

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Caracterização de proteoformas expressas no estágio larval do parasito
Echinococcus granulosus

Tese de Doutorado

Karina Rodrigues Lorenzatto

Porto Alegre, maio de 2015

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Caracterização de proteoformas expressas no estágio larval do parasito
Echinococcus granulosus

Tese submetida ao Programa de Pós-Graduação em
Biologia Celular e Molecular da UFRGS, como
requisito parcial para a obtenção do grau de Doutor
em Biologia Celular e Molecular

Karina Rodrigues Lorenzatto

Orientador: Dr. Henrique Bunselmeyer Ferreira

Porto Alegre, maio de 2015

Este trabalho foi realizado no Laboratório de Genômica Estrutural e Funcional e no Laboratório de Biologia Molecular de Cestódeos do Centro de Biotecnologia da UFRGS, e no *Proteomics Center of Excellence* da Northwestern University (EUA), sendo financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), e pela Fundação de Apoio a Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

Ao Jônadas.

À minha mãe Maria e irmã Juliana.

AGRADECIMENTOS

Ao orientador, Dr. Henrique Bunselmeyer Ferreira, pela dedicação, pelo exemplo, e pela amizade.

Ao Dr. Arnaldo Zaha, pelo exemplo, apoio e pela amizade.

Ao Dr. Diego Bonatto e ao Dr. Alexandre Bailão, pela participação em minha comissão de acompanhamento no PPGBCM e pelas sugestões ao longo do trabalho.

Ao Dr. Rodolfo Paredes e ao Dr. Norbel Galanti pela colaboração no trabalho.

À Dr. Karina M. Monteiro, pelas sugestões e amizade.

Ao Dr. Neil Kelleher (*Northwestern University*, EUA) pela oportunidade de trabalhar em seu laboratório.

À equipe do *Proteomics Center of Excellence* (*Northwestern University*, EUA), em especial ao Paul Thomas e à Ioanna Ntai, pelo auxílio nos experimentos.

À equipe do laboratório do Dr. Neil, pelo auxílio e amizade, em especial ao Kyunggon Kim, que é para mim não somente um exemplo profissional mas também pessoal. Agradeço também aos colegas brasileiros, Luís e Rafael, pelas sugestões e pela amizade.

Ao Centro de Microscopia Eletrônica (CME) da UFRGS, pelo suporte na microscopia confocal.

À Unidade de Química de Proteínas e Espectrometria de Massas (Uniprote-MS), do Centro de Biotecnologia da UFRGS, em especial à Jozi, pelo auxílio nas análises de LC-MS/MS e pela amizade.

Ao laboratório de espectrometria de massas (proposta MAS-14831), do Laboratório Nacional de Biociências, Campinas, SP, em especial à Adriana Paes Leme e à Romênia Domingues, pelo auxílio nas análises.

Ao Dr. José Cláudio F. Moreira do Departamento de Bioquímica da UFRGS e à Dr. Fernanda Caregnato, pelo auxílio nos ensaios enzimáticos.

À ex-colega, Dr. Marbella, pelo auxílio na modelagem da EgFBA1 e pela amizade.

À ex-aluna de IC Gabi, agora mestranda, pelo apoio nos experimentos e pela coragem em ingressar no mundo da biologia computacional, estimulando a minha aproximação do tema.

Aos colegas do laboratório 210, Alemão, Andrey, Fernanda, Gabi W., Jéssica, Lauren, Natália, Veri, Vinicius e também aos ex-colegas Karin e Kendi, pela convivência, pelo auxílio e pela amizade.

Aos colegas dos laboratórios 206 e 204, Aline, Bruna, Cris, Edi, Fernanda Z., Filipe, Guilherme B., Guilherme S., Helier, Jeferson, João, Martín, Nina, Sérgio, Tati e também às ex-colegas Carol, Cibele e Daiani, pelo apoio, pelas sugestões e pela amizade.

A todos os colegas dos demais laboratórios do Centro de Biotecnologia, pelo auxílio.

Aos professores do PPGBCM, por todos os ensinamentos.

Aos membros da banca examinadora, por terem aceitado o convite para avaliar este trabalho.

Aos secretários do PPGBCM, Luciano e Silvinha, pela competência e amizade.

À funcionária Terezinha, pela disposição.

Ao CNPq, à CAPES e à FAPERGS, pelo apoio financeiro.

Aos amigos Aline, Dani, Nice e Marcelo, pela parceria constante.

À minha mãe Maria e irmã Juliana, pelo incentivo, paciência e apoio em todos os momentos.

Ao Jônadas, pelo incentivo, carinho e pela compreensão nos momentos de ausência.

A todos que, de alguma forma, contribuíram para a realização deste trabalho.

ÍNDICE

LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES.....	8
LISTA DE FIGURAS	10
RESUMO.....	11
ABSTRACT	12
1. INTRODUÇÃO	13
1.1 Proteoformas	13
1.2 Proteínas <i>moonlighting</i>	15
1.3 Funções <i>moonlighting</i> em parasitos	15
1.4 <i>Echinococcus granulosus</i>	17
1.5 Hidatidose cística	20
1.6 Interações <i>Echinococcus</i> -hospedeiro	21
2. JUSTIFICATIVAS.....	23
3. OBJETIVOS	24
4. MATERIAIS E MÉTODOS E RESULTADOS	25
4.1 Frutose-bifosfato-aldolase e enolase de <i>Echinococcus granulosus</i> : genes, padrões de expressão e interações proteicas de duas potenciais proteínas <i>moonlighting</i>	26
4.2 Frutose-bifosfato-aldolase e enolase: potenciais proteínas <i>moonlighting</i> de <i>E. granulosus</i>	42
4.3 Análise proteômica <i>top down</i> e <i>bottom up</i> de frações subcelulares de protoescólices de <i>E. granulosus</i> contendo proteínas de baixa massa	49
4.4 Análise proteômica <i>top down</i> e <i>bottom up</i> de frações subcelulares de protoescólices de <i>E. granulosus</i> contendo proteínas de maior massa	77
5. DISCUSSÃO	81
6. CONCLUSÕES E PERSPECTIVAS	90
7. REFERÊNCIAS BIBLIOGRÁFICAS	91
8. <i>CURRICULUM VITAE</i>	101

LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES

AcN: acetilação N-terminal
BLAST: *Basic Local Alignment Search Tool*
CG: camada germinativa
CoA: coenzima A
DNA: ácido desoxirribonucleico
DTT: ditioneitol
EgEno: enolase de *E. granulosus*
EgFBA: frutose-bifosfato-aldolase de *Echinococcus granulosus*
EMN: excisão da metionina N-terminal
ES: excreção/secreção
FBA: frutose-bifosfato-aldolase
Fru-1,6-P2: frutose 1,6-bifosfato
GELFrEE: *Gel-Eluted Liquid Fraction Entrapment Electrophoresis*
GO: *Gene ontology*
GST: glutationa-S-transferase
HSP: proteína de choque térmico
IPP: interação proteína-proteína
IUPAC: *International Union of Pure and Applied Chemistry*
kDa: quilodalton(s)
 K_m : constante de Michaelis
LC-MS/MS: cromatografia líquida acoplada à espectrometria de massas em tandem
MAP: metionina aminopeptidase
MCT: modificação cotraducional
MPT: modificação pós-traducional
MS: espectrometria de massas
NADH: dinucleotídeo de nicotinamida e adenina
NAT: N-acetiltransferase
NCBI: *National Center for Biotechnology Informaion*
PBS: solução salina tamponada com fosfato
PGA: fosfoglicerato

pH: potencial hidrogeniônico

PSC: protoescólicas

rEgEno: proteína EgEno recombinante

rEgFBA: proteína EgFBA recombinante

RNA: ácido ribonucleico

SDS: dodecilsulfato de sódio

SDS-PAGE: eletroforese em gel de poliacrilamida na presença de SDS

TCA: ácido tricloroacético

TEV: *tobacco etch virus*

TFA: ácido trifluoroacético

U: unidade de atividade enzimática

UV: ultravioleta

VEs: vesículas extracelulares

WHO: *World Health Organization*

LISTA DE FIGURAS

Figura 1. Origens do termo proteoforma para descrever a variabilidade biológica..	14
Figura 2. Ciclo vital do <i>E. granulosus</i>	18
Figura 3. Estágio larval do parasito <i>E. granulosus</i>	19
Figura 4. Distribuição geográfica de <i>E. granulosus</i> e da hidatidose cística.....	21
Figura 5. Árvore filogenética de isoformas de FBA de vertebrados e invertebrados.....	44
Figura 6. Árvore filogenética de isoformas de enolase de vertebrados e invertebrados..	45
Figura 7. Caracterização cinética das enzimas glicolíticas recombinantes..	46
Figura 8. Classificação funcional do GO das proteínas de interação da rEgFBA1 e da rEgEno1.....	48
Figura 9. Fracionamento por GELFrEE das amostras nucleares e citosólicas de protoescólices de <i>E. granulosus</i>	78
Figura 10. Distribuição das proteínas de maior massa identificadas em frações nucleares e citosólicas de protoescólices de <i>E. granulosus</i> através da abordagem <i>bottom up</i>	78
Figura 11. Visão global do proteoma de frações nucleares e citosólicas de protoescólices de <i>E. granulosus</i>	80
Figura 12. Perfis de transcrição das isoformas de EgFBA e EgEno nos diferentes estágios do ciclo de vida do <i>E. granulosus</i>	82

RESUMO

O termo proteoformas refere-se a todas as formas moleculares na qual variantes proteicas codificadas por um único gene podem ser encontradas, incluindo mudanças devido a variação genética, *splicing* alternativo e modificações cotraducionais e pós-traducionais (MCTs e MPTs, respectivamente). Em *Echinococcus granulosus*, o agente causador da hidatidose cística, a caracterização de proteoformas ainda não foi sistematicamente abordada e pode revelar mecanismos moleculares relevantes para o estabelecimento do parasito nos hospedeiros intermediários. A caracterização tanto de proteoformas da frutose-bifosfato-aldolase (FBA) e enolase, potencialmente *moonlighting*, como de outras proteoformas expressas em frações subcelulares do estágio larval deste parasito foi realizada. Em relação às enzimas glicolíticas, várias evidências indicam a EgFBA1 e a EgEno1 como sendo proteínas *moonlighting*: (i) as localizações não usuais destas proteínas no tegumento de protoescólices e em produtos de excreção/secreção (ES) *in vivo* (líquido hidático) e *in vitro*; (ii) características estruturais destas proteínas, com destaque para a identificação de um sítio de ligação à F-actina na estrutura da EgFBA1; (iii) a habilidade destas proteínas de interagirem com várias proteínas não relacionadas à glicólise. Em relação ao repertório de proteínas expressas em frações subcelulares, frações nucleares e citosólicas de protoescólices de *E. granulosus* foram analisadas através de proteômica *top down* e *bottom up*. De modo geral, 525 proteínas foram identificadas, sendo 224 e 156 proteínas exclusivamente detectadas em frações nucleares e citosólicas, respectivamente. Além da identificação de proteínas, nossa abordagem de proteômica *top down* também permitiu a caracterização de MCTs e MPTs, destacando a excisão da metionina N-terminal e a acetilação N-terminal como modificações conservadas nas proteínas de *E. granulosus*. Nossos resultados representam as primeiras evidências de funções alternativas realizadas por enzimas glicolíticas em *E. granulosus* e sugerem que proteínas multifuncionais devem desempenhar papéis importantes na interação parasito-hospedeiro. Os proteomas nuclear e citosólico de protoescólices também foram descritos, revelando novos aspectos da biologia do parasito, incluindo proteoformas com MCTs e MPTs, as quais devem desempenhar papéis críticos na regulação de processos celulares deste parasito.

