com os plasmídeos construídos (figura 12b, para análises do promotor citrato sintase, e 12c, para análises do promotor isocitrato liase). Considerando que estamos trabalhando com um provável promotor bidirecional e que o vetor construído contém um único gene repórter, os fragmentos de DNA foram clonados em ambas as orientações, gerando 20 construções (tabela 2, plasmídeos 1- 20).

Dos 20 plasmídeos que foram avaliados por transformação, 11 não se desenvolveram na presença do antibiótico [pHygAD (5' \rightarrow 3' TrCit), pHygAC' (5' \rightarrow 3' TrCit), pHygFB (5' \rightarrow 3' TrIcl), pHygBD (5' \rightarrow 3' TrCit), pHygDB (5' \rightarrow 3' TrIcl), pHygBC' (5' \rightarrow 3' TrIcl), pHygC'B (5' \rightarrow 3' TrIcl), pHygFC (5' \rightarrow 3' TrIcl), pHygCD (5' \rightarrow 3' TrIcl), pHygCD (5' \rightarrow 3' TrIcl). Os nove demais produziram transformantes estáveis que germinaram na presença do antibiótico [pHygAF (5' \rightarrow 3' TrCit), pHygFA (5' \rightarrow 3' TrIcl), pHygAE (5' \rightarrow 3' TrCit), pHygBA (5' \rightarrow 3' TrCit), pHygDA (5' \rightarrow 3' TrCit), pHygC'A (5' \rightarrow 3' TrIcl), pHygBF (5' \rightarrow 3' TrCit), pHygCF (5' \rightarrow 3' TrCit) e pHygCE (5' \rightarrow 3' TrCit)]. A medida relativa do crescimento radial de cada transformante estável (figura 12) foi determinada a partir do crescimento desses transformantes em meio APG (glicose 5 mM) / Hyg (100 μ g.mL⁻¹) (figura 13).

Como mostrado na figura 12 (b e c), o fragmento de -1 a -647 (647 pb), contido nos plasmídeos pHygAF (5' \rightarrow 3' TrCit) e pHygFA (5' \rightarrow 3' TrIcl), e o fragmento de -162 a -647 (486 pb), contido nos plasmídeos pHygAE (5' \rightarrow 3' TrCit) e pHygEA (5' \rightarrow 3' TrIcl), dirigem a transcrição de ambos os genes, pela comprovação do crescimento dos transformantes com essas construções em placas APG (glicose 5mM))/ Hyg (100 μ g.mL⁻¹).

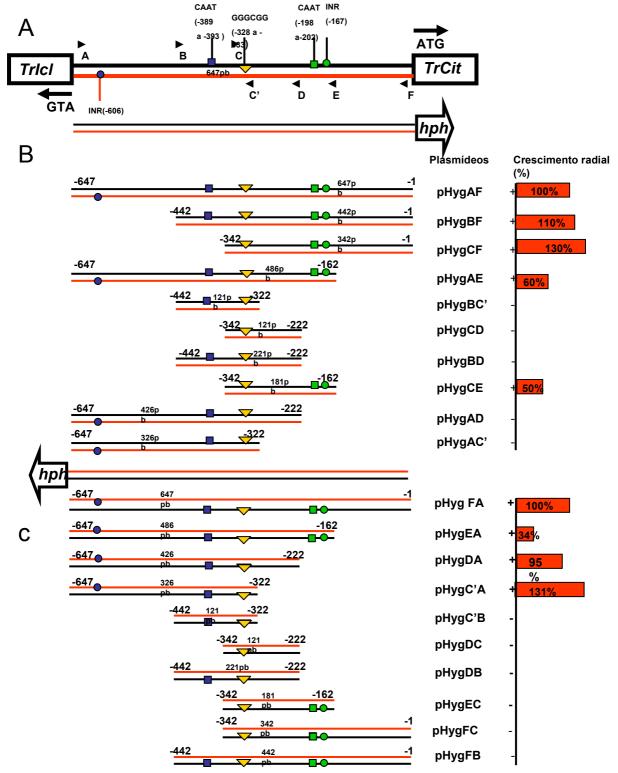


Figura 12. Análise da região intergênica entre os genes citrato sintase (*TrCit*) e isocitrato liase (*TrIcl*) de *Trichoderma reesei*. A) Esquema da organização divergente dos genes *TrIcl* e *TrCit*. As setas acompanhadas de letras indicam a orientação dos "primers" empregados na amplificação de fragmentos. B e C) Análise da atividade transcricional dos fragmentos de DNA amplificados por PCR e clonados no vetor pHyg.na orientação do gene Citrato sintase e Isocitrato liase respectivamente. A posição das bases é indicada por números nas extremidades de cada fragmento. Transformantes obtidos foram avaliados pelo seu crescimento. + = Crescimento, - = Não há crescimento. O = Indica INR; □ =Indica caixa CAAT; ▽ = Indica caixa GC.

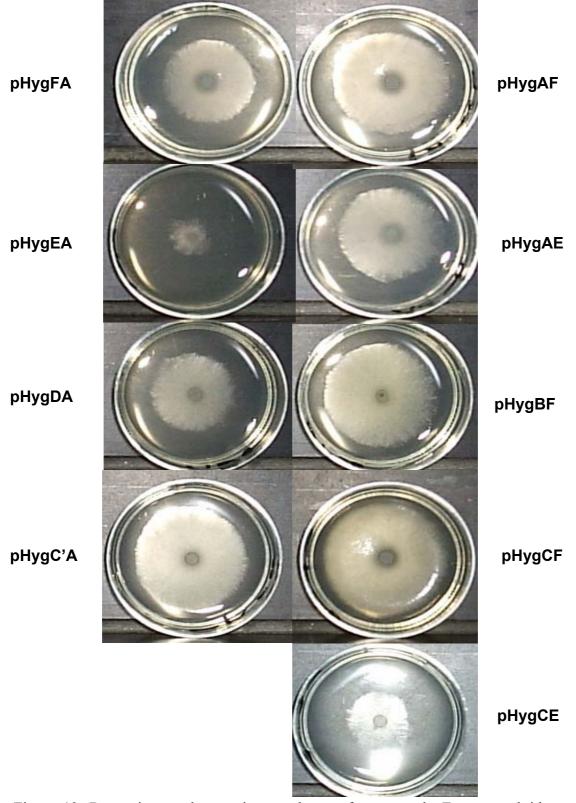


Figura 13. Determinação do crescimento de transformantes de *T. reesei*, obtidos por biobalística. Culturas após 48 h/28 °C de crescimento, em meio APG (glicose 5 mM)/Hyg (100 μ g/mL).

Observa-se também que os transformantes obtidos com plasmídeo levando o fragmento na orientação do gene *TrCit* (pHygAF e pHygAE) apresentaram um crescimento radial maior quando comparados aos transformantes nos quais o fragmento estava na orientação inversa, no sentido de *TrIcl*, (pHygFA e pHygEA) (figura 13). Esses resultados foram confirmados pela análise da expressão gênica por meio de northern blot, empregando RNA total da cepa "selvagem" QM9414 de *T. reesei* crescida em diferentes fontes de carbono; os resultados mostram que o nível de expressão do gene *TrCit* é maior que o do gene *TrIcl* (figura 10).

A análise da figura 12 (b e c) também mostra que o plasmídeo pHygC'A, que contém o fragmento -322 a -647, de 326 pb, dirige a transcrição do gene *TrIcl*, e que o plasmídeo pHygCE, que contém o fragmento -162 a -342, de 181 pb, dirige a transcrição do gene *TrCit*. Ambos estes fragmentos apresentam em comum uma seqüência de 21 pb, localizada entre -328 a -333, onde está presente a seqüência GGGCGG, conhecida como caixa GC.

Para determinar a importância da caixa GC na transcrição de ambos os genes, ela foi deletada em quatro plasmídeos construídos (tabela 2, plasmídeos 21, 22, 23 e 24). Os plasmídeos construídos: pHygdAF ($5' \rightarrow 3'$ TrCit) e pHygdFA ($5' \rightarrow 3'$ TrIcl), que contêm um fragmento de 641 pb, pHygdHF, de 327 pb ($5' \rightarrow 3'$ TrCit), e pHygdGA, de 314 pb ($5' \rightarrow 3'$ TrIcl), não produzem transformantes estáveis que germinem na presença do antibiótico (figura 14).