ABSTRACT

The term proteoforms refers to all molecular forms in which the protein product of a single gene can be found, including changes due to genetic variation, alternatively spliced RNA transcripts and co-translational and post-translational modifications (CTMs and PTMs, respectively). In *Echinococcus granulosus*, the causative agent of hydatid disease, the characterization of proteoforms has never been systematically addressed and it may reveal molecular mechanisms underlying the parasite biology. Here, the characterization of fructose-bisphosphate aldolase (FBA) and enolase proteoforms, potentially moonlighting, and also, the identification and characterization of proteoforms expressed in subcellular fractions of *E. granulosus* larval stage were performed. In relation to the glycolytic enzymes, several evidences indicate EgFBA1 and EgEno1 as moonlighting proteins: (i) their unusual locations in protoscolex tegument and in *in vivo* (hydatid fluid) and *in vitro* excretory/secretory (ES) products; (ii) their structural features, highlighting the F-actin binding site in the EgFBA1 structure; (iii) their ability to interact with several proteins non-related to glycolysis. In relation to the repertoire of proteins expressed in subcellular fractions, nuclear and cytosolic fractions from *E. granulosus* protoscoleces were analyzed by top down and bottom up proteomics for protein identification and characterization. Altogether, 525 proteins were identified, with 224 and 156 proteins exclusively detected in nuclear and cytoplasmic fractions, respectively. Besides protein identification, our top down approach also provided CTMs and PTMs characterization, highlighting N-terminal methionine excision and N-terminal acetylation as conserved modifications in *E. granulosus* proteins. Overall, our results provided the first experimental evidences of alternative functions performed by glycolytic enzymes in *E. granulosus* and suggested that multifunctional proteins might play important roles in host-parasite interplay. We also provided the description of nuclear and cytosolic proteomes which revealed new aspects of parasite biology, highlighting proteoforms with CTMs and PTMS which might be playing critical roles in regulating cellular processes occurring in this parasite species.

1. INTRODUÇÃO

1.1 Proteoformas

O sequenciamento de genomas revelou que não existe uma relação entre o número de genes e a complexidade fenotípica, sugerindo que parte da complexidade dos sistemas biológicos poderia estar relacionada com variações proteicas (Patrushev & Minkevich, 2008; Pruitt *et al.*, 2007). Em nível de DNA, RNA e proteína, a complexidade pode ser gerada através de variações alélicas, *splicing* alternativo, modificações cotraducionais e pós-traducionais (MCTs e MPTs, respectivamente) (Smith *et al.*, 2013). Estes eventos criam proteínas distintas que modulam uma ampla variedade de processos biológicos, desde sinalização celular até regulação gênica.

A análise proteômica surgiu como uma ferramenta essencial para revelar a complexidade das variantes proteicas (Mann & Kelleher, 2008). Neste contexto, a espectrometria de massas emergiu como uma plataforma-chave para a análise proteômica, com os propósitos de identificar proteínas e caracterizar alterações relacionadas com a produção de diferentes variantes proteicas (Olsen & Mann, 2013). Essencialmente, duas abordagens contrastantes referidas como *bottom up* e *top down* têm sido utilizadas para atender estes propósitos (Kellie *et al.*, 2010). Na abordagem *bottom up*, as proteínas são digeridas usando tripsina ou outras proteases, e os peptídeos gerados são então analisados por cromatografia líquida acoplada a espectrometria de massas em tandem (LC-MS/MS). Na abordagem *top down*, a identificação de proteínas é obtida diretamente da fragmentação de proteínas intactas (Ahlf *et al.*, 2013). Quando disponível, a *top down* fornece dados mais ricos tanto para a identificação precisa como para a caracterização da composição molecular (Kelleher *et al.*, 2014; Tran *et al.*, 2011).

A disponibilidade de tecnologias para capturar a variabilidade proteica tem gerado uma nomenclatura diversa para referir-se a complexidade de proteínas. Esta variabilidade tem sido chamada de diferentes maneiras, incluindo formas, isoformas, e variantes proteicas (Prabakaran *et al.*, 2012; Smith *et al.*, 2013). Isoformas é um termo amplamente utilizado, mas geralmente de forma incorreta, conforme a definição estabelecida pela *International Union of Pure and Applied Chemistry* (IUPAC), para a qual isoformas referem-se a variantes produzidas devido a diferenças genéticas e não a diferenças em

nível proteico (Schlüter *et al.*, 2009). Recentemente, o termo proteoformas foi proposto para designar todas as diferentes formas moleculares nas quais produtos proteicos de um único gene podem ser encontrados (Figura 1) (Smith *et al.*, 2013).

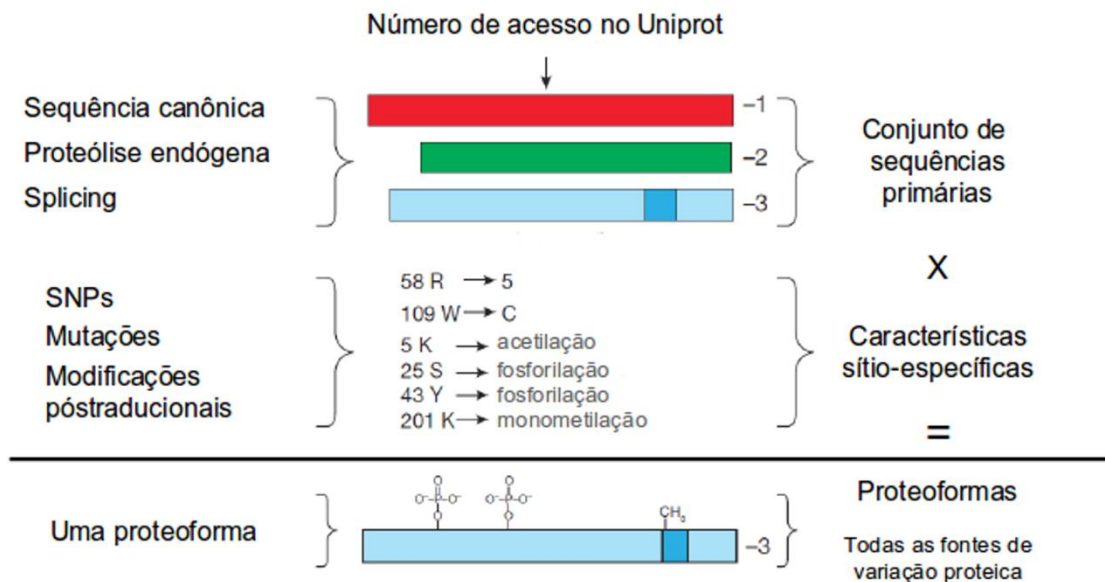


Figure 1. Origens do termo proteoforma para descrever a variabilidade biológica. O UniProt é um banco de dados não redundante, de modo que existe um esforço para manter um único número de acesso para cada gene. SNP, polimorfismo de nucleotídeo único (do inglês single nucleotide polymorphism). Modificado de (Smith *et al.*, 2013).

A caracterização de proteoformas nem sempre é um processo trivial. Para tornar este cenário ainda mais complexo, outro aspecto a ser considerado dentro do universo de proteoformas são as proteoformas *moonlighting*, as quais exibem propriedades multifuncionais (Jeffery, 2009). Os avanços na instrumentação em espectrometria de massas devem cada vez mais favorecer a captura da complexidade proteica. Considerando que as proteoformas são reguladoras primárias de funções celulares, esforços para a caracterização delas são fundamentais para entender mecanismos moleculares associados a diferentes processos biológicos.

1.2 Proteínas *moonlighting*

Proteínas *moonlighting* consistem em uma classe de proteínas multifuncionais na qual uma única cadeia polipeptídica realiza múltiplas funções (Jeffery, 1999). Desde a definição do termo *moonlighting*, centenas de proteínas com esta propriedade têm sido identificadas nos três domínios da vida (Henderson & Martin, 2014; Mani *et al.*, 2015). Exemplos conhecidos de proteínas *moonlighting* incluem diversas classes de proteínas, como receptores, enzimas, fatores de transcrição e proteínas ribossômicas. Entre as funções *moonlighting* destas proteínas destacam-se regulação transcricional, ligação a receptores, funções relacionados com apoptose e outras funções reguladoras (Khan *et al.*, 2014; Mani *et al.*, 2015).

As proteínas *moonlighting* conhecidas utilizam diversos mecanismos para troca de funções (Jeffery, 2009). Algumas proteínas podem ter funções diferentes quando expressas em diferentes locais no interior da célula. Por exemplo, várias proteínas citosólicas são também encontradas na superfície celular, onde elas podem funcionar como receptores (Kunert *et al.*, 2007; Luo *et al.*, 2009; Modun *et al.*, 2000). Outras proteínas *moonlighting* trocam de função quando são secretadas (Niinaka *et al.*, 1998), quando sofrem MPTs (Pal-Bhowmick *et al.*, 2007) ou quando formam complexos multiproteicos (Zheng *et al.*, 2003).

Muitas das proteínas multifuncionais conhecidas são enzimas altamente conservadas e, especialmente enzimas envolvidas na glicólise têm sido descritas com funções *moonlighting* (Huberts & van der Klei, 2010). Existem evidências de que 7 de 10 enzimas da via glicolítica exibem atividades *moonlighting* (Sriram *et al.*, 2005). Dependendo do organismo, as funções alternativas das enzimas glicolíticas podem variar desde funções estruturais e reguladoras (Gancedo & Flores, 2008; Starnes *et al.*, 2009; Torres-Machorro *et al.*, 2015) até, em organismos parasitos, funções de interação com e invasão do hospedeiro (Ginger, 2014; Gómez-Arreaza *et al.*, 2014; Karkowska-Kuleta & Kozik, 2014).

1.3 Funções *moonlighting* em parasitos

Em relação a parasitos, o conceito de funções *moonlighting* é interessante, uma vez que a simplificação biológica é uma das principais características geralmente associadas com a adaptação ao parasitismo (Collingridge *et al.*, 2010). A extensa descrição de

proteínas *moonlighting* em bactérias torna claro que elas são uma das principais beneficiárias deste mecanismo de evolução (Henderson & Martin, 2011). Mas, proteínas multifuncionais também têm sido descritas em muitos outros parasitos e muitas delas desempenham papéis cruciais para interação com seus hospedeiros (Gómez-Arreaza *et al.*, 2014; Marcos *et al.*, 2014).

Dentre as proteínas já descritas como tendo funções *moonlighting* em organismos parasitas, destacam-se enzimas da via glicolítica, como a frutose-bifosfato-aldolase (FBA) e a enolase. Além de apresentarem funções não glicolíticas na regulação do citoesqueleto, no transporte celular, na transcrição e na apoptose (Díaz-Ramos *et al.*, 2012; Pancholi, 2001), estas enzimas desempenham papéis importantes em processos-chave na biologia de organismos parasitas, como motilidade, adesão e invasão, entre outros (Jewett & Sibley, 2003; Marcos *et al.*, 2014; Pomel *et al.*, 2008). Essas proteínas também apresentam capacidade de ligação a macromoléculas do hospedeiro, como o plasminogênio (Bernal *et al.*, 2004; Díaz-Martín *et al.*, 2013; Marcilla *et al.*, 2007), cuja ligação promove a degradação de componentes de matriz extracelular, sendo essencial para a disseminação de bactérias e fungos nos tecidos do hospedeiro (Chaves *et al.*, 2015; Haile *et al.*, 2006; Nogueira *et al.*, 2010) e, no caso de helmintos, parece contribuir para o estabelecimento dos parasitos no hospedeiro (de la Torre-Escudero *et al.*, 2010; Marcilla *et al.*, 2007), impedindo a formação de coágulos ao redor dos parasitos (Ramajo-Hernández *et al.*, 2007). Adicionalmente, enzimas glicolíticas têm sido comumente identificadas no tegumento e nos produtos de excreção/secreção (ES) de diversos parasitos, onde apresentam propriedades antigênicas e de interação com o hospedeiro (de la Torre-Escudero *et al.*, 2010; Harnett, 2014; Liu *et al.*, 2012; Zhang *et al.*, 2014).

Enzimas da via glicolítica de *Echinococcus granulosus*, como a FBA (EgFBA) e a enolase (EgEno), foram identificadas em diferentes componentes do cisto hidático (Aziz *et al.*, 2011; Cui *et al.*, 2013; Monteiro *et al.*, 2010). Estudos proteômicos detectaram a presença das proteínas EgFBA e EgEno em protoescólices, camada germinativa e líquido hidático, demonstrando também, em alguns casos, caráter antigênico para estas proteínas. A presença de proteínas como essas na camada germinativa (principal interface com o hospedeiro) e no líquido hidático (produtos de ES) de *E. granulosus*, inclusive fora do contexto intracelular comumente associado a componentes da via glicolítica, sugere a participação dessas enzimas em outros processos fisiológicos, dentre os quais aqueles de

interação com o hospedeiro.