A região intergênica *TrIcl/TrCit* foi analisada computacionalmente com a finalidade de se identificar sítios de ligação putativos para fatores de transcrição. Foi utilizado o servidor web do programa TESS ("Transcription Element Search System", http://www.cbil.upenn.edu/tess; Overton & Schug, 1997). A análise identificou mais de 100 sítios de ligação putativos para fatores de transcrição, dentre os quais sítios de ligação para o fator de transcrição Sp1, entre outros (figura 15).

Como mostrado anteriormente, a região intergênica *TrIcl/TrCit* é considerada uma região rica em GC e as regiões ricas em GC apresentam sítios de ligação putativos para a proteína Sp1. A análise do "*Trichoderma reesei* EST database and mitochondrial genome" (http://trichoderma.iq.usp.br) permitiu a identificação do clone TrESTA0638, que apresenta 42% de similaridade com a proteína Sp1 humana (NP 612482) e de rato

(NP_036787) (figura 16a). O gene do putativo fator de transcrição "zinc-finger" *TrZnFSp1*, de 1500 pb, contém 3 éxons (de, respectivamente, 478, 140 e 726 pb) e 2 íntrons (de, respectivamente, 66 e 90 pb) (figura 16b). A clonagem e expressão da seqüência codificadora da proteína recombinante TrZnFSp1, de 1.344 pb, produz uma proteína de 448 aminoácidos, com um peso molecular estimado de 48,4 kD (figura 17a).

A clonagem e expressão da putativa proteína Sp1 de *T. reesei* (TrZnFSp1) permitiu a análise funcional da caixa GC, localizada entre -328 a -333 na região intergênica *TrIcl/TrCit*, pelo ensaio de retardamento de mobilidade eletroforética (EMSA). A figura 17b mostra a interação do putativo fator de transcrição TrZnFSp1 com o oligonucleotídeo específico de 21 pares de bases, de seqüência CTGATTTGAGGGCGGTGACGG, que contém a caixa GC (GGGCGG) central da região intergênica *TrIcl/TrCit*.

As reações para o ensaio EMSA foram montadas e sujeitas a eletroforeses em gel de poliacrilamida não desnaturante (Materiais e Métodos), na seguinte ordem: coluna 1 (controle negativo), apenas o oligonucleotídeo específico marcado radioativamente; coluna 2 (controle positivo), proteína recombinante TrZnFSp1 e oligonucleotídeo específico marcado radioativamente; coluna 3 (competidor específico), proteína recombinante TrZnFSp1, competidor específico (oligonucleotídeo específico nãomarcado) e oligonucleotídeo específico marcado radioativamente; coluna 4 (competidor não específico), proteína recombinante TrZnFSp1, competidor não-específico AP2 (oligonucleotídeo não-específico não-marcado; tabela 1, oligo 20) e oligonucleotídeo específico marcado radioativamente. Na análise da figura 17b, observamos uma banda intensa (coluna 2), devido à formação do complexo proteína:DNA entre o oligonucleotídeo contendo a caixa GC e o putativo fator de transcrição TrZnFSp1. Na presença do competidor específico (coluna 3), a intensidade da banda foi reduzida, mas não foi afetada pela presença do competidor não-específico (coluna 4).

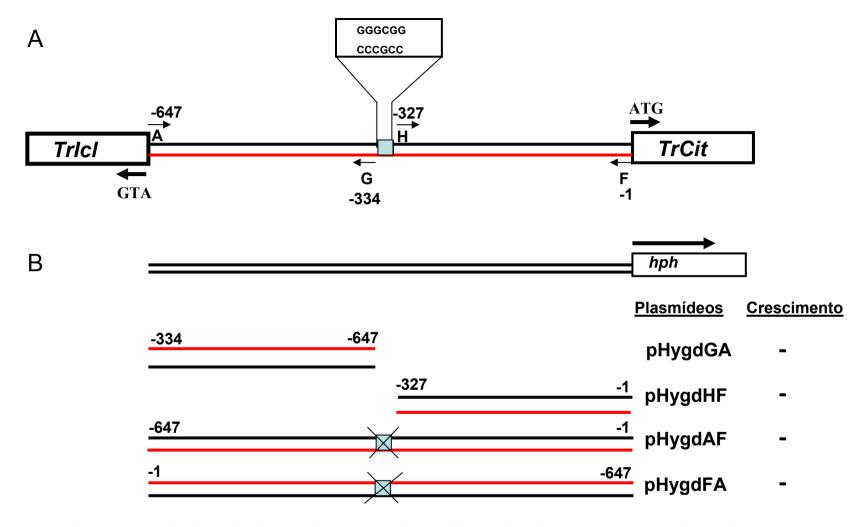


Figura 14. Análise da região intergênica *TrCit/TrIcl* após deleção da caixa GC. A) Esquema da organização dos genes divergentes *TrCit* e *TrIcl* e localização da caixa GC (GGGCGG). As setas acompanhada de letras indicam a orientação dos primers empregados na amplificação dos fragmentos. B) Análise da atividade transcricional dos fragmentos de DNA amplificados e inseridos no vetor pHyg. A posição de cada fragmento é indicado por números nas extremidades. (-) Não foram obtidos transformantes.

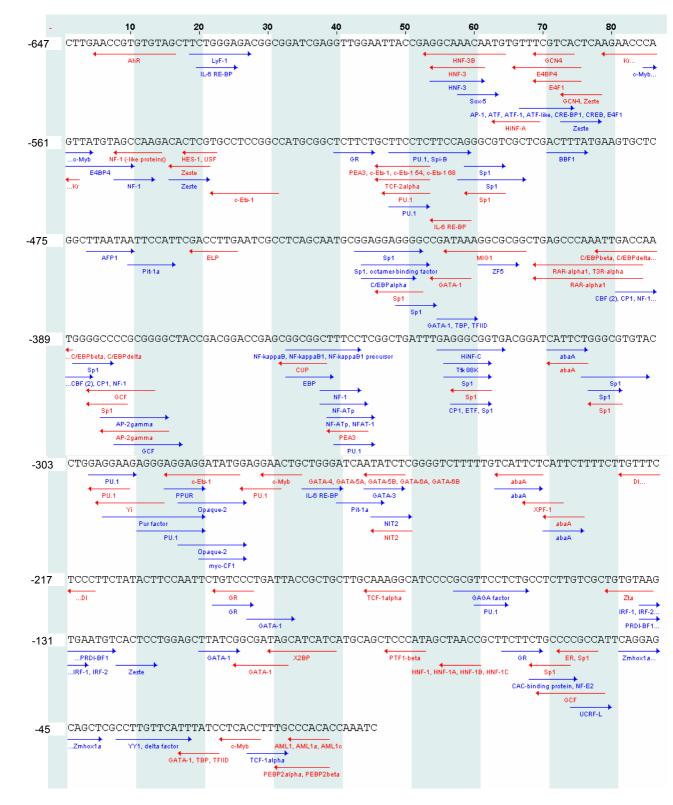


Figura 15. Predição de possíveis sítios de ligação de fatores de transcrição, na região intergênica de 647 pb, entre os genes citrato sintase (*TrCit*) e isocitrato liase (*TrIcI*) de *T. reesei*. A predição foi realizada empregando o servidor do programa TESS: "Transcription Element Search System" (http://www.cbil.upenn.edu/tess) .

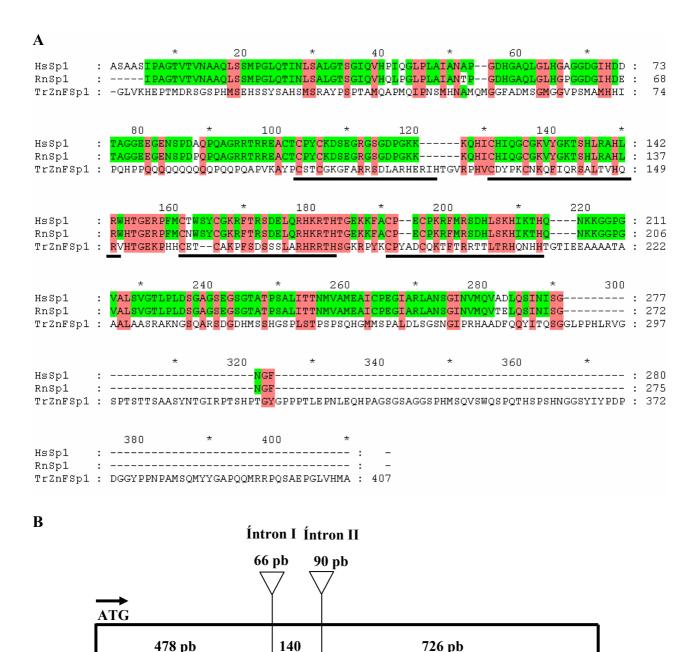


Figura 16. Análise da putativa proteína TrZnFSp1. A) Alinhamento da seqüência de aminoácidos do domínio "zinc-finger" das proteínas HsSp1 (Sp1 de *Homo sapiens*, NP_612482), RnSp1 (Sp1 de *Rattus norvegicus*, NP_036787) e TrZnFSp1. Os aminoácidos que são idênticos ou conservados são marcados num fundo colorido. Números na coluna direita indicam o último aminoácido em cada linha da respectiva seqüência de aminoácidos. Os domínios de ligação ao DNA estão sublinhados, B) Estrutura gênica de TrZnFSp1.