1.4 *Echinococcus granulosus*

O gênero *Echinococcus* pertence à classe Cestoda (ordem Cyclophyllidea, família Taeniidae). As duas principais espécies de importância médica e veterinária são *E. granulosus* e *Echinococcus multilocularis*, que causam, respectivamente, a hidatidose cística e a hidatidose alveolar (McManus & Thompson, 2003). Devido à distribuição cosmopolita e o importante impacto tanto na saúde de seres humanos como de animais domésticos, *E. granulosus* é considerada a espécie mais relevante (Romig, 2003).

Uma importante característica da biologia do *E. granulosus* é que ele apresenta considerável variabilidade genética (Alvarez Rojas *et al.*, 2014; McManus, 2013). Atualmente, dez genótipos são reconhecidos (G1-G10), sendo alguns considerados até mesmo espécies diferentes, referidos conjuntamente como *Echinococcus sensu lato* (*s. l.*) (Alvarez Rojas *et al.*, 2014). Uma nova proposição taxonômica retém o nome *E. granulosus sensu stricto* (*s. s.*) para G1-G3. A variação intraespecífica em *E. granulosus* (*s. l.*) deve estar associada com diferentes padrões de ciclo de vida, especificidade aos hospedeiros, taxas de desenvolvimento, antigenicidade, dinâmica de transmissão, sensibilidade a agentes quimioterapêuticos e patologia (McManus, 2009; Thompson & McManus, 2002).

Como todos os cestódeos, as espécies do gênero *Echinococcus* requerem dois hospedeiros mamíferos para completar os seus ciclos vitais (Romig, 2003). No caso de *E. granulosus*, o verme adulto vive no intestino delgado do hospedeiro definitivo, que geralmente é o cão doméstico. Os ovos eliminados pelo verme adulto são infectivos para os hospedeiros intermediários que, em geral, são ungulados domésticos (especialmente bovinos, ovinos e suínos). O homem é infectado acidentalmente, devido ao manuseio de hospedeiros definitivos contaminados e à transferência de ovos das mãos para a boca, ou ainda pela ingestão de alimentos contaminados (Eckert & Deplazes, 2004). Nos hospedeiros intermediários, os ovos eclodem liberando a oncosfera, que, nas vísceras do hospedeiro (geralmente fígado ou pulmões), se diferencia na forma larval patogênica do parasito, o metacestódeo ou cisto hidático, causando a hidatidose cística. O ciclo vital de *E. granulosus* é fechado quando os protoescólices, formas pré-adultas do parasito presentes no interior do cisto, são ingeridos por um hospedeiro definitivo, podendo diferenciar-se em

vermes adultos (Figura 2).

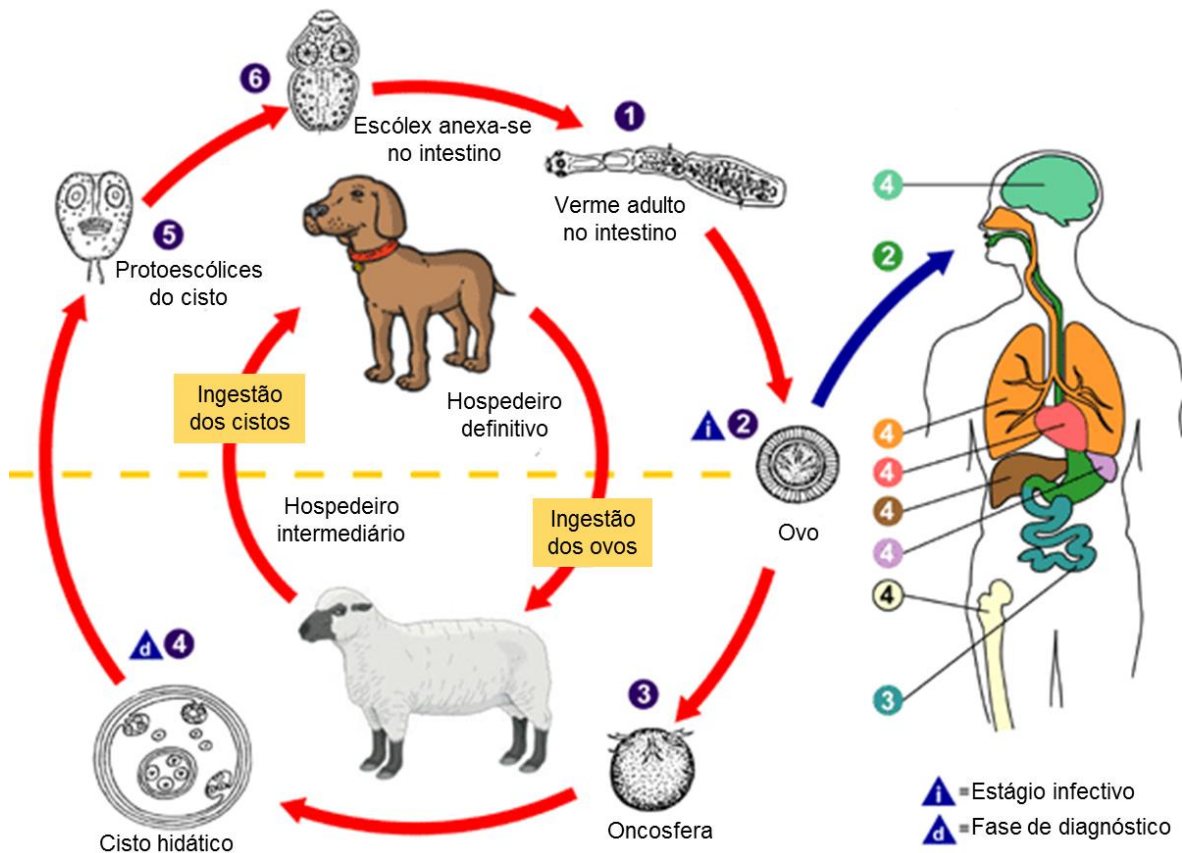


Figure 2. Ciclo vital do *E. granulosus*. (1) O verme adulto vive no intestino delgado do hospedeiro definitivo (principalmente o cão doméstico). (2) As proglótides grávidas liberam os ovos, os quais são eliminados nas fezes. Após a ingestão por um hospedeiro intermediário adequado (bovinos, ovinos, o homem, etc.), os ovos eclodem no intestino delgado e liberam a (3) oncosfera que penetra na parede do intestino e migra através do sistema circulatório para vários órgãos, especialmente fígado e pulmões. Nestes órgãos, a oncosfera diferencia-se em um (4) cisto hidático o qual aumenta gradualmente de tamanho, produzindo (5) protoescólices no seu interior. O hospedeiro definitivo é infectado pela ingestão de cistos existentes nos órgãos dos hospedeiros intermediários. Após a ingestão, os (6) protoescólices evaginam, fixam-se à mucosa intestinal e se transformam em vermes adultos. Os seres humanos são infectados pela ingestão de ovos, com a liberação de oncosferas no intestino delgado e o desenvolvimento de cistos (4) em vários órgãos. (Fonte: <http://www.cdc.gov/parasites/echinococcosis/biology.html>).

O cisto hidático é unilocular e pode alcançar um tamanho grande, de 20 cm de diâmetro ou até mais (Rogan *et al.*, 2006) (Figura 3A). A cavidade do cisto é preenchida com um líquido, denominado líquido hidático, o qual é uma mistura complexa de moléculas derivadas do parasito e componentes do soro do hospedeiro (Holcman & Heath, 1997; Juyi *et al.*, 2013). A camada mais interna do cisto, a camada germinativa, é uma camada celular indiferenciada proliferativa, responsável tanto pela formação de protoescólices (Figura 3B), como pela síntese da camada laminar, uma camada mais

externa, acelular, rica em carboidratos (Díaz *et al.*, 2011). Os protoescólices surgem a partir de evaginações decorrentes da proliferação e diferenciação de células de qualquer região da camada germinativa (Galindo *et al.*, 2002). Estas se diferenciam em uma região anterior (com ventosas e acúleos) e uma região posterior ou caudal (indiferenciada) bem no início do processo de crescimento.

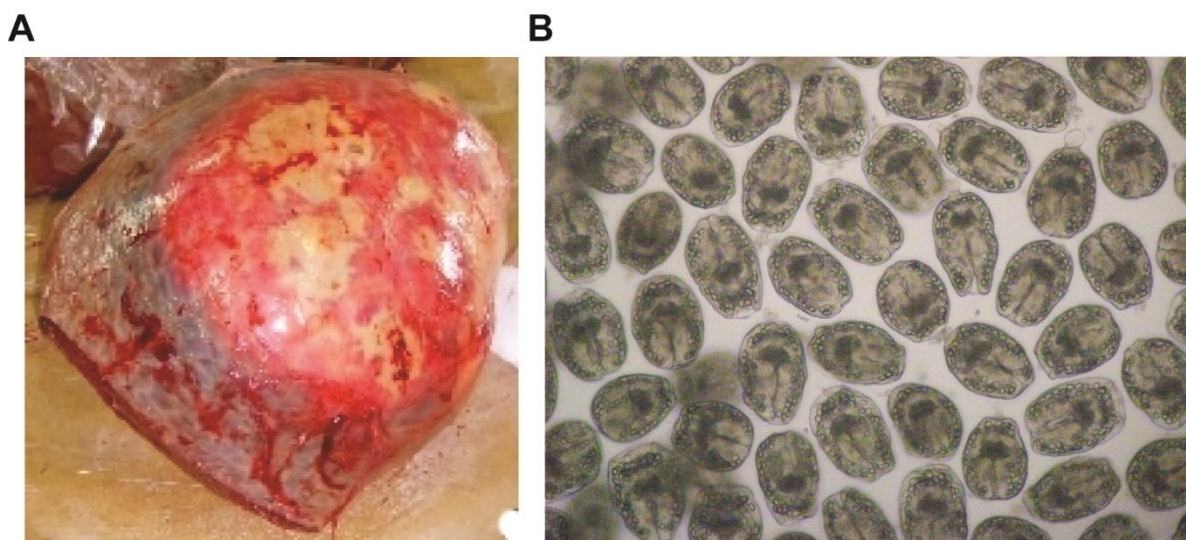


Figure 3. Estágio larval do parasito *E. granulosus*. **A:** Aspecto macroscópico de um cisto hidático em pulmão bovino (foto de A. Teichmann); **B:** Protoescólices liberados de um cisto hidático (foto de J. A. Debarba).

Durante a formação do protoescólex, diferentes tipos celulares desenvolvem-se e organizam-se em territórios celulares com funções específicas (acúleos, ventosas, etc.) (Galindo *et al.*, 2002; Galindo *et al.*, 2008). Além da base proliferativa, o aumento no tamanho das células e a secreção de material extracelular devem estar envolvidos durante o processo de formação de protoescólices, os quais geralmente são liberados no lúmen do cisto. O desenvolvimento de protoescólices é particularmente interessante devido à capacidade de diferenciação em duas direções desta forma do parasito, dependendo do hospedeiro: (i) se ingerido por um hospedeiro definitivo, um protoescólex se desenvolverá em um verme adulto; (ii) se a ruptura de um cisto no hospedeiro intermediário produzir a disseminação de protoescólices, cada um deles se desdiferenciará, dando origem a novos cistos, chamados de cistos secundários (Zhang *et al.*, 2005). A elucidação das bases moleculares desta plasticidade de desenvolvimento é central para o entendimento da natureza do parasitismo em cestódeos (Thompson & Jenkins, 2014; Thompson & Lymbery,

2013).

1.5 Hidatidose cística

A hidatidose cística, uma zoonose comum em animais de pecuária, também é um problema de saúde pública humana, especialmente em países em desenvolvimento (Mandal & Mandal, 2012). Ela é considerada uma das 17 doenças negligenciadas prioritizadas pela *World Health Organization* (WHO) (http://www.who.int/neglected_diseases/diseases/en/). A distribuição global da hidatidose mudou pouco desde 2010 (Figura 4) e, a América do Sul está entre uma das áreas consideradas altamente endêmicas. Os custos anuais associados com a hidatidose cística são estimados em US\$ 3 bilhões, incluindo custos para tratamento de casos e perdas na produção pecuária (Budke *et al.*, 2006). Em regiões nas quais a hidatidose é considerada endêmica, a incidência em humanos pode exceder 50 casos por 100.000 pessoas-ano (Mandal & Mandal, 2012). Na pecuária, a taxa de hidatidose encontrada em abatedouros em áreas hiperendêmicas da América Latina variam de 20 a 95% dos animais abatidos (Moro *et al.*, 2011).