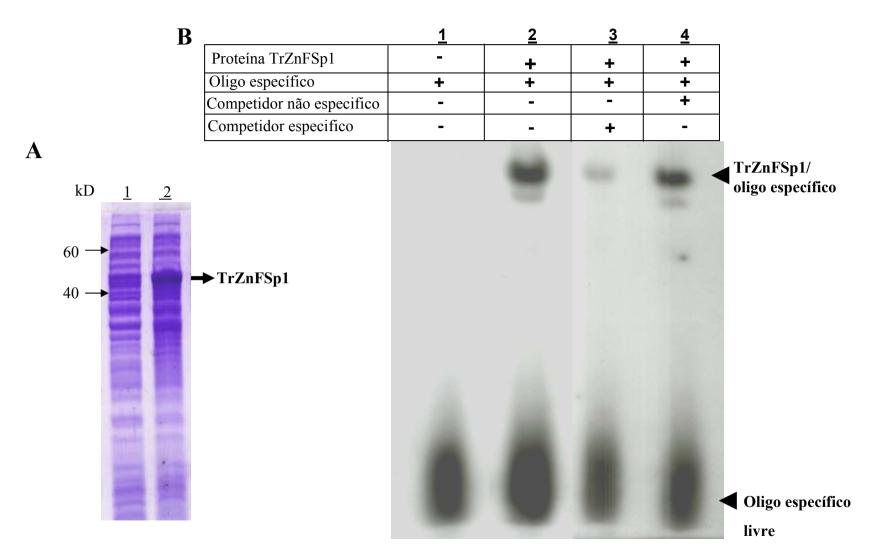


Figura 17. Análise de ensaio de mudança de mobilidade eletroforética. A) Eletroforese em gel de poliacrilamida-SDS 10%, da expressão da proteína recombinante TrZnFSp1, linha 1)10 uL bactéria não induzida; linha 2)10 uL bactéria induzida com IPTG. B) Análise da interação da proteína recombinante TrZnFSp1 ao oligonucleotídeo específico contendo a caixa GC. Eletroforese em gel de poliacrilamida não denaturante 5%, Linha 1) controle negativo, linha 2) controle positivo, linha 3) competidor específico, e linha 4) competidor não específico.

V. Discussão.

O seqüenciamento do genoma de numerosos seres vivos ("Genomes OnLine Database", http://wit.integratedgenomics.com/GOLD/), como o do ser humano (Lander *et al.*, 2001; Venter *et al.*, 2001) e *S. cerevisiae* (Goffeau *et al.*, 1996), marcou o advento da era genômica, e tornou possível a análise e comparação do genoma de diferentes organismos. Os dados obtidos têm resultado na determinação da estrutura e organização dos genomas, identificando e localizando genes e regiões intergênicas (Lawrence *et al.*, 2003).

No presente trabalho, descrevemos a caracterização e análise funcional da região intergênica *TrIcl/TrCit* (647 pb) contendo um promotor bidirecional que regula a transcrição divergente dos genes citrato sintase (*TrCit*) e isocitrato liase (*TrIcl*) do fungo filamentoso *Trichoderma reesei*.

5.1 Genoma e rede de regulação transcricional.

No genoma humano (Takai et al., 2004) foi estabelecido que a proporção entre genes transcritos divergentes e os direcionados no mesmo sentido numa mesma fita (ver figura 1) está em acordo com a suposição de que os genes nos eucariotos estão localizados aleatoriamente no genoma, ou seja, para qualquer gene a orientação de um gene localizado 5' na mesma direção ou na direção oposta é de ~ 50%. Esta forma de organização no genoma, de genes transcritos divergentes, sugere que sua expressão possa ser co-regulada (Trinklein et al., 2004) e seu estudo é de grande interesse, pois pode revelar como a transcrição de diferentes genes é coordenada (Koyanagi et al., 2005). Genes transcritos divergentes têm sido identificados em diferentes organismos, por exemplo, Homo sapiens, Mus musculus, Rattus norvegicus, Fugu rubripes, Ciona intestinalis, Drosophila melanogaster, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Arabidopsis thaliana (Koyanagi et al., 2005), Aspergillus nidulans (Muro-Pastor et al., 1999), Penicillium chrysogenum, Aspergillus nidulans e Penicillium nalgiovense (Martín, 2000), mostrando ser um tipo de organização gênica conservada durante a evolução.

Trichoderma reesei é um fungo filamentoso empregado como organismo modelo para a análise genômica da regulação de redes metabólicas (Chambergo *et al.*, 2002; Bonaccorsi *et al.*, 2006). Este fungo se destaca por sua capacidade de produzir e secretar um sistema multienzimático que hidrolisa eficientemente celulose a glicose (Béguin, 1990). Nosso laboratório seqüenciou o cDNA e o genoma mitocondrial e identificou mais de 10% dos genes expressos em *T. reesei* (Chambergo *et al.*, 2002), que estão disponíveis no website "Trichoderma reesei EST database and Mitochondrial genome" (http://trichoderma.iq.usp.br). Em 2005 foi liberado o "Draft genome *Trichoderma reesei* v 1.0" (http://gsphere.lanl.gov/trire1/trire1.home.html), que identifica um aproximado de 8000 genes putativamente expressos neste organismo.

Durante a análise do "*Trichoderma reesei* EST database and Mitochondrial genome" foi identificado o gene citrato sintase (*TrCit*). A análise da região 5′ flanqueadora "upstream" de *TrCit* apontou a presença de um gene, transcrito de forma divergente, que codificaria uma putativa proteína isocitrato liase (*TrIcl*) (ver figura 5). Ambos os genes seriam transcritos a partir de uma região intergênica de 647 pb.

A sequência de 647 pb, embora não apresente consenso TATAAA ("TATA less promoter"), apresenta sítios INR, seu conteúdo G + C é de 54% (região rica em GC) e contém uma ilha CpG (CpG observado / CpG esperado = 0.98). Essas características são comuns em promotores de genes constitutivos (Seki *et al.*, 2002). Uma análise computacional recente estabeleceu que apenas um aproximado de 30% a 50% dos promotores humanos de RNA pol II apresentam caixa TATA (Gross *et al.*, 2006). Alguns genes transcritos a partir de promotores sem caixa TATA têm sido identificados (Seki et al., 2002; Badran *et al.*, 2005).

A transcrição é um processo regulado pela interação concertada de fatores gerais de transcrição, ativadores, ou repressores, que se ligam em seqüências específicas, e proteínas acessórias que modulam a acessibilidade à cromatina. A proteína TBP ("TATA box binding protein") é um fator geral de transcrição que se liga a promotores de RNA pol II que contenham, ou não, a seqüência consenso TATA, indicando que o processo de transcrição é dependente de TBP. Assim, outros componentes da maquinaria de transcrição, associados a TBP, devem promover a ligação de TBP aos promotores sem

caixa TATA, onde reconhecem uma estrutura diferente da caixa TATA, que participa na montagem do complexo de pré-iniciação (PIC) (Kim *et al.*, 2004).

Seqüências ricas em G + C apresentam grupos de dinucleotídeos CpGs, ou "ilhas CpG", presentes nos promotores de ~40% dos genes de mamíferos (Takai *et al.*, 2002). Ilhas CpG são definidas como regiões de DNA de 200 pb com um alto conteúdo G + C (> 50%) e uma relação CpG observado / CpG esperado (Obs_{CpG}/Exp_{CpG}) maior, ou igual, a 0,6 (Takai *et al.*, 2002; Wang *et al.*, 2004). A presença de ilhas CpG nas regiões 5'de alguns genes está associada ao efeito da metilação sobre a expressão gênica. Ilhas CpG não metiladas podem ser necessárias, porém não suficientes, para a transcrição. Assim, a presença de ilhas CpG não metiladas na maioria dos genes constitutivos ("housekeeping genes") pode ser considerada como uma indicação de que o gene está potencialmente ativo (Wang *et al.*, 2004).

A transcrição divergente de *TrCit* e *TrIcl* a partir da região intergênica de 647 pb sugere que o promotor requerido para expressão de ambos os genes seja compartilhado. O programa TESS identificou mais de 100 sítios de ligação putativos para diferentes fatores de transcrição, entre eles, 6 caixas GC e 1 caixa CAAT.

A identificação computacional de sítios de ligação de fatores de transcrição e a análise de deleção são as primeiras etapas no estudo de uma região promotora. Diversos programas (MatInd/MatInspector/FastM http://www.gsf.de/biodv/matinspector.html; MATRIX SEARCH 1.0 chenq@boulder.colorado.edu; PatSearch 1.1 http://transfac.gbfbraunschweig.de/cgi-bin/patSearch/patsearch.pl; Signal Scan http://bimas.dcrt.nih.gov/ **TESS** http://www.cbil.upenn.edu/tess/; **TFSEARCH** molbio/signal/; http://pdap1.trc.rwcp. or.jp/research/db/FSEARCH.html, entre outros) disponíveis na internet realizam a predição de sítios de ligação putativos em sequências de DNA por meio da análise comparativa da sequência dada com sequências consenso predefinidas para sítios de ligação de proteínas. Entretanto as matrizes utilizadas na análise correspondem a fatores de transcrição identificados em procariontes, vírus, vertebrados, nematódeos, artrópodes, plantas e levedura, faltando um programa que faça a análise utilizando matrizes feitas a partir de dados obtidos de fatores de transcrição de fungos filamentosos

Para determinar a região de DNA essencial para a expressão dos dois genes, foi adotada uma análise de deleção com uma série de plasmídeos construídos na qual fragmentos de DNA da região intergênica *TrIcl/TrCit* foram clonados "upstream" ao gene repórter *hph*, no vetor pHyg.

Na análise de deleção, fragmentos de DNA de diferentes tamanhos são clonados "upstream" de um gene repórter e a atividade promotora é determinada a partir da transcrição do gene repórter. Empregando esta metodologia, foi possível determinar a importância da caixa CCAAT e dos sítios de ligação para PacC (GCCARG) e CreA (SYGGRG) presentes na região promotora que regula a transcrição divergente dos genes pcbAB e pcbC, da via de biossíntese de penicilina, nos fungos filamentosos Penicillium chrysogenum, P. nalgiovense e Aspergillus nidulans, utilizando como gene repórter o gene lacZ (Martín, 2000). Também foi estabelecido que um fragmento de 165 pb é suficiente para regular a transcrição bidirecional do gene diidrofolato redutase (DHFR) e do gene "upstream" transcrito em sentido divergente empregando como repórter o gene cloranfenicol acetiltransferase (Shimada et al., 1989).

Em nosso trabalho (figura 12 e figura 14) é demonstrada, a importância das caixas GC, caixas CAAT e sítios INR na expressão do gene repórter. Os fragmentos de 647 pb (de -1 a -647) e de 486 pb (de -161 a -647), que contêm as caixas GC, caixas CAAT e sítios INR, foram clonados no plasmídeo pHyg em ambas as orientações (5'\rightarrow3' e 3'\rightarrow 5') e, dirigem a expressão do gene hph, permitindo o crescimento do fungo transformado na presença do antibiótico higromicina B. Na análise da seqüência na orientação 5'→3' no sentido do gene TrCit, quando o fragmento de -1 a -221 (221 pb) que contém o sítio INR (localizado em -167) e a caixa CAAT (localizada -198 a -202) é deletado, não se observa crescimento de transformantes. Porém, os fragmentos dos plasmídeos pHygBF, pHygCF e pHygCE, que contêm o sítio INR (-167), caixa CAAT (-198 a -202) e caixa GC (-328 a -333), permitem o crescimento do fungo transformado na presença do antibiótico higromicina B, indicando que para a expressão na orientação do gene TrCit não são importantes a caixa CAAT localizada entre -389 a -393 e o sitio INR localizado em -606, como mostrado pela análise do plasmídeo pHygCE. Esse resultado mostra que a sequência mínima entre -162 e -342 contém os elementos promotores necessários para a transcrição do gene na orientação TrCit.

Na análise da seqüência na orientação 5'→3' no sentido do gene *TrIcl*, quando o fragmento de -443 a -647 (205 pb) que contém o sitio INR (localizado em - 606) é deletado, não se observa crescimento de transformantes. Porém, os fragmentos dos plasmídeos pHygDA e pHygC'A, que contêm caixa CAAT (-389 a -393), caixa GC (-328 a -333) e INR (-606), permitem o crescimento do fungo transformado na presença do antibiótico higromicina B, assinalando que para a expressão na orientação do gene *TrIcl* não são importantes a caixa CAAT localizada entre -198 a -202 e o sitio INR localizado em -167, como mostrado pela análise do plasmídeo pHygC'A. Esse resultado mostra que a seqüência mínima entre -322 a -647 contém os elementos promotores necessários para a transcrição na orientação do gene *TrIcl*.

Os fragmentos inseridos nos plasmídeos pHygCE e pHygC'A, que contêm os elementos promotores mínimos necessários para a transcrição do gene repórter, têm em comum uma sequência de 21 pb, localizada entre -322 a -342, onde está presente a sequência GGGCGG, ou caixa GC. A deleção das 5 bases nucleotídicas da caixa GC (-328 a -333), localizada no meio da região intergênica, não produz transformantes estáveis que germinem na presença do antibiótico (figura 14). Esse resultado mostra o papel essencial que esta caixa GC tem na transcrição gênica divergente dos genes *TrCit* e *TrIcI* a partir de sua região intergênica.

Uma região promotora sem caixa TATA, rica em G + C e que apresenta ilha CpG inclui sítios de ligação para o fator de transcrição Sp1 (consenso GGGCGG, ou caixa GC) (Oei *et al.*, 2004), que recruta o fator geral de transcrição TFIID ativando o promotor (Kolell *et al.*, 2002). Em *T. reesei*, foi identificada uma proteína com similaridade a proteína "zinc-finger" Sp1 (*TrZnFSp1*). A figura 17 mostra a interação da proteína recombinante *TrZnFSp1* com a seqüência do oligonucleotídeo de 21 pb (-322 a -342), através da análise EMSA, compartilhado pelos plasmídeos pHygCE e pHygC'A (ver figura 12), que contém a caixa GC (-328 a -333), confirmando a importância e funcionalidade desta região para ligar fatores de transcrição necessários para a expressão dos genes. Sp1 é um fator de transcrição ubíquo, expresso em diversas células de mamíferos, que liga em caixas GC e está relacionado à manutenção das ilhas CpG livre de metilação e interage com TBP para estabilizar o complexo de pré-iniciação (Torigoe *et al.*, 2003).

A análise EMSA é baseada na formação de um complexo proteína:DNA, que migra mais lentamente que as moléculas livres do DNA alvo específico, em eletroforese em gel de poliacrilamida não desnaturante. A interação específica proteína:DNA é comprovada através da análise de três reações:

- a proteína interage com a sequência específica do DNA alvo marcado radioativamente, sendo o complexo proteína:DNA de mobilidade menor que a observada para o DNA alvo livre, observado como um sinal intenso em filme radiográfico;
- 2) a proteína interage com uma mistura de DNA alvo específico marcado e nãomarcado radioativamente (competidor específico), resultando na observação de um sinal de mesma mobilidade que o anterior, mas intensidade menor, devido à interação com o DNA não-marcado; e
- 3) a proteína interage com uma mistura de DNA alvo específico marcado radioativamente e DNA alvo não-específico não-marcado radioativamente (competidor não-específico), que leva à observação de um sinal de iguais mobilidade e intensidade às da primeira reação no filme radiográfico.

Os experimentos de EMSA com a proteína recombinante TrZnFSp1 e o fragmento de 21 pb contendo a caixa GC (Figura 17) satisfizeram as três condições indicadas acima. A intensidade da banda formada pela proteína recombinante e o fragmento de 21 pb contendo a caixa GC foi fortemente reduzida pela presença do competidor específico; por outro lado, o competidor não específico não afetou a intensidade da banda. Os resultados obtidos, portanto, indicam uma interação específica entre a proteína TrZnFSp1 e o oligonucleotídeo de 21 pb da região central do promotor bidirecional da região intergênica *TrIcl/TrCit*.