O diagnóstico precoce da hidatidose cística é difícil, uma vez que a doença é tipicamente assintomática por um longo período depois da infecção (Moro & Schantz, 2009). O diagnóstico é geralmente baseado em dados epidemiológicos e na combinação de métodos radiológicos e imunodiagnóstico. Apesar de limitações associadas à disponibilidade restrita de antígenos melhor caracterizados, os métodos imunológicos ainda são as alternativas mais eficientes para a detecção da infecção pelo estágio larval (Siracusano *et al.*, 2012a). Além disso, os testes sorológicos oferecem vantagens em relação aos radiológicos, incluindo o diagnóstico e tratamento precoces, os quais poderiam aumentar a eficiência da quimioterapia (Zhang *et al.*, 2012). Recentemente, a purificação do antígeno 5 (Ag5) nativo e a utilização em imunoenaios demonstrou-se como um método imunodiagnóstico promissor (Pagnozzi *et al.*, 2014). A caracterização limitada do repertório antigênico do parasito também limita o desenvolvimento de vacinas, embora algumas proteínas com potencial para utilização em vacinação contra a hidatidose cística ou contra a hidatidose alveolar já tenham sido identificadas (Alvarez Rojas *et al.*, 2013; Larrieu *et al.*, 2013).

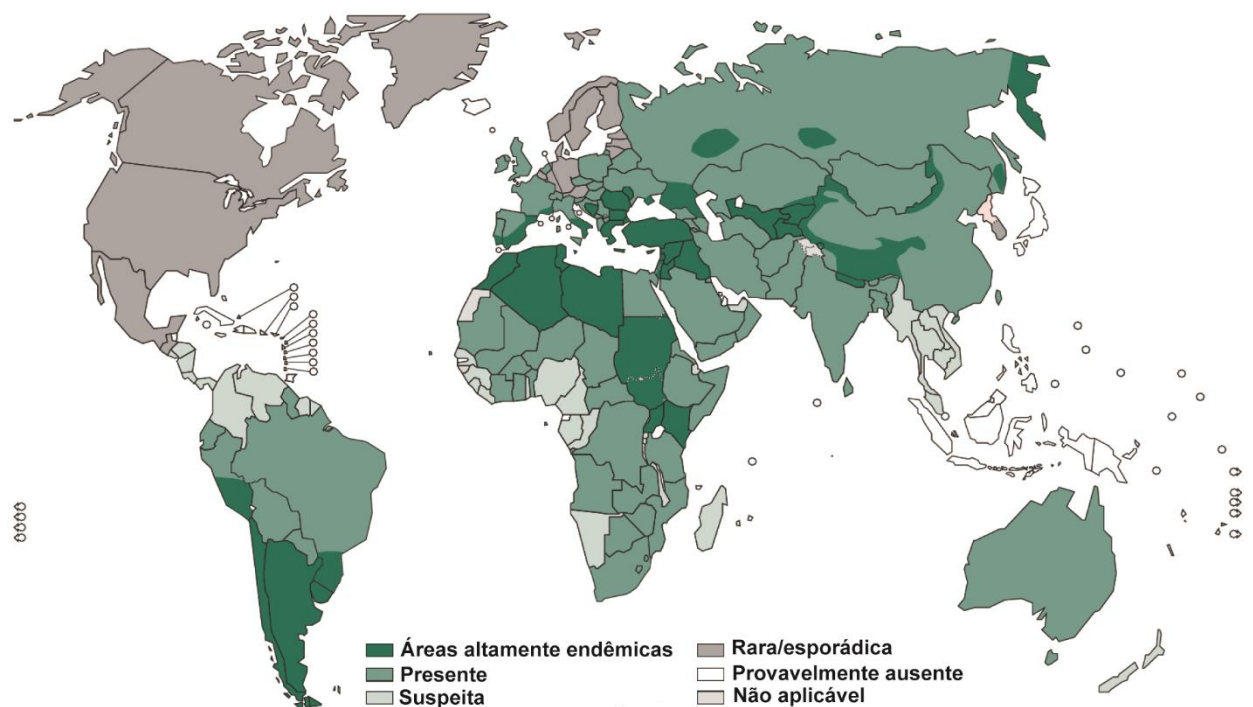


Figure 4. Distribuição geográfica de *E. granulosus* e da hidatidose cística (WHO 2012, adaptado).

1.6 Interações *Echinococcus*-hospedeiro

O cisto hidático é provavelmente a maior estrutura parasitária alojada em tecidos de mamíferos e a sua localização e imunogenicidade resultam num grande potencial para desencadeamento de uma resposta inflamatória por parte do hospedeiro (Siracusano *et al.*, 2012b). Contudo, o cisto hidático é capaz de sobreviver por décadas no corpo do hospedeiro intermediário, devido à secreção e exposição na sua superfície de moléculas que neutralizam ou modulam a resposta imune do hospedeiro (Siracusano *et al.*, 2012a). O repertório global de proteínas expressas no estágio larval patogênico de *E. granulosus* tem sido revelado por vários estudos proteômicos tanto do nosso grupo como de outros (Aziz *et al.*, 2011; Chemale *et al.*, 2003; Cui *et al.*, 2013; Monteiro *et al.*, 2010). Porém, a caracterização mais aprofundada de proteínas ainda é bastante focada em proteínas imunodominantes (Siracusano *et al.*, 2008; Zhang *et al.*, 2012) como o antígeno B (AgB) (Jiang *et al.*, 2012; Monteiro *et al.*, 2012) e o antígeno 5 (Ag5) (Li *et al.*, 2012; Pagnozzi *et al.*, 2014), os quais estão entre os componentes imunodominantes presentes no líquido hidático, onde são encontrados os produtos de ES do parasito (Aziz *et al.*, 2011; Monteiro *et al.*, 2010; Virginio *et al.*, 2012).

Além de proteínas imunodominantes, várias outras têm sido identificadas nos componentes de interação de *E. granulosus* com o hospedeiro (no líquido hidático e na camada germinativa) (Aziz *et al.*, 2011; Monteiro *et al.*, 2010). Surpreendentemente, diversas proteínas intracelulares, como a tiorredoxina-peroxidase, a actina, a 14-3-3, a FBA e a enolase tem sido repetidamente detectadas como produtos de ES da forma larval do parasito. Essas proteínas citosólicas foram inicialmente consideradas como antígenos “ocultos”, mas a detecção delas em líquido hidático indica que elas estão acessíveis para potencialmente induzir ou modular a resposta imune do hospedeiro.

2. JUSTIFICATIVAS

A hidatidose cística é uma zoonose negligenciada, mundialmente distribuída, e responsável por uma alta morbidade e mortalidade (Brunetti & White, 2012). A plasticidade de desenvolvimento exibida por parasitos do gênero *Echinococcus* é certamente uma das características relevantes para a interação com diferentes hospedeiros. Apesar disso, mecanismos moleculares envolvidos no controle da plasticidade destes parasitos, são ainda pouco compreendidos. Alguns esforços têm sido feitos com intuito de melhor compreender tanto aspectos básicos da biologia deste parasito como aspectos intimamente relacionados ao contexto parasito-hospedeiro. Além de estudos transcritômicos do estágio larval do *E. granulosus* (Pan *et al.*, 2014; Parkinson *et al.*, 2012), cabe destacar o recente sequenciamento dos genomas tanto do *E. granulosus* como de outros 3 cestódeos parasitas (Tsai *et al.*, 2013; Zheng *et al.*, 2013). Porém, apesar dos estudos genômicos e transcritômicos revelarem vários aspectos relevantes para a biologia do parasito, muitos aspectos ainda precisam ser melhor explorados. Estudos proteômicos do *E. granulosus*, por exemplo, são de grande valia tanto para confirmar dados preditos pelos estudos genômicos como para revelar novos aspectos da biologia do parasito não acessíveis em nível genômico ou transcritômico. Neste sentido, o nosso grupo tem se dedicado à análise proteômica do estágio larval de *E. granulosus* (Chemale *et al.*, 2003; Monteiro *et al.*, 2010), no contexto da infecção do hospedeiro intermediário. No que diz respeito a produtos de ES, tanto nosso grupo como outros têm identificado proteínas importantes para a relação parasito-hospedeiro (Aziz *et al.*, 2011; Virginio *et al.*, 2012). Embora a identificação de proteínas do parasito tenha contribuído para entender alguns aspectos da biologia do parasito, a identificação não tem sido suficiente para revelar mecanismos moleculares envolvidos no controle de processos biológicos destes parasitos. Deste modo, o estudo de aspectos responsáveis pela geração de variabilidade proteica, como a presença de proteínas *moonlighting* e de MCTs e MPTs constituem uma abordagem importante para revelar novos mecanismos moleculares relevantes para a biologia destes parasitos.

3. OBJETIVOS

A caracterização de proteoformas expressas no estágio larval de *E. granulosus*, incluindo proteoformas potencialmente multifuncionais da EgFBA e EgEno, pode contribuir para um melhor entendimento da biologia do parasito. Esse estudo representa os primeiros passos na caracterização mais aprofundada de proteínas deste parasito, e destaca a importância da caracterização de proteoformas para elucidação de mecanismos moleculares envolvidos no controle de processos biológicos, no estabelecimento e desenvolvimento da fase larval de *E. granulosus*.

Os objetivos específicos deste trabalho são:

1. Investigar possíveis mecanismos associados à multifuncionalidade da EgFBA e da EgEno;
2. Caracterizar MCTs e MPTs de proteoformas expressas no estágio larval de *E. granulosus*;
3. Identificar proteoformas expressas em frações subcelulares do estágio larval de *E. granulosus*;

4. MATERIAIS E MÉTODOS E RESULTADOS

Na seção 4.1, métodos e resultados referentes à predição do genes de EgFBA e EgEno, à análise do padrão de expressão nos diferentes componentes do cisto da EgFBA1 e da EgEno1, à modelagem 3D e a um ensaio de interação proteína-proteína da EgFBA1, são apresentados na forma de artigo publicado.

Na seção 4.2, métodos e resultados referentes à análise filogenética das isoformas de EgFBA e EgEno, aos ensaios enzimáticos e aos ensaios de interação proteína-proteína (sendo os dois últimos realizados com as proteínas rEgFBA1 e rEgEno1), são complementares aos apresentados na seção 4.1.

Na seção 4.3, métodos e resultados referentes às análises de proteômica *top down* e *bottom up* de frações subcelulares de protoescólices de *E. granulosus*, contendo proteínas de baixa massa (< 30 kDa), são apresentados na forma de artigo.

Na seção 4.4, métodos e resultados referentes às análises de proteômica *top down* e *bottom up* de frações subcelulares de protoescólices de *E. granulosus*, contendo proteínas de maior massa (> 30 kDa), são complementares aos apresentados na seção 4.3.

4.1 Frutose-bifosfato-aldolase e enolase de *Echinococcus granulosus*: genes, padrões de expressão e interações proteicas de duas potenciais proteínas *moonlighting*

Referência do artigo original publicado:

Karina Rodrigues Lorenzatto, Karina Mariante Monteiro, Rodolfo Paredes, Gabriela Prado Paludo, Marbella Maria da Fonsêca, Norbel Galanti, Arnaldo Zaha, Henrique Bunselmeyer Ferreira

Fructose-bisphosphate aldolase and enolase from Echinococcus granulosus: genes, expression patterns and protein interactions of two potential moonlighting proteins. Gene, 506: 76-84, 2012.

Contribuição dos autores:

K.R.L: delineamento experimental, execução de todos os experimentos (exceto a modelagem molecular), e redação do manuscrito; K.M.M, N.G, A.Z., H.B.F.: delineamento experimental, análise e discussão de resultados e revisão do manuscrito; R.P.: auxílio nos experimentos de imunofluorescência e revisão do manuscrito; G.P.P: auxílio nos experimentos de imunoblot e *cross-linking* de proteínas; M. M. F: realizou a modelagem molecular.