A localização de uma caixa GC funcional, na região central de um promotor, que seria compartilhada para regular a expressão de ambos os genes é uma característica comum de promotores compartilhados. Johnston *et al.* (1984) mostraram que uma região de 108 pb, rica em GC, e localizada no meio da região entre os genes *GAL1* e *GAL10* é necessária para a expressão de ambos os genes. Os resultados obtidos mostram que a região intergênica *TrIcl/TrCit* rica em GC, que contém uma ilha CpG, caixas GC, caixas CAAT, seqüência INR e sem caixa TATA, age como um promotor bidirecional (figura 3b), que controla a expressão dos genes *TrCit* e *TrIcl*, dispostos cabeça-cabeca (figura 1b) no genoma do fungo filamentoso *Trichoderma reesei*.

Os dados obtidos são de grande importância, pois indicam que o genoma de T. reesei, um fungo filamentoso, apresenta uma organização e talvez regulação transcricional similar a outros organismos superiores. Os dados obtidos a partir da análise do genoma de organismos localizados em diferentes níveis taxonômicos e evolutivos têm contribuído para determinar a distribuição de genes entre grupos taxonômicos e a organização dos genes nos genomas (Lawrence, 2003), e analisar a estratégia combinatorial de regulação gênica necessária para gerar os diversos padrões de expressão, devido a apenas um número limitado de fatores de transcrição poder ser codificado pelo genoma (Smale, 2001). Assim, o genoma humano com um número aproximado de 30 a 40 mil genes tem mostrado como um limitado número de genes é capaz de desenvolver um complexo sistema de vida. Tal feito talvez seja possível devido à participação de um sistema de regulação transcricional mais complexo e preciso. Ainda não se sabe como esse sistema de regulação evoluiu, mas, acredita-se que a forma de organização do genoma pode refletir o sistema de regulação transcricional dos genes. Assim, o conhecimento da evolução da organização do genoma pode revelar a evolução e os mecanismos de regulação transcricional (Koyanagi et al., 2005).

O genoma de procariotos, na maioria dos casos, é altamente organizado em unidades transcricionais denominadas operons nos quais genes são co-expressos a partir dum simples promotor e que, de forma geral, contêm genes funcionalmente relacionados, implicando numa co-regulação. Em eucariotos, os genes são transcritos individualmente e estão arranjados em cromossomos sem aparente organização por função ou padrão de expressão (Blumenthal, 1998; Koyanagi *et al.*, 2005). Entretanto alguns grupos de genes

duplicados, relacionados funcionalmente, ou são co-regulados como, por exemplo, os genes humanos da hemoglobina e imunoglobulinas (Blumenthal, 1998) ou podem formar pares de genes divergentes (localizados cabeça-cabeça) co-expressos a partir de uma região promotora como, por exemplo, um grande número de genes constitutivos no genoma humano (Adachi *et al.*, 2002; Takai *et al.*, 2004; Trinklein *et al.*, 2004; Koyanagi *et al.*, 2005). Esta última forma de organização encontrada em diversos organismos é de grande interesse pois pode mostrar como a transcrição de diferentes genes é coordenada.

Células eucarióticas empregam uma grande rede de fatores que regulam a transcrição em resposta a sinais extracelulares ou de desenvolvimento. Para iniciar a transcrição de um gene específico, uma série de eventos deve ocorrer de forma concertada: abertura da cromatina empacotada, ligação de fatores ativadores de transcrição a seqüências especificas no DNA do gene e interação coordenada entre a maquinaria de transcrição e a RNA pol II. O início da transcrição requer a montagem ordenada dos fatores gerais de transcrição e a RNA pol II num complexo de pré-iniciação, montado no centro do promotor. O centro do promotor é funcionalmente definido como a mínima região de DNA necessária para dirigir *in vitro* um baixo nível de inicio de transcrição na ausência de ativadores. Nesta região podem ser encontrados elementos de seqüência consenso conhecida, como a caixa TATA (TATAAA), elemento BRE ("TFIIB recognition element", G/C G/C G/A CGCC), sítio de inicio de transcrição INR (YYAN T/A YY), elemento DPE ("downstream promoter element", A/G G A/T C/T G/A/C) e elemento MTE ("motif ten element", AACGGAACGG) (Gross *et al.*, 2006).

A análise da prevalência destes elementos nos promotores em genomas de diferentes organismos apontou o sítio consenso TATA em 33% de 1.941 promotores de Drosophila e em 32% de 1.000 promotores de humano enquanto 80% dos promotores, de ambos os organismos apresentam seqüência consenso INR. Análise de promotores humanos depositados na base de dados de promotores eucarióticos (EPD, "Eukaryotic Promoter database") e na base de dados de sítios de início de transcrição (DBTSS, "Database of Transcriptional Start Sites"), mostra os seguintes dados sobre os promotores: 32% contêm caixa TATA, (60% com caixa TATA e sítio INR; 25% com caixa TATA e sitio consenso BRE), dos promotores sem caixa TATA: 45% sem caixa TATA e com sítio INR e 55% sem caixa TATA e com sítio BRE e aproximadamente

25% dos promotores humanos não apresenta elementos conhecidos no promotor. Estes dados mostram a ocorrência combinatorial de elementos na região do promotor que, unida à presença de ilhas CpG e regiões ricas em GC, e a organização e estrutura do genoma, demonstra e justifica a presença de um complexo *sistema* de regulação transcricional (Smale, 2001; Takai *et al.*, 2003; Wang *et al.*, 2004; Gross *et al.*, 2006). Aproximadamente metade dos genes de *S. cerevisiae* apresenta promotores numa região intergênica "upstream" compartilhada com outros genes transcritos em sentido divergente (Kim *et al.*, 2004). Além disso, muitos promotores são ligados por vários reguladores de transcrição, por meio de uma rede de reguladores de transcrição que controla outros reguladores de transcrição, de uma forma altamente coordenada sugerindo que a regulação gênica em eucariotos é regulada pela interação combinada de uma rede de reguladores (Lee *et al.*, 2002). Em *S. cerevisiae*, um eucarioto unicelular, um aproximado de mais de 200 proteínas regulam a transcrição de aproximadamente 6.200 genes.

5.2 Organização gênica e metabolismo.

A análise estrutural e funcional da região intergênica *TrIcl/TrCit*, mostra que ela contém os elementos promotores necessários para dirigir a transcrição divergente de ambos os genes. A sequência codificadora de *TrCit*, interrompida por 2 íntrons, de 73 e 78 pb, codifica uma proteína de 474 aminoácidos, com peso molecular previsto de 52,3 kD. A similaridade da proteína TrCIT é especialmente alta com as proteínas citrato sintase (Cit1p, Cit2p e Cit3p) de *S. cerevisiae* e do fungo filamentoso *P. anserina* – 80% similaridade em um alinhamento completo das proteínas de acordo com o programa FASTA.

O alinhamento múltiplo de TrCIT com outras proteínas citrato sintase de microorganismos eucarióticos representativos revelou a presença de motivos conservados e característicos de proteínas citrato sintase.

A sequência codificadora de *TrIcl*, interrompida por 2 íntrons, de 30 e 62 pb, codifica uma proteína de 596 aminoácidos, com peso molecular previsto de 65,4 kD. A similaridade da proteína TrICL é especialmente alta com as proteínas metilisocitrato liase do fungo filamentoso *Aspergillus nidulans* e de *S. cerevisiae*, 83,8% e 80%,

respectivamente, em um alinhamento completo das proteínas de acordo com o programa FASTA.

O alinhamento múltiplo de TrICL com outras proteínas isocitrato liase de microorganismos eucarióticos representativos revelou a presença de motivos conservados e característicos de proteínas isocitrato liase.

A análise comparativa do perfil de expressão entre os transcritos divergentes TrCit e TrIcl aponta para um padrão de expressão assimétrico (ver figura 10), no qual, um dos transcritos divergentes é mais expresso que o outro. As vias metabólicas nas quais os genes TrCit e TrIcl estariam envolvidos seriam o ciclo de TCA (ciclo de Krebs ou do ácido cítrico) e o ciclo do glioxilato, respectivamente. Como mostrado na figura 18, existe uma relação entre ambos ciclos o que justificaria a transcrição divergente coregulada de ambos os genes.

A organização de pares de genes transcritos de forma divergente parece ser um arranjo comum dos genes constitutivos (Adachi *et al.*, 2002), ainda que não exista necessariamente uma relação de função entre eles (Koyanagi *et al.*, 2005). Os genes humanos da APEX nuclease (envolvida na reparação de DNA) e da OSGEP (uma glicoprotease) são transcritos de forma divergente a partir de um promotor bidirecional sem caixa TATA, que apresenta ilha CpG, uma caixa CCAAT e sítios de ligação para os fatores de transcrição UBS e Sp1, sendo que ambos os genes apresentam igual padrão de expressão em diferentes tecidos (Seki *et al.*, 2002). Por outro lado em camundongo e no homem os genes ACACA, envolvido no metabolismo de ácidos graxos, e TADA2L, componente de um complexo que modifica cromatina, são transcritos de forma divergente a partir de um promotor bidirecional sem caixa TATA, que apresenta ilha CpG, uma caixa E ("E-box") e sítios de ligação para os fatores de transcrição AP-2 e Sp1, mas são transcritos de forma assimétrica em diferentes tecidos (Travers *et al.*, 2005).

Assim como a evolução produziu diversos mecanismos regulatórios, as diversas vias metabólicas presentes nos seres vivos refletem a adaptação dos organismos a distintas condições ambientais e nutricionais. A análise parcial do genoma de *T. reesei* (http://gsphere.lanl.gov/trire1/trire1. home.html) identifica uma série de vias envolvidas no metabolismo de diversos substratos

Em organismos aeróbios, como *T. reesei* (Chambergo *et al.*, 2002), o ciclo de Krebs (ou dos ácidos tricarboxílicos, TCA), é uma via central do metabolismo que proporciona equivalentes redutores a cadeia respiratória, através da descarboxilação oxidativa de acetil-CoA, e também participa nos processos de síntese de heme, aminoácidos e glicose. A assimilação de moléculas de dois carbonos, como etanol e acetato, através do ciclo de glioxilato para a síntese de glicose depende da atividade do ciclo de TCA. Assim, a disposição cabeça-cabeça e padrão de expressão do par de genes divergentes *TrCit* e *TrIcl* parece ter uma relação funcional, uma vez que o funcionamento normal do ciclo de TCA é acompanhado por um nível constitutivo de funcionamento do ciclo de glioxilato.

No arranjo divergente dos genes, existe um curto espaço para ligação da RNA pol II, a partir do qual deve ocorrer a transcrição em ambas as orientações, o que motiva a interação de várias moléculas de RNA pol II no mesmo sítio do DNA estabelecendo uma competição entre os promotores compartilhados, afetando a frequência de início de transcrição de um ou de ambos os promotores. Essa forma de interferência transcricional produz uma expressão gênica assimétrica, na qual um promotor forte (agressivo) reduz a expressão de um promotor mais fraco (sensível). Por outro lado, a ligação de uma proteína reguladora pode ativar / reprimir o inicio de transcrição em ambas as orientações ou ativar a transcrição numa direção enquanto reprime a expressão na direção oposta (Beck *et al.*, 1988; Callen *et al.*, 2004).

T. reesei, assim como outros organismos que se desenvolvem no solo, devem adaptar seu metabolismo como um mecanismo de resposta celular a diferentes condições ambientais. Os fungos degradam compostos de 6 carbonos através da via glicolítica e compostos de dois carbonos através da via de glioxilato. O ciclo do ácido cítrico (ou dos ácidos tricarboxílicos, TCA) é a via central do metabolismo, com vias degradativas chegando até ele e com vias anabólicas se principiando nele. Os fungos filamentosos habitam o solo, onde compostos de 2 carbonos (acetato) são abundantes devido a decomposição de vegetais, degradação de aminoácidos e fermentação de carboidratos. Portanto, os fungos filamentosos e outros organismos devem ter a habilidade metabólica para crescer e utilizar estes compostos como fonte de carbono e energia, por meio da regulação e da inter-relação entre suas diversas vias metabólicas (figura 18).

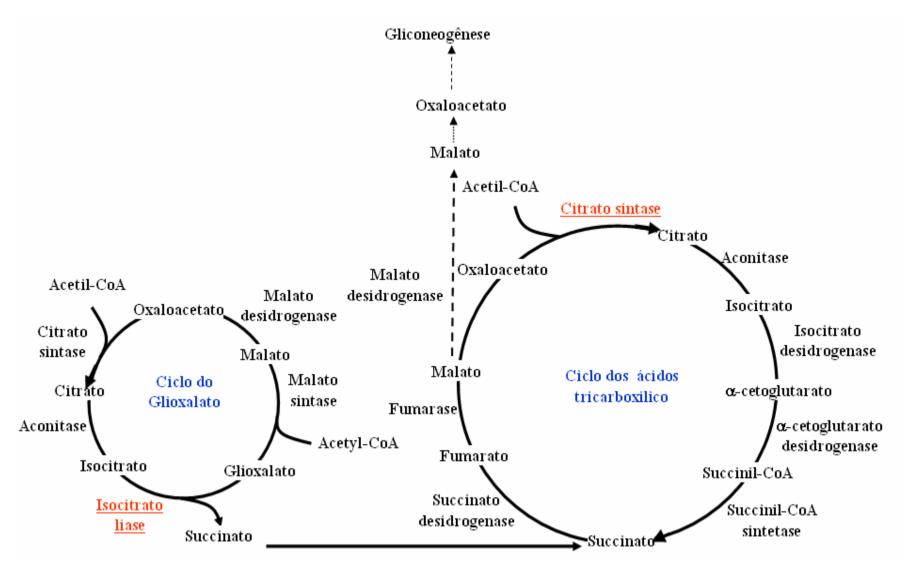


Fig 18.Relações entre os ciclos do glioxalato e do ácido cítrico. As reações do ciclo do glioxalato misturamse com aquelas do ciclo do ácido cítrico, a medida que os intermediários passam entre estes ciclos.

No metabolismo de compostos de 2 carbonos (acetato ou etanol), através da via de glioxilato, a enzima isocitrato liase cliva isocitrato formando glioxilato e succinato, sendo o succinato metabolizado por enzimas do ciclo de TCA. Em *S. cerevisiae*, as enzimas do ciclo de TCA e glioxilato são reprimidas em meio com alta concentração de glicose (DeRisi *et al.*, 1997), enquanto que em *T. reesei* as enzimas do ciclo de TCA estão ativas, ou são parcialmente reprimidas, em meio com alta concentração de glicose (Chambergo *et al.*, 2002). A análise dos dados justifica a organização e co-expressão dos genes *TrCit* e *TrIcI* de *T. reesei*, controlados por um promotor bidirecional, envolvidos no ciclo de TCA e glioxilato, respectivamente.

VI. Conclusões

Neste trabalho, descrevemos: i) a caracterização estrutural dos genes *TrCit* e *TrIcl*, envolvidos em vias distintas do metabolismo do fungo filamentoso *T. reesei*; ii) a estrutura e organização da região intergênica *TrIcl/TrCit*; iii) o padrão transcricional destes genes em diferentes fontes de carbono; e iv) o possível envolvimento de um putativo fator Sp1 na co-expressão dos dois genes.

A proteína TrCIT apresenta alta similaridade com proteínas citrato sintase, enquanto a proteína TrICL apresenta alta similaridade com proteínas isocitrato liase. Também são apresentadas a estrutura e a organização do gene TrZnFSp1, que codifica uma proteína com similaridade ao fator de transcrição Sp1.

Os genes *TrCit* e *TrIcl* estão organizados numa orientação cabeça-cabeça, sendo separados por uma região promotora intergênica (*TrIcl/TrCit*) de 647 pb, rica em G + C, que apresenta uma ilha CpG, seqüência INR, caixas GC, caixas CAAT e sítios de ligação para diversos fatores de transcrição; não foi identificada uma caixa TATA nesta região intergênica. Por meio de uma análise de deleção na região promotora, demonstrou-se que uma caixa GC, localizada na região central do promotor, é necessária para a expressão gênica divergente. Utilizando-se a técnica de retardamento da mobilidade eletroforética (EMSA), mostrou-se que o putativo fator de transcrição TrZnFSp1 se liga especificamente a um fragmento de 21 pb contendo esta caixa GC.

Os resultados mostraram que os genes *TrCit* e *TrIcl* são transcritos de forma divergente, a partir de um promotor bidirecional de 647 pb; além disso, os resultados indicaram que uma caixa GC, localizada no centro do promotor, é essencial para a expressão divergente de ambos os genes. Os transcritos de *TrCit* e *TrIcl* estariam envolvidos, respectivamente, nas vias metabólicas do ciclo de TCA e glioxalato.

Finalmente, este trabalho irá contribuir para uma maior compreensão da organização e estrutura do genoma e, em particular, de promotores bidirecionais, assim como dos mecanismos metabólicos e de regulação transcricional de fungos filamentosos em geral.