Fructose-bisphosphate aldolase and enolase from *Echinococcus granulosus*: Genes, expression patterns and protein interactions of two potential moonlighting proteins

Karina Rodrigues Lorenzatto ^a, Karina Mariante Monteiro ^a, Rodolfo Paredes ^c, Gabriela Prado Paludo ^a, Marbella Maria da Fonsêca ^d, Norbel Galanti ^e, Arnaldo Zaha ^{a,b}, Henrique Bunselmeyer Ferreira ^{a,b,*}

^a Laboratório de Genômica Estrutural e Funcional and Laboratório de Biologia Molecular de Cestódeos, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

^b Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências, UFRGS, Porto Alegre, RS, Brazil

^c Escuela de Medicina Veterinaria, Facultad de Ecología y Recursos Naturales, Universidad Andrés Bello, Santiago, Chile

^d Laboratório Nacional de Computação Científica, Petrópolis, RJ, Brazil

^e Programa de Biología Celular y Molecular, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Santiago, Chile

ARTICLE INFO

Article history:

Accepted 17 June 2012

Available online 28 June 2012

Keywords:

Glycolytic enzymes

Moonlighting functions

Cestode

Host–parasite interplay

ABSTRACT

Glycolytic enzymes, such as fructose-bisphosphate aldolase (FBA) and enolase, have been described as complex multifunctional proteins that may perform non-glycolytic moonlighting functions, but little is known about such functions, especially in parasites. We have carried out *in silico* genomic searches in order to identify FBA and enolase coding sequences in *Echinococcus granulosus*, the causative agent of cystic hydatid disease. Four FBA genes and 3 enolase genes were found, and their sequences and exon–intron structures were characterized and compared to those of their orthologs in *Echinococcus multilocularis*, the causative agent of alveolar hydatid disease. To gather evidence of possible non-glycolytic functions, the expression profile of FBA and enolase isoforms detected in the *E. granulosus* pathogenic larval form (hydatid cyst) (EgFBA1 and EgEno1) was assessed. Using specific antibodies, EgFBA1 and EgEno1 were detected in protoscolex and germinal layer cells, as expected, but they were also found in the hydatid fluid, which contains parasite's excretory–secretory (ES) products. Besides, both proteins were found in protoscolex tegument and *in vitro* ES products, further suggesting possible non-glycolytic functions in the host–parasite interface. EgFBA1 modeled 3D structure predicted a F-actin binding site, and the ability of EgFBA1 to bind actin was confirmed experimentally, which was taken as an additional evidence of FBA multifunctionality in *E. granulosus*. Overall, our results represent the first experimental evidences of alternative functions performed by glycolytic enzymes in *E. granulosus* and provide relevant information for the understanding of their roles in host–parasite interplay.

© 2012 Elsevier B.V. Open access under the [Elsevier OA license](http://creativecommons.org/licenses/by/3.0/).

1. Introduction

The larval stage (hydatid cyst or metacestode) of *Echinococcus granulosus* is the causative agent of cystic hydatid disease, a worldwide zoonosis of important medical and economic impacts (Budke et al.,

2006; Moro and Schantz, 2009). The hydatid cyst is the largest parasite structure that lodges within mammalian tissues, and its location and immunogenicity result in a large potential for eliciting inflammation (Siracusano et al., 2008). Nevertheless, the hydatid cyst is able to survive decades in the intermediate host, secreting and exposing molecules that downregulate host responses against the parasite.

The repertoire of known proteins from the hydatid cyst was initially restricted to immunodominant antigens (Carmena et al., 2006; Zhang and McManus, 2006), but, more recently, proteomic surveys have significantly increased the number of *E. granulosus* proteins identified in cyst components (Aziz et al., 2011; Monteiro et al., 2010). Both, in antigen characterization studies and in proteomic surveys, many cytosolic proteins, including glycolytic enzymes, were also consistently found in extracellular host-interacting components (e.g. tegument, and hydatid fluid) (Aziz et al., 2011; Chemale et al., 2003; Monteiro et al., 2010; Rodrigues et al., 1993), and the ectopic presence of these proteins was taken as evidence of potential moonlight functions.

Moonlighting proteins refer to a subset of multifunctional proteins in which two or more different functions are performed by one poly

Abbreviations: cDNA, DNA complementary to RNA; DAPI, 4',6-diamidino-2-phenylindole; EgEno, *E. granulosus* enolase; EgFBA, *E. granulosus* FBA; ELISA, enzyme-linked immunosorbent assay; ES, excretory–secretory products; EST, expressed sequence tag; FBA, fructose-bisphosphate aldolase; Fru 1,6-P2, fructose 1,6-bisphosphate; GST, glutathione S-transferase; IgG, immunoglobulin G; IPTG, isopropyl β-D-thiogalactopyranoside; LC-MS/MS, liquid chromatography–tandem mass spectrometry; ORF, open reading frame; rEgEno1, recombinant *E. granulosus* enolase isoform 1; rEgFBA1, recombinant *E. granulosus* FBA isoform 1; SDS-PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; Sulfo-SBED, sulfosuccinimidyl-2-[6-(biotinamido)-2-(p-azido-benzamido) hexanoamido] ethyl-1,3-diithiopropionate; TEV, tobacco etch virus.

* Corresponding author at: Laboratório de Genômica Estrutural e Funcional, Centro de Biotecnologia, UFRGS, Caixa Postal 15005, 91501-970 Porto Alegre, RS, Brazil. Tel.: +55 51 3308 7768; fax: +55 51 3308 7309.

E-mail address: henrique@cbiot.ufrgs.br (H.B. Ferreira).

peptide chain (Jeffery, 1999). Many of the currently known moonlighting proteins are highly conserved enzymes, and enzymes involved in sugar metabolism particularly appear to moonlight (Huberts and Van Der Klei, 2010). It has been suggested that at least 7 out of the 10 glycolytic enzymes exhibit various non-glycolytic activities (Sriram et al., 2005), whose functions can vary from structural to regulatory, depending on the organism (Gancedo and Flores, 2008; Starnes et al., 2009).

Fructose-bisphosphate aldolase (FBA), enolase and other glycolytic enzymes have been commonly identified in the tegument and in excretory–secretory (ES) products of various parasites, showing antigenic properties and potential for mediating interactions with the host (Bernal et al., 2004; Huang et al., 2009; Marcilla et al., 2007; McCarthy et al., 2002). In parasites, both FBA and enolase have been described as moonlighting proteins, with involvement in key processes, such as motility, adhesion, invasion, differentiation and development (Jewett and Sibley, 2003; Labbé et al., 2006; Pal-Bhowmick et al., 2007; Pomel et al., 2008). These proteins have the ability to bind host macromolecules, such as plasminogen (Bernal et al., 2004; Marcilla et al., 2007), promoting degradation of extracellular matrix components and favoring the pathogen spread in host tissues (Haile et al., 2006; Nogueira et al., 2010). In helminths, the plasminogen-binding ability appears to contribute to parasite establishment in the host (Marcilla et al., 2007), preventing clot formation around the parasites (Ramajo-Hernández et al., 2007).

Here, we surveyed *Echinococcus* genome and transcriptome databases in order to identify FBA and enolase genes and their encoded isoforms. *E. granulosus* FBA isoform 1 (EgFBA1) and enolase isoform 1 (EgEno1) were cloned and expressed in *Escherichia coli*, and antisera raised against the recombinant proteins were used to determine the expression profile of these enzymes in hydatid cyst components. The EgFBA1 3D structure was modeled and a predicted EgFBA1–actin interaction was experimentally confirmed. Possible moonlighting functions of these enzymes in the host–parasite interplay are discussed.

2. Materials and methods

2.1. Parasite material

E. granulosus hydatid cysts (G1 strain, typed according to Bowles et al., 1992) were obtained from livers and lungs of cattle slaughtered in a local abattoir (Cooperleo, São Leopoldo, Brazil), and individual fertile cysts were processed essentially as described by Monteiro et al. (2010). Animal slaughtering was conducted according to Brazilian laws and under supervision of the *Serviço de Inspeção Federal* (Brazilian Sanitary Authority) of the Brazilian *Ministério da Agricultura, Pecuária e Abastecimento*. Condemned viscera at post-mortem meat inspection due to the presence of hydatid cysts were collected at the abattoir and dissected in the laboratory, following protocols approved by the Ethical Committee of the *Universidade Federal do Rio Grande do Sul* (UFRGS). Protoscoleces were collected by aspiration, washed several times in PBS and treated with pepsin 0.1% in Hanks salt solution pH 2.0 for 15 min to eliminate remnants of the germinal layer and dead protoscoleces.

The protoscolex tegument extract was isolated using a freeze/thawing/vortexing method according to Roberts et al. (1983). The protein extraction was performed essentially as described by Mulvenna et al. (2010b), with the isolated teguments successively washed three times in each of the following buffers and temperatures: (i) 40 mM Tris, pH 7.4 at 4 °C; (ii) 5 M urea in 40 mM Tris, pH 7.4 at room temperature; and (iii) 0.1% SDS, 1% Triton X-100 in 40 mM Tris, pH 7.4 at room temperature.

To obtain excretory–secretory (ES) products, protoscoleces were cultivated similarly as described by Virginio et al. (2012), with some modifications. Protoscoleces were cultured in 4 ml of RPMI medium for 48 h, at 37 °C. The culture supernatant was collected and 10-fold

concentrated using an Amicon Ultra-4 3 kDa MWCO centrifugal filter device (Millipore).

2.2. Nucleotide and amino acid sequence analyses

FBA and enolase expression sequence tag (EST) clusters were obtained from the LophoDB Database (<http://nema.cap.ed.ac.uk/Lopho/LophoDB.php>) (Fernández et al., 2002). Related genomic sequences were recovered from *E. granulosus* and *E. multilocularis* genome data available on Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/resources/downloads/helminths/echinococcus-granulosus.html>; <http://www.sanger.ac.uk/resources/downloads/helminths/echinococcus-multilocularis.html>) and on GeneDB (Logan-Klumpler et al., 2012; www.genedb.org). Searches for FBA and enolase genes in the available contigs/supercontigs were performed using the *Echinococcus* Blast Server (<http://www.sanger.ac.uk/cgi-bin/blast/submitblast/Echinococcus>). Recovered contigs and scaffolds (referred at Table S1) were submitted to GeneMark-E (Lomsadze et al., 2005), Fgenesh (Salamov and Solovyev, 2000) and GeneScan (Burge and Karlin, 1998) to determine protein-coding potential and exon–intron gene structure. *In silico* gene predictions were compared to *E. multilocularis* gene-models available on GeneDB (<http://www.genedb.org/Homepage/Emultilocularis>), with *in silico* exon predictions manually adjusted according to *E. multilocularis* gene-models based on RNA evidence.

DNA and protein sequences were aligned using ClustalW2 (Larkin et al., 2007) and Clustal Omega (Sievers et al., 2011), respectively. Alignments were edited using GeneDoc Ver. 2.6.002 (Nicholas and Nicholas, 1997). DNA sequence translations and predictions of protein molecular masses and isoelectric points were performed using tools available on ExpASY website (<http://expasy.org/>).

2.3. RNA extraction, cDNA synthesis and cloning

Total RNA was extracted from *E. granulosus* protoscoleces using the Trizol reagent (Invitrogen). The obtained RNA was treated with DNase I, RNase-free (Fermentas) and reverse transcribed using M-MuLV Reverse Transcriptase (Fermentas) and Oligo(dT)₁₈ Primer (Fermentas). The cDNA synthesis was followed by PCR amplification with High Fidelity DNA Polymerase (Fermentas) using gene-specific primers for EgFBA1 (5'-ATGGCTCGTTTGGTCCCTAC-3' and 5'-GTAGGCGTGGTGGCCAC-3') and EgEno1 (5'-ATGTCCATCTTAAAGATCCA-3' and 5'-CAAAGGATTGCGGAAGT-3'), which were designed according to EgFBA1 and EgEno1 EST clusters. In the first PCR amplification, 27 nt recombination tags FrecI (5'-TCTGAAAACCTGTATTTTCAGGGAGAA-3') and RreI (5'-CGCGGAGGCAGATCGTCAGTCAGTCA-3'), matching pGEX-TEV, were added to the ends of the 5' and 3' gene-specific primers, respectively. Secondary PCR amplification was performed using the primary reaction as template and the following primers: FrecII (5'-TGGTTCGGTGGATCTGAAAACCTGTATTTTCAGGGAGAAATTCGGGGT-3') and RreII (5'-GGTTTTCACCGTCATACCGAAACGCGGAGGCAGATCGTCAGTCAGTCA-3'). The final PCR products were tagged with 50 bp matching pGEX-TEV in their 5' and 3' ends, respectively.

The EgFBA1 and EgEno1 open reading frames (ORFs) (1086 bp and 1299 bp, respectively) were cloned by *in vivo* homologous recombination according to Parrish et al. (2004), with minor modifications. Briefly, the vector pGEX-TEV, a version of pGEX-4-T1 (GE Healthcare) modified in our laboratory to include a cleavage site for tobacco etch virus (TEV) protease (D. M. Vargas, K. M. Monteiro, A. Zaha and H. B. Ferreira, unpublished data), was linearized with XhoI, dephosphorylated with shrimp alkaline phosphatase and purified from agarose gel using GFX Purification Kit (GE Healthcare). Vector DNA and PCR products were mixed (1:1 molar ratio) and transformed into *E. coli* KC8 cells. Transformant colonies were screened for recombinants by colony PCR, and cloned sequences were confirmed by sequencing using the Kit Dyanamic ET Dye Terminator Cycle Sequencing (GE Healthcare) in a MEGABACE 1000 Sequencing System.

2.4. Expression and purification of recombinant proteins

Recombinant protein expression and purification were carried out according to Moitinho-Silva et al. (2012), with some modifications. Briefly, recombinant plasmids carrying the *EgFBA1* and *EgEno1* ORFs were transformed into BL21-CodonPlus-RP and BL21-CodonPlus-Ril (Stratagene). Cells were grown in 1 L of CircleGrow medium (MP Biomedicals) at 37 °C until a OD₆₀₀ of 0.8. Cultures were then induced with 0.1 mM isopropyl beta-D-1-thiogalactopyranoside (IPTG) for 5 h at 20 °C. Glutathione S-transferase (GST)-tagged recombinant proteins were purified by affinity chromatography on Glutathione Sepharose 4B (GE Healthcare) and cleaved with TEV protease for 16 h at 34 °C to be released from the GST moiety. Eluted protein concentration was measured using a Qubit™ quantitation fluorometer and Quant-it™ reagents (Invitrogen) and proper protein purification was verified by SDS-PAGE. The identity of each purified recombinant protein was confirmed by liquid chromatography–tandem mass spectrometry (LC-MS/MS), using a Waters nanoACQUITY UPLC system coupled to a Waters Micromass Q-TOF Micro mass spectrometer (Waters).

2.5. Antisera production and antibody purification

For antisera production, rabbits were immunized by subcutaneous injection with 150 µg of each recombinant protein in complete Freund's adjuvant (Sigma). Immunization was followed by two (rEgFBA1) or three (rEgEno1) boosters of 150 µg protein in incomplete Freund's adjuvant (Sigma) every three weeks. Sera titers were checked by ELISA one week after each injection using the recombinant proteins as antigens.

IgG antibody purification from non-immune control sera and from polyclonal sera from immunized animals was carried out in HiTrap Protein G HP columns (GE Healthcare), according to the manufacturer's protocol.

2.6. Immunoblots

Protein extracts from cyst components (protoscolexes, germinal layer and hydatid fluid), protoscolex tegumental fractions and ES products used in immunoblots were obtained from individual cysts and processed as described in Section 2.1. Protein samples were resolved by 12% SDS-PAGE and transferred to nitrocellulose membranes (Hybond™-ECL™, GE Healthcare) at 15 V for 30 min in a Trans-Blot SD Semi-Dry Transfer System (Bio-Rad). Membranes were processed essentially as described by Monteiro et al. (2010). Rabbit polyclonal sera anti-rEgFBA1 (diluted 1:10,000 v/v) or anti-rEgEno1 (diluted 1:5000 v/v) were used as primary antibodies, and horseradish peroxidase (HRP)-labeled anti-rabbit IgG (ECL™, GE Healthcare) was used as secondary antibody (1:9000 v/v dilution). Antigen–antibody complexes were detected using ECL detection reagent (GE Healthcare) and imaged using a VersaDoc imaging system (Bio-Rad).

2.7. Immunofluorescence

E. granulosus protoscolexes and the cyst wall tissues were fixed, dehydrated and embedded in paraffin as described by Paredes et al. (2007). Sections 5 µm thick were mounted, rehydrated, and blocked with 1% bovine serum albumin, 0.05% Tween in PBS for 1 h at 37 °C. Processed sections were then incubated in a humid chamber for 1 h at 37 °C with purified primary IgG antibodies (Section 2.5) diluted (1:50 v/v in blocking solution), followed by three washes with PBS and incubation in a humid chamber for 1 h at 37 °C with goat anti-rabbit IgG conjugated with Alexa Fluor 488 (Invitrogen) diluted (1:100 v/v in blocking solution). After three additional washes in PBS, sections were incubated with 50 mM DAPI for 20 min at 37 °C, mounted with Fluoromount and observed under the confocal

microscope (Olympus FluoView 1000). Images were digitally captured and processed using the Olympus Fluoview Ver. 2.1c and the Olympus Fluoview Ver. 3.0 Viewer softwares, respectively.

2.8. EgFBA1 modeling

A 3D molecular model of EgFBA1 was built by comparative modeling. Search for templates and generation of molecular models were conducted as described elsewhere (da Fonsêca et al., 2011). Templates used for EgFBA1 modeling were FBAs from *Drosophila melanogaster* (Hester et al., 1991) (PDB code: 1FBA), from rabbit muscle (Blom and Sygusch, 1997) (PDB code: 1ADO) and liver (PDB code: 1FDJ), and from human muscle (Dalby et al., 1999) (PDB code: 4ALD). The alignment between EgFBA1 and the templates was used to generate EgFBA 3D structure using Modeller Ver. 9.9 (Eswar et al., 2007). The final model was selected for evaluation using the following servers: PROCHECK through PDBsum (Laskowski, 2009), and PROSA-web (Wiederstein and Sippl, 2007). Visualization and manipulation of molecular images were performed with Pymol Ver. 1.2 (www.pymol.org).

2.9. Protein cross-linking and mass spectrometry for protein identification

The sulfo-succinimidyl-2-[6-(biotinamido)-2-(p-azido-benzamido) hexanoamido] ethyl-1,3-dithiopropionate (Sulfo-SBED) cross-linker (Pierce) was used for protein cross-linking experiments, according to manufacturer's instructions. Briefly, 1 mg of rEgFBA1 was labeled with a 5-fold molar excess of Sulfo-SBED and Sulfo-SBED-labeled rEgFBA1 was incubated for 30 min with a protoscolex protein extract (Section 2.1) under UV irradiation produced by a Boitton UV lamp (365 nm, 6 W) at a distance of 5 cm. Sulfo-SBED disulfide bonds from rEgFBA1–Sulfo-SBED-interacting protein complexes were cleaved with dithiothreitol and biotin-labeled proteins were recovered by affinity chromatography using a monomeric avidin column (Pierce). The recovered interacting proteins were submitted to trypsin digestion, desalted and analyzed by LC-MS/MS as described by Monteiro et al. (2010). Database searching for protein identifications was performed in local *E. granulosus* and platyhelminths EST databases and in *E. multilocularis* annotated sequences available on GeneDB, using Mascot software Ver. 2.2.1.

3. Results

3.1. Echinococcus FBA and enolase genes and encoded proteins

EST clusters for FBA (EGC00369) and enolase (EGC03002 and EGC04828, corresponding to partial 5' and 3' ORF ends, respectively), available on LophoDB Database, were initially used as search queries to recover homologous sequences in the available *E. granulosus* genome sequences. As a result, four FBA putative paralog genes and three enolase putative paralog genes were found. These FBA and enolase genes were named *EgFBA1-4* and *EgEno 1-3*. Ortholog gene sequences from *E. multilocularis*, recovered from GeneDB, were respectively named *EmFBA1-4* and *EmEno1-3*. Structure of *E. granulosus* and *E. multilocularis* FBA and enolase genes and genomic distribution of these genes in both species are shown in Table 1 and Supplementary Fig. S1 and Table S1.

E. granulosus and *E. multilocularis* orthologs for FBA (e.g. *EgFBA1* and *EmFBA1*) and enolase genes (e.g. *EgEno1* and *EmEno1*) are all well conserved in terms of both exon and intron sequences, size and relative positions, with minor differences in intron size and sequence found only between *EgFBA3* and *EmFBA3*, between *EgFBA4* and *EmFBA4* and between *EgEno3* and *EmEno3* (Table 1 and Supplementary Figs. S1, S2 and S3). Almost all predicted introns display canonical splice sites consistent with the GT-AG rule in *E. granulosus* and *E. multilocularis* FBA and enolase genes. Only three FBA introns showed non-canonical splice

Table 1
Structure and characteristics of *E. granulosus* and *E. multilocularis* putative FBA and enolase genes and their encoded proteins.

Gene	Gene length ^a	ORF ^b	E 1	I 1	E 2	I 2	E 3	I 3	E 4	I 4	E 5	I 5	E 6	I 6	E 7	Protein	Length (aa)	Mass (Da)	pI
<i>EgFBA1</i>	1235	1092	540	76	259	67	293	–	–	–	–	–	–	–	–	EgFBA1 ^c	363	39666.3	8.31
<i>EmFBA1</i>	1235	1092	540	76	259	67	293	–	–	–	–	–	–	–	–	EmFBA1 ^d	363	39667.2	8.03
<i>EgFBA2</i>	1218	1095	543	34	459	89	93	–	–	–	–	–	–	–	–	EgFBA2	364	39967.7	6.27
<i>EmFBA2</i>	1218	1095	543	34	459	89	93	–	–	–	–	–	–	–	–	EmFBA2	364	39967.7	6.27
<i>EgFBA3</i>	2973	1122	327	1033	243	302	259	86	200	430	93	–	–	–	–	EgFBA3	373	41605.8	7.97
<i>EmFBA3</i>	2886	1122	327	993	243	279	259	86	200	406	93	–	–	–	–	EmFBA3	373	41373.5	7.06
<i>EgFBA4</i>	2302	1101	163	39	189	66	197	436	259	481	200	179	93	–	–	EgFBA4	380	41506.4	6.38
<i>EmFBA4</i>	2294	1101	163	33	189	66	197	436	259	479	200	179	93	–	–	EmFBA4	380	41555.5	6.57
<i>EgEno1</i>	1449	1302	310	78	635	69	357	–	–	–	–	–	–	–	–	EgEno1 ^e	433	46561.2	6.48
<i>EmEno1</i>	1449	1302	310	78	653	69	357	–	–	–	–	–	–	–	–	EmEno1	433	46560.2	6.48
<i>EgEno2</i>	1296	1296	–	–	–	–	–	–	–	–	–	–	–	–	–	EgEno2	431	46880.2	5.17
<i>EmEno2</i>	1296	1296	–	–	–	–	–	–	–	–	–	–	–	–	–	EmEno2	431	46771.1	5.16
<i>EgEno3</i>	3319	1317	85	1046	228	244	68	72	289	71	278	339	234	230	135	EgEno3	438	47676.7	6.34
<i>EmEno3</i>	3280	1317	85	1032	228	244	68	72	289	71	278	317	234	227	135	EmEno3	438	47693.8	6.19

All sequence sizes in bp. E: exon; I: intron.

^a Gene length refers to the sum of exons and introns.

^b ORF length refers to the sum of exons.

^c Sequence derived from EST cluster EGC00369 <http://www.nematodes.org/NeglectedGenomes/Lopho/LophDB.php>.

^d EmFBA1 corresponds to the *E. multilocularis* putative fructose-bisphosphate aldolase deposited in GenBank under access number CAC18550.1.

^e EgEno1 corresponds to the *E. granulosus* enolase sequence deposited in GenBank under access number GU080332.

sites GC-AG (intron 2 of *EgFBA2* and *EmFBA2*, and introns 2 and 3 of *EgFBA4* and *EmFBA4*).

EGC00369 EST cluster corresponds to *EgFBA1*, while EGC03002 and EGC04828 EST clusters correspond to *EgEno1*. No ESTs corresponding to the other *EgFBA* and *EgEno* genes have been found, and, therefore, *EgFBA1* and *EgEno1* and their protein products (*EgFBA1* and *EgEno1*, respectively) became the focus of this study. Sizes, molecular masses and isoelectric points of the proteins encoded by the *E. granulosus* and *E. multilocularis* FBA and enolase genes are presented in Table 1.