Perspectivas

Na continuação deste trabalho, pretendemos determinar as bases responsáveis pela ligação do fator de transcrição TrZnFSp1 à caixa GC. Ensaios de EMSA estão sendo realizados com substituições de bases específicas no fragmento de 21 pb que contém a caixa GC.

A atividade do promotor bidirecional dos genes *TrCit* e *TrIcl* será analisada, em termos quantitativos, com o uso dos genes repórteres luciferase (Firely e Renilla Luciferase) e Gus (Beta -glucoronidase). A construção dos plasmídeos necessários para a realização deste trabalho se encontra em andamento.

Finalmente, pretendemos deletar as duas caixas INR (localizadas -167, na fita senso, e a -606, na fita anti-senso) e as caixas CAAT para a análise da correlação entre elas e do papel desempenhado por cada uma na expressão dos genes *TrCit* e *TrIcl*.

Esperamos que os resultados descritos nesta Dissertação de Mestrado, juntamente com os que pretendemos obter na continuação deste trabalho, possam descrever, de modo adequado e detalhado, a organização e o funcionamento do promotor bidirecional dos genes *TrCIcI/TrCit* e oferecendo uma contribuição relevante para o estudo do genoma e dos mecanismos de controle transcricional em fungos filamentosos.

VIII. Referências

- 1. Adachi, N. & Lieber, M. R. (2002) Bidirectional gene organization: a common architectural feature of the human genome, Cell. 109, 807-9.
- 2. Alam, J. & Cook, J. L. (1990) Reporter genes: application to the study of mammalian gene transcription, Anal Biochem. 188, 245-54.
- 3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool, J Mol Biol. 215, 403-10.
- 4. Ausbel, F. M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., & Smith, J.S.K. (1992) Short protocols in Molecular Biology, 2da edn.
- Badran, B. M., Kunstman, K., Stanton, J., Moschitta, M., Zerghe, A., Akl, H., Burny, A., Wolinsky, S. M. & Willard-Gallo, K. E. (2005) Transcriptional regulation of the human CD3 gamma gene: the TATA-less CD3 gamma promoter functions via an initiator and contiguous Sp-binding elements, J Immunol. 174, 6238-49.
- 6. Beck, C. F. & Warren, R. A. (1988) Divergent promoters, a common form of gene organization, Microbiol Rev. 52, 318-26.
- 7. Beguin, P. (1990) Molecular biology of cellulose degradation, Annu Rev Microbiol. 44, 219-48.
- 8. Blumenthal, T. (1998) Gene clusters and polycistronic transcription in eukaryotes, Bioessays. 20, 480-7.
- 9. Blumenthal, T. (2004) Operons in eukaryotes, Brief Funct Genomic Proteomic. 3, 199-211.
- Bonaccorsi, E. D., Ferreira, A. J., Chambergo, F. S., Ramos, A. S., Mantovani, M. C., Farah, J. P., Sorio, C. S., Gombert, A. K., Tonso, A. & El-Dorry, H. (2006) Transcriptional response of the obligatory aerobe Trichoderma reesei to hypoxia and transient anoxia: implications for energy production and survival in the absence of oxygen, Biochemistry. 45, 3912-24.
- 11. Callen, B. P., Shearwin, K. E. & Egan, J. B. (2004) Transcriptional interference between convergent promoters caused by elongation over the promoter, Mol Cell. 14, 647-56.
- 12. Chambergo, F. S., Bonaccorsi, E. D., Ferreira, A. J., Ramos, A. S., Ferreira, J. R. J. J. R., Abrahao-Neto, J., Farah, J. P. & El-Dorry, H. (2002) Elucidation of the metabolic fate of glucose in the filamentous fungus Trichoderma reesei using expressed sequence tag (EST) analysis and cDNA microarrays, J Biol Chem. 277, 13983-8.

- 13. Clark, F. & Thanaraj, T. A. (2002) Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human, Hum Mol Genet. 11, 451-64.
- 14. Craig, J. M. & Bickmore, W. A. (1994) The distribution of CpG islands in mammalian chromosomes, Nat Genet. 7, 376-82.
- 15. DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale, Science. 278, 680-6.
- El-Gogary, S., Leite, A., Crivellaro, O., Eveleigh, D. E. & el-Dorry, H. (1989) Mechanism by which cellulose triggers cellobiohydrolase I gene expression in Trichoderma reesei, Proc Natl Acad Sci U S A. 86, 6138-41.
- 17. Eveleigh, D. E. (1985) Trichoderma in Biology of Industrial Microorganism (Demain, A. L. a. S., N.A., ed), Benjamin/Cummings, Menlo Park, CA.
- Freitag, M. & Sachs, M. 1995 A simple dot blot assay to measure hygromycin B phosphotransferase activity in whole cell extracts of Neurospora crassa. Fungal Genetics Newsletter 42:26-28
- 19. Gams, W. & Bissett, J. (1998) Morphology and identification of Trichoderma. In *Trichoderma & Gliocladium* (Christian P Kubicek & Gary E Harman, eds), pp. 3-34.
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S. G. (1996) Life with 6000 genes, Science. 274, 546, 563-7.
- 21. Gordon, D., Abajian, C. & Green, P. (1998) Consed: a graphical tool for sequence finishing, Genome Res. 8, 195-202.
- 22. Gross, P. & Oelgeschlager, T. (2006) Core promoter-selective RNA polymerase II transcription, Biochem Soc Symp, 225-36.
- 23. Hansen, J. J., Bross, P., Westergaard, M., Nielsen, M. N., Eiberg, H., Borglum, A. J., Mogensen, J., Kristiansen, K., Bolund, L. & Gregersen, N. (2003) Genomic structure of the human mitochondrial chaperonin genes: HSP60 and HSP10 are localised head to head on chromosome 2 separated by a bidirectional promoter, Hum Genet. 112, 436.

- 24. Haurie, V., Perrot, M., Mini, T., Jeno, P., Sagliocco, F. & Boucherie, H. (2001) The transcriptional activator Cat8p provides a major contribution to the reprogramming of carbon metabolism during the diauxic shift in Saccharomyces cerevisiae, J Biol Chem. 276, 76-85.
- 25. Henrique-Silva, F., el-Gogary, S., Carle-Urioste, J. C., Matheucci, E., Jr., Crivellaro, O. & El-Dorry, H. (1996) Two regulatory regions controlling basal and cellulose-induced expression of the gene encoding cellobiohydrolase I of Trichoderma reesei are adjacent to its TATA box, Biochem Biophys Res Commun. 228, 229-37.
- 26. Herrera-Estrella, A., Goldman, G. H. & Van Montagu, M. (1990) High-efficiency transformation system for the biocontrol agents, Trichoderma spp, Mol Microbiol. 4, 839-43.
- 27. Herrera-Estrella, A., Goldman, G. H., van Montagu, M. & Geremia, R. A. (1993)

 Electrophoretic karyotype and gene assignment to resolved chromosomes of

 Trichoderma spp, Mol Microbiol. 7, 515-21.
- 28. Johnston, M. & Davis, R. W. (1984) Sequences that regulate the divergent GAL1-GAL10 promoter in Saccharomyces cerevisiae, Mol Cell Biol. 4, 1440-8.
- 29. Kim, J. & Iyer, V. R. (2004) Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles, Mol Cell Biol. 24, 8104-12.
- 30. Kolell, K. J. & Crawford, D. L. (2002) Evolution of Sp transcription factors, Mol Biol Evol. 19, 216-22.
- 31. Koyanagi, K. O., Hagiwara, M., Itoh, T., Gojobori, T. & Imanishi, T. (2005) Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system, Gene. 353, 169-76.
- 32. Kruglyak, S. & Tang, H. (2000) Regulation of adjacent yeast genes, Trends Genet. 16, 109-11.
- 33. Lander, E. S.Linton, L. M.Birren, B.Nusbaum, C.Zody, M. C.Baldwin, J.Devon, K.Dewar, K.Doyle, M.FitzHugh, W.Funke, R.Gage, D.Harris, K.Heaford, A.Howland, J.Kann, L.Lehoczky, J.LeVine, R.McEwan, P.McKernan, K.Meldrim, J.Mesirov, J. P.Miranda, C.Morris, W.Naylor, J.Raymond, C.Rosetti, M.Santos, R.Sheridan, A.Sougnez, C.Stange-Thomann, N.Stojanovic, N.Subramanian, A.Wyman, D.Rogers, J.Sulston, J.Ainscough, R.Beck, S.Bentley, D.Burton, N.Cheng, J. F.Olsen, A.Lucas, S.Elkin,

- C.Uberbacher, E.Frazier, M., et al. (2001) Initial sequencing and analysis of the human genome, Nature. 409, 860-921.
- 34. Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. (1992) CpG islands as gene markers in the human genome, Genomics. 13, 1095-107.
- 35. Lawrence, J. G. (2003) Gene organization: selection, selfishness, and serendipity, Annu Rev Microbiol. 57, 419-40.
- 36. Lee, J. A., Suh, D. C., Kang, J. E., Kim, M. H., Park, H., Lee, M. N., Kim, J. M., Jeon, B. N., Roh, H. E., Yu, M. Y., Choi, K. Y., Kim, K. Y. & Hur, M. W. (2005) Transcriptional activity of Sp1 is regulated by molecular interactions between the zinc finger DNA binding domain and the inhibitory domain with corepressors, and this interaction is modulated by MEK, J Biol Chem. 280, 28061-71.
- 37. Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J. B., Volkert, T. L., Fraenkel, E., Gifford, D. K. & Young, R. A. (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae, Science. 298, 799-804.
- 38. Luttik, M. A., Kotter, P., Salomons, F. A., van der Klei, I. J., van Dijken, J. P. & Pronk, J. T. (2000) The Saccharomyces cerevisiae ICL2 gene encodes a mitochondrial 2-methylisocitrate lyase involved in propionyl-coenzyme A metabolism, J Bacteriol. 182, 7007-13.
- 39. Martin, J. F. (2000) Molecular control of expression of penicillin biosynthesis genes in fungi: regulatory proteins interact with a bidirectional promoter region, J Bacteriol. 182, 2355-62.
- 40. Meier-Noorden, M., Flindt, S., Kalinke, U. & Hinz, T. (2004) A CpG-rich bidirectional promoter induces the T-cell death-associated gene 51 and downregulates an inversely oriented transcript during early T-cell activation, Gene. 338, 197-207.
- 41. Muro-Pastor, M. I., Gonzalez, R., Strauss, J., Narendja, F. & Scazzocchio, C. (1999) The GATA factor AreA is essential for chromatin remodelling in a eukaryotic bidirectional promoter, Embo J. 18, 1584-97.

- 42. Nicholas, K., Nicholas, H. J. & Deerfield. (1997) GeneDoc: Analysis and visualization of genetic variation. in, EMBNEW,
- 43. Oei, S. L., Babich, V. S., Kazakov, V. I., Usmanova, N. M., Kropotov, A. V. & Tomilin, N. V. (2004) Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters, Genomics. 83, 873-82.
- 44. Overton, J. S. G. C. (1997) TESS: Transcription Element Search Software on the WWW in, Computational Biology and Informatics Laboratory School of Medicine University of Pennsylvania, Pennsylvania.
- 45. Patton, J., Block, S., Coombs, C. & Martin, M. E. (2005) Identification of functional elements in the murine Gabp alpha/ATP synthase coupling factor 6 bi-directional promoter, Gene. 369, 35-44.
- 46. Punt, P. J., Oliver, R. P., Dingemanse, M. A., Pouwels, P. H. & van den Hondel, C. A. (1987) Transformation of Aspergillus based on the hygromycin B resistance marker from Escherichia coli, Gene. 56, 117-24.
- 47. Rice, P., Longden, I. & Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite, Trends Genet. 16, 276-7.
- 48. Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989) Molecular Cloning: A laboratory Manual, Cold Spring Harbor, NY.
- 49. Samuels, G. J. (1996) Trichoderma: a review of biology and systematics of the genus, Mycology Research, 923-935.
- 50. Schneider, K. & Beck, C. F. (1986) Promoter-probe vectors for the analysis of divergently arranged promoters, Gene. 42, 37-48.
- 51. Schneider, K. & Beck, C. F. (1987) New expression vectors for identifying and testing signal structures for initiation and termination of transcription, Methods Enzymol. 153, 452-61.
- 52. Seki, Y., Ikeda, S., Kiyohara, H., Ayabe, H., Seki, T. & Matsui, H. (2002) Sequencing analysis of a putative human O-sialoglycoprotein endopeptidase gene (OSGEP) and analysis of a bidirectional promoter between the OSGEP and APEX genes, Gene. 285, 101-8.
- 53. Shimada, T., Fujii, H. & Lin, H. (1989) A 165-base pair sequence between the dihydrofolate reductase gene and the divergently transcribed upstream gene is sufficient for bidirectional transcriptional activity, J Biol Chem. 264, 20171-4.

- 54. Smale, S. T. (2001) Core promoters: active contributors to combinatorial gene regulation, Genes Dev. 15, 2503-8.
- 55. Staben, C., Jensen, B., Singer, M. & Pollock, J. (1989) Use of a bacterial Hygromycin B resistance gene as a dominant selectable marker in Neurospora crassa transformation., Fungal Genetics Newsletter. 36, 79-81.
- 56. Sussman, H. (2001) Choosing the Best Reporter Assay in The Scientist pp. 25-32
- 57. Takai, D. & Jones, P. A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22, Proc Natl Acad Sci U S A. 99, 3740-5.
- 58. Takai, D. & Jones, P. A. (2004) Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome, Mol Biol Evol. 21, 463-7.
- 59. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25, 4876-82.
- 60. Torigoe, T., Izumi, H., Yoshida, Y., Ishiguchi, H., Okamoto, T., Itoh, H. & Kohno, K. (2003) Low pH enhances Sp1 DNA binding activity and interaction with TBP, Nucleic Acids Res. 31, 4523-30.
- 61. Travers, M. T., Cambot, M., Kennedy, H. T., Lenoir, G. M., Barber, M. C. & Joulin, V. (2005) Asymmetric expression of transcripts derived from the shared promoter between the divergently oriented ACACA and TADA2L genes, Genomics. 85, 71-84.
- 62. Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P. & Myers, R. M. (2004) An abundance of bidirectional promoters in the human genome, Genome Res. 14, 62-6.
- 63. Venter, J. C.Adams, M. D.Myers, E. W.Li, P. W.Mural, R. J.Sutton, G. G.Smith, H. O.Yandell, M.Evans, C. A.Holt, R. A.Gocayne, J. D.Amanatides, P.Ballew, S.Clark, A. G.Nadeau, J.McKusick, V. A.Zinder, N.Levine, A. J.Roberts, R. J.Simon, V. A.Neelam, B.Nusskern, D.Rusch, D. B.Salzberg, S.Shao, W.Shue, B.Sun, J.Wang, Z.Wang, A.Wang, X.Wang, J.Wei, M.Wides, R.Xiao, C.Yan, C., et al. (2001) The sequence of the human genome, Science. 291, 1304-51.
- 64. Wang, Y. & Leung, F. C. (2004) An evaluation of new criteria for CpG islands in the human genome as gene markers, Bioinformatics. 20, 1170-7.

CURRICULUM VITAE

Dados pessoais:

Nome: Estela Ynés Valencia Morante.

Nascimento: 28/07/1966

Naturalidade: Lambayeque – Perú.

Formação Acadêmica:

Escolar Colégio Nacional Daniel Alcides Carrion, 1974 -1978

Colégio Nacional San Pedro 1979 -1980

Colégio Inmaculada de la Merced 1981-1983

Graduação Bachiller em Ciências - Biologia outorgado 16 de Dezembro de 1992

Faculdade de Ciências Biológicas

Universidade Nacional Pedro Ruiz Gallo.

Lambayeque -Perú 1985 - 1992

Licenciada em: Biologia especialidade Microbiologia - Parasitologia

Outorgado o titulo 6 de setembro de 1994.

Participação em eventos científicos

1. EL-DORRY, H., CHAMBERGO, F.S, VALENCIA, E.Y.

Elucidation of the metabolic fate of glucose in *Trichoderma reesei* using EST analysis and cDNA microarray: a step toward metabolic engineering of the regulatory networks for anaerobic metabolism of glucose

Metabolic engineering IV: Applied System Biology. Tuscany, Italia. 2002.

2. **ESTELA Y. VALENCIA**; FELIPE S. CHAMBERGO; AUGUSTO S.P. RAMOS AND HAMZA ELDORRY.

Analysis of a bidirectional promoter controlling the expression of the citrate synthase and isocitrate lyase genes in Trichoderma reesei.

XXXIV Reuniao Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2005, Águas de Lindóia, SP.

XXXI Reunião Anual da SBBq Programa e Resumos, em CD, 2005