The alignment of *E. granulosus* and *E. multilocularis* FBA and enolase deduced amino acid sequences (Supplementary Fig. S4) showed high conservation between both paralog and ortholog proteins. Considering *E. granulosus* paralogs, *EgFBA1* presented 72% of amino acid identity (84% similarity) with *EgFBA2*, 51% identity (71% similarity) with *EgFBA3*, and 66% identity (76% similarity) with *EgFBA4* (Supplementary Fig. S4A), while *EgEno1* presented 74% identity (85% similarity) with *EgEno2*, and 66% identity (80% similarity) with *EgEno3* (Supplementary Fig. S4B). The alignments between the *E. granulosus* FBA and enolase deduced protein sequences with their *E. multilocularis* counterparts, in turn, showed levels of identity/similarity between 97%/97% and 100%/100% in the comparisons between the ortholog sequences of these two species. *EgFBA1* and *EgEno1* amino acid sequences were also aligned with ortholog sequences from other eukaryotes (data not shown). *EgFBA1* showed identity/similarity levels of 74%/84% to *Schistosoma mansoni* aldolase, 65%/74% to *Caenorhabditis elegans* aldolase Ce2 and 64%/76% to human aldolase A. *EgEno1*, in turn, showed identity/similarity levels of 74%/88% to *S. mansoni* enolase, 74%/87% to human enolase α , and 74%/85% to *C. elegans* enolase enol-1.

3.2. Recombinant protein expression and purification

EgFBA1 and *EgEno1* ORFs were expressed in *E. coli* as GST-tagged proteins. *E. coli* strains BL21-CodonPlus-RP and BL21-CodonPlus-RIL were tested as hosts, with the first being more suitable for r*EgFBA1* expression, and the second for r*EgEno1* expression (Supplementary Fig. S5). Attempts to express these proteins at 37 °C resulted in reduced solubility (data not shown) and, therefore, protein expression was carried out at 20 °C. Purified r*EgFBA1* and r*EgEno1* presented the expected molecular masses of 39.6 and 46.5 kDa, respectively. About 4 mg of r*EgFBA1* and 3 mg of r*EgEno1* per liter of culture were obtained. The identity of these proteins was confirmed by LC-MS/MS mass spectrometry analysis (data not shown).

3.3. *EgFBA1* and *EgEno1* detection in hydatid cyst components

Rabbit monospecific polyclonal antibodies against *EgFBA1* and *EgEno1* were used in immunoblot experiments to probe samples from different metacestode components. Based on the *E. granulosus* transcriptomic and proteomic data (LophDB database and Monteiro et al., 2010, respectively) *EgFBA1* and *EgEno1* were assumed as the only detectable *EgFBA* and *EgEno* isoforms in the metacestode. This assumption is also corroborated by *E. multilocularis* RNAseq data (Wellcome Trust Sanger Institute, unpublished data), which indicate that *EgFBA1* and *EgEno1* orthologs (*EmFBA1* and *EmEno1*, respectively) correspond to the predominantly expressed isoforms in the metacestode stage.

In the *E. granulosus* metacestode, *EgFBA1* and *EgEno1* presence was demonstrated in protoscolexes, in the germinal layer, and in the hydatid fluid as bands corresponding to the expected molecular masses (of 39.6 and 46.5 kDa, respectively) (Fig. 1). Immunofluorescence assays were used to refine this result in protoscolexes and in the cyst wall, showing that these proteins are widely distributed in all tissues, with confirmation of their expected cytoplasmic localization (Figs. 2D and H, middle and lower panels). Differential expression of *EgFBA1* and *EgEno1* in the cyst wall was also revealed by immunofluorescence labeling. Although the localization of both proteins in the germinal layer is quite similar (Fig. 2H, middle and lower panel), *EgEno1*, but not *EgFBA1*, was detected in the laminated layer (Fig. 2H, lower panel), which separates the cyst germinal layer from the surrounded host-produced adventitial layer.

To further investigate the presence of *EgFBA1* and *EgEno1* in parasite's host-interacting components we also analyzed protoscolex tegument fractions and ES products (Fig. 3). Both *EgFBA1* and *EgEno1* were found in different tegumental fractions (Fig. 3A). Their detection in protein fractions extracted with urea is indicative of their association to cytoskeletal components. Likewise, the detection of *EgEno1* as a faint band in the protein fractions extracted with SDS/Triton X-100 suggests that this protein has some degree of surface association. Additionally, these enzymes were detected as ES products from protoscolexes in culture (Fig. 3B). Overall, these results are indicative that these enzymes are distributed in important host-interaction components.

3.4. *EgFBA1* modeling

In the attempt to find structural features that could be indicative of moonlighting functions, we have modeled *EgFBA1*. To generate

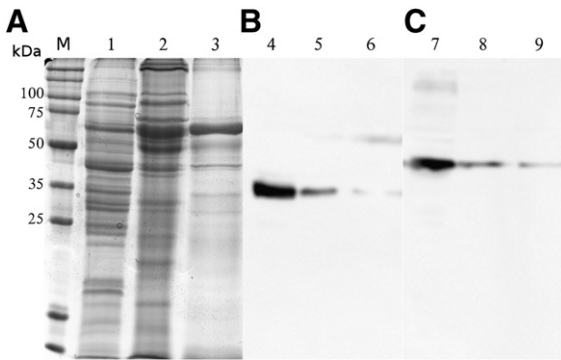


Fig. 1. Detection of EgFBA1 and EgEno1 in *E. granulosus* hydatid cyst components. (A) Parasite protein extracts resolved by 12% SDS-PAGE and stained with Coomassie R-250. Immunoblots were probed with polyclonal serum against EgFBA1 (B) and EgEno1 (C). Molecular mass markers are indicated (kDa). All lanes were loaded with 40 μ g of protein. Lanes 1, 4 and 7: protoscolex extract; lanes 2, 5 and 8: germinal layer extract; lanes 3, 6 and 9: hydatid fluid.

an EgFBA1 3D model, templates with identities ranging from 63% (1FDJ) to 67% (1FBA) were used, and these high identity levels allowed the construction of a model with high confidence. The obtained structure presented a $(\alpha/\beta)_8$ barrel fold architecture (Fig. 4), consistent with those of its templates. The hydrogen bonding between the phosphate groups of the fructose 1,6-bisphosphate (Fru 1,6-P2) substrate to FBA involves the following amino acids Asp34, Ser36, Ser38, Lys108, Lys147, Arg149, Glu188, Glu190, Lys230, Ser272, Arg304 and Leu357 (Dalby et al., 1999), all conserved in EgFBA1. Besides the binding site for Fru 1,6-P2, EgFBA1 also appears to possess a F-actin binding site, since five amino acids previously confirmed to be involved in such interaction (Asp34, Arg43, Lys108, Arg149 and Lys230; Wang et al., 1996) are conserved. Except for Arg43, all these amino acids are also involved in Fru 1,6-P2 binding (see above). Ramachandran plotting showed that in the structure 99% of the EgFBA1 model amino acids lie in favored and additional allowed regions. ProSA-web was used to test for overall model quality and provided a z-score of -10.42 , which falls within the range of values observed for experimentally determined structures of similar lengths.

3.5. Confirmation of predicted EgFBA1 interactions with other proteins

To functionally investigate the EgFBA1 binding properties, we have performed an *in vitro* protein–protein cross-linking interaction assay, using Sulfo-SBED. Mass spectrometry analysis of EgFBA1-interacting proteins recovered from protoscolex extracts allowed the identification of 3 proteins (Table 2). The actin identification indicates the functionality of the *in silico* predicted EgFBA1 actin-binding site (see Section 3.4).

4. Discussion

The glycolytic pathway and its enzymes, which convert glucose to pyruvic acid using the oxidative potential of NAD^+ , are among the most ancient molecular metabolic networks (Canback et al., 2002). Several reports have shown that these enzymes perform a number of functions in addition to their innate glycolytic function and play important roles in several biological and pathophysiological processes (Kim and Dang, 2005; Sirover, 1999; Sriram et al., 2005). FBA and

enolase are examples of multifunctional proteins and they have been described performing moonlighting functions in different biological contexts, including parasitism (Ramajo-Hernández et al., 2007; Tunio et al., 2010).

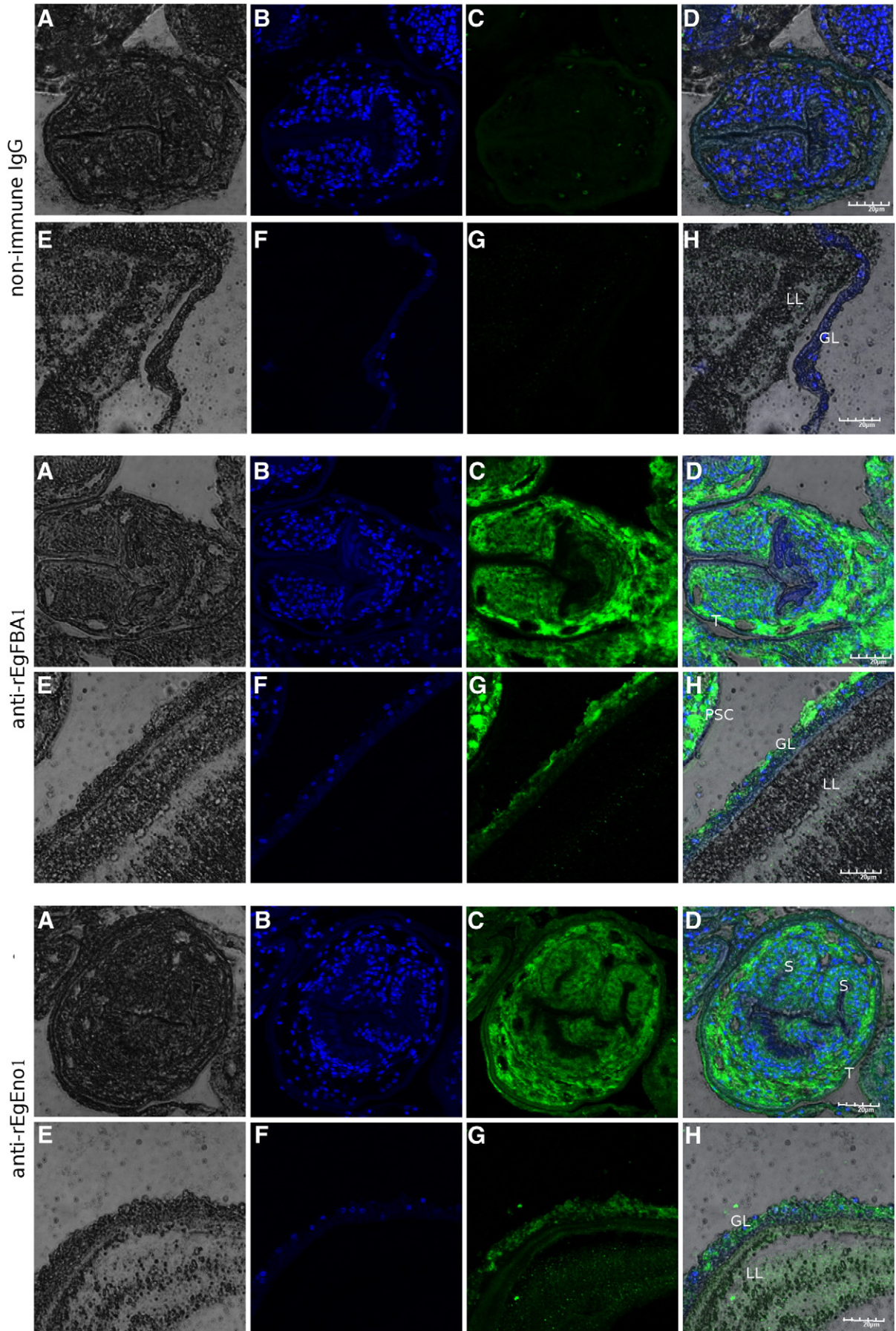
Aldolase primary function is the catalysis of the reversible cleavage of Fru 1,6-P2 to dihydroxyacetone phosphate and glyceraldehyde 3-phosphate. Its functional form is either homo- or hetero-tetrameric. Evidence for EgFBA1 oligomerization was provided by our protein cross-linking assay, which identified this enzyme isoform as one of its ligands. In vertebrates, there are three isozymes (encoded by different genes) that have different activities and vary in their tissue distribution: aldolase A (predominantly present in muscle), aldolase B (in liver) and aldolase C (in brain) (Kukita et al., 1988; Shiokawa et al., 2002). In invertebrates, the presence of more than one aldolase isozyme has also been described. In *D. melanogaster* there are three isozymic forms of aldolase which are generated from a single gene by alternative splicing (Kai et al., 1992). *C. elegans* and *Clonorchis sinensis* present two and three isozymes, respectively (Cho et al., 2006; Inoue et al., 1997), and similarly to vertebrates, each of these isozymes is encoded by different genes.

Enolase primary function is the catalysis of the reversible transformation between 2-phospho-glyceric acid and phosphoenolpyruvic acid. In vertebrates, there are three genes encoding three enolase isoenzymes (α , β and γ), which usually function as homodimers and are expressed in different tissues (McAleese et al., 1988). In invertebrates, there are evidences for the presence of more than one enolase gene with differential regulation. For instance, two enolase genes have been reported for the silkworm *Antheraea pernyi* (Liu et al., 2010) and for *Toxoplasma gondii* (Dzierszinski et al., 2001). In *S. japonicum*, southern blot analysis suggested that this parasite also contains more than one enolase gene (Waine et al., 1993).

In *E. granulosus*, our *in silico* analysis indicates the presence of four FBA genes and three enolase genes. *EgFBA2*, *EgFBA3*, *EgFBA4*, *EgEno2* and *EgEno3* are here being reported for the first time, since previous transcriptional (Fernández et al., 2002) and proteomic (Monteiro et al., 2010) studies of hydatid cyst samples have only detected *EgFBA1* and *EgEno1* products, and no genomic surveys for *E. granulosus* FBA and enolase encoding genes have been conducted so far. *EgFBA2–4* and *EgEno2–3* are structurally integrous, at least regarding their exon–intron structures, and, therefore, are likely functional, assuming that previous failing in the detection of their RNA or protein products could be explained by their possible expression solely or preferentially in stages other than the larval stage, in which only *EgFBA1* and *EgEno1* products were detected. Further studies with *E. granulosus* oncospheres or adult worms will be necessary to address this issue, and, for the moment, only *EgFBA1* and *EgEno1*, expressed in the parasite's pathogenic larval form, including the host interface, will be further discussed here.

As a first approach to gather evidence of possible moonlighting functions for EgFBA1 and EgEno1, we immunolocalized these proteins in components of the parasite larval stage. The used antibodies were produced against recombinant versions of EgFBA1 and EgEno1, the only EgFBA and EgEno isoforms whose expression was so far detected in *E. granulosus* metacystode material (Fernández et al., 2002; Monteiro et al., 2010). In the cyst wall, the distribution of EgEno1 calls attention due to its presence not only in germinal layer, but also in laminated layer. The laminated layer is an acellular, carbohydrate-rich sheath secreted by the germinal layer, and was evolutionarily designed for maintaining the physical integrity of metacystodes and for protecting germinal layer cells from host

Fig. 2. Immunolocalization of EgFBA1 and EgEno1 in *E. granulosus* protoscolexes (A–D) and cyst wall (E–H). Parasite sections (5 μ m) were incubated with purified non-immune IgG (1:50 v/v) (upper panel), purified IgG anti-rEgFBA1 (1:50 v/v) (middle panel) and purified IgG anti-rEgEno1 (1:50 v/v) (lower panel). Recognition of immune complexes was achieved using Alexa 488-conjugated secondary antibodies. A and E: bright field images; B and F: DAPI nuclei staining; C and G: antibody staining; D and H: bright field merges of DAPI and antibody staining. GL: germinal layer; LL: laminated layer; PSC: protoscolex; S: sucker; T: tegument. Scale bar: 20 μ m.



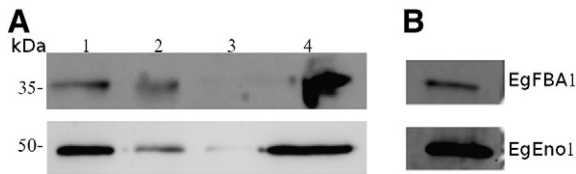


Fig. 3. Detection of EgFBA1 and EgEno1 in protoscolex tegument and ES products. Immunoblots were probed with polyclonal serum against EgFBA1 or EgEno1, as indicated. (A) Immunoblots of tegumental proteins extracted according to their solubility (20 μ g of protein/lane). Lanes 1, 2 and 3: tegumental fractions extracted with 40 mM Tris, 5 M urea and 0.1% SDS/1% Triton X-100, respectively; lane 4: non-tegumental fraction. (B) Immunoblots of culture supernatants containing protoscolex *in vitro* ES products (1 μ g protein/lane). Molecular masses are indicated on the left side of the figure.

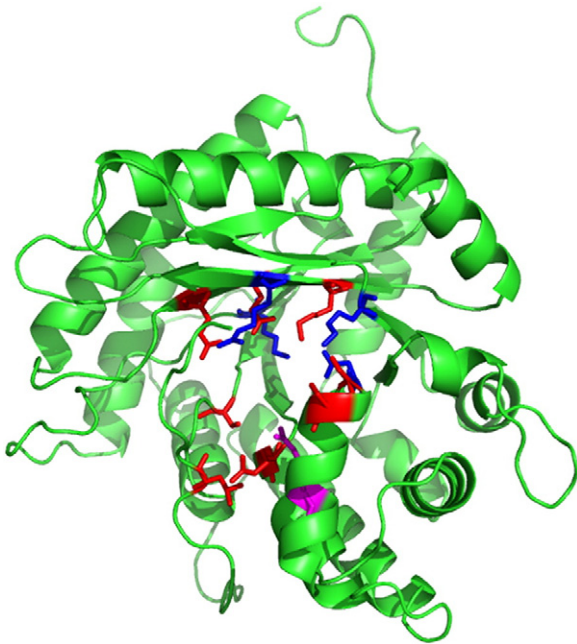


Fig. 4. 3D structure of EgFBA1. Highlighted in green are α -helices, β -sheet and loops. Amino acids involved in Fru 1,6-P2 binding (Asp34, Ser36, Ser38, Lys108, Lys147, Arg149, Glu188, Glu190, Lys230, Ser272, Arg304 and Leu357) are shown in red; amino acids proposed to interact with F-actin (Asp34, Arg43, Lys108, Arg149 and Lys230) are shown in magenta; and amino acids involved in interactions with both Fru 1,6-P2 and actin (Asp34, Lys108, Arg149 and Lys230) are shown in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

immunity (Díaz et al., 2011). Therefore, the detection of EgEno1 in the laminated layer is indicative of its direct involvement with survival mechanisms in the parasite–host interface.

In protoscolexes, immunostaining revealed a wide distribution of both EgFBA1 and EgEno1, which is compatible with their primary function as glycolytic enzymes in the cytoplasm. However, in addition to EgEno1 cytoplasmic localization, its detection in the protoscolex tegumental fraction extracted with detergent confirms its association with surface membranes. It is known that proteins that are expressed or secreted on the surface of pathogens are potential players in host–parasite molecular crosstalk (Vivona et al., 2008). Although protoscolexes are not directly exposed to the intermediate host in the hydatid cyst, their surface proteins are accessible to interaction with host molecules that permeate the cyst and can become directly accessible for interactions with host tissues upon cyst rupture and consequent cyst content leakage. In this later scenario, protoscolex tegumental proteins are assumed to be important for the recognition of host molecules and mediate the activation of signaling pathways

Table 2

Proteins identified as EgFBA1-binding in the cross-linking assay by searches in *E. granulosus* and *E. multilocularis* databases.^a

Protein	Accession number	Organism	MASCOT score ^b	Sequence coverage (%)
1 Actin	AAC37175	<i>E. granulosus</i>	23	2
	EmW_000061200	<i>E. multilocularis</i>	29	4
2 Fructose-bisphosphate aldolase	EGC00369	<i>E. granulosus</i>	106	25
	EmW_000905600	<i>E. multilocularis</i>	103	15
3 Phosphoenolpyruvate carboxykinase	EGC04068	<i>E. granulosus</i>	33	14
	EGC04111	<i>E. granulosus</i>	29	7
	EmW_000292700	<i>E. multilocularis</i>	37	7

^a Searches were performed in local *E. granulosus* and platyhelminths EST databases and in *E. multilocularis* annotated sequences available on GeneDB database.

^b MASCOT score is $-10 \times \log(P)$, where P is the probability that the observed match a random event. Individual scores >22 indicate identity or extensive homology for $p \leq 0.05$.

involved in dedifferentiation of protoscolexes into secondary hydatid cysts (Breijo et al., 1998). Similarly, upon ingestion by the definitive host, tegumental proteins are likely to be important for protoscolex fixation in the small intestine and differentiation into the adult worm. Therefore, the EgEno1 presence at parasite surface favors the interaction with host tissues and molecules, and might have important non-glycolytic functions in these different contexts.

Although mechanisms regulating the presence of the EgEno1 in protoscolex tegument surface are still unknown, this location is evocative of moonlighting functions. Enolase has been found in the tegument of adult *Schistosoma* and *Opisthorchis* worms (Mulvenna et al., 2010a, 2010b), suggesting that this enzyme location in the adult stage could be usual in parasitic flatworms. Another possibility, not mutually exclusive, is that different localizations could be regulated by the presence of substrates. In situations of high metabolic activities, EgFBA1 and EgEno1 would be especially needed in the cytoplasm, but, in situations of reduced metabolic activity, they could be directed to other locations and used for other functions (Tovy et al., 2010). Regarding parasite evolution, this kind of enzyme multifunctionality may have contributed in some degree for biological streamlining associated with adaptation to parasitism (Collingridge et al., 2010).

Equally noticeable is the presence of both EgFBA1 and EgEno1 in ES products. Both proteins have already been identified by proteomic analyses of *E. granulosus* hydatid fluid (Aziz et al., 2011; Monteiro et al., 2010) and *in vitro* ES products of protoscolexes (Virginio et al., 2012). Here, these proteins were detected accordingly, both in the hydatid fluid (containing ES products of the germinal layer and protoscolex) and in the ES products from *in vitro* cultures of protoscolexes. The absence of a typical leader peptide sequence in N-terminal end of EgFBA1 and EgEno1 excludes the involvement of the classical secretory pathway for surface transport. Therefore, the presence of these glycolytic enzymes in the ES products could be explained either by parasite tegument shedding or by a so far unknown transport mechanism. On the other hand, to access whether the parasite/host pair is determinant in the presence of cytoplasmic proteins in the hydatid fluid, other parasite/host combination (G5/cattle and G1/sheep) will be addressed in future studies.

The presence of glycolytic enzymes in the ES products has been previously described in many helminth parasites (Bernal et al., 2004; Guillou et al., 2007; Marcilla et al., 2007; Stadelmann et al., 2010; Wang et al., 2011) and is, therefore, unlikely to have occurred by chance. For example, the presence of *E. multilocularis* phosphoglucose isomerase in the vesicle fluid seems to be related to cellular proliferation and/or differentiation of the germinal layer-associated tissue (Stadelmann et al., 2010). In *C. sinensis*, the inhibition of the enolase in the ES products affects the parasite growth (Wang et al., 2011). Although the exact mechanism responsible for releasing of

these glycolytic enzymes remains a challenging question, the relevance of this process is evident.

Besides EgFBA1 and EgEno1 localization, some structural features additionally support multifunctionality of these proteins. The EgFBA1 conserved F-actin binding site, modeled and experimentally confirmed here, is suggestive of interactions with the cytoskeleton, which is consistent with our protein–protein interaction result where the actin was identified as one of the EgFBA1 partner proteins. The actin binding ability of aldolase has been shown in previous studies for both non-parasite (Wang et al., 1996) and parasite organisms (Starnes et al., 2009), and, in apicomplexan parasites, aldolase was described as a member of a multiprotein complex connecting adhesins to cytoskeleton, providing a model linking adhesion to motility (Jewett and Sibley, 2003; Starnes et al., 2009). For EgEno1, a structural study (Gan et al., 2010) predicted a transmembrane domain. Accordingly, this protein is extracted from the tegumental fraction with detergent. Enolase surface localization has also been described for several parasites, favoring the interaction of enolase with host proteins, like plasminogen (de la Torre-Escudero et al., 2010).

In conclusion, the results presented here showed that, in both *E. granulosus* and *E. multilocularis*, four FBA isoforms and three enolase isoforms were predicted. In *E. granulosus*, only one EgFBA and one EgEno isoform were detected in the pathogenic larval form. EgFBA1 and EgEno1 expression patterns in hydatid cyst components showed that these proteins are exposed in the host–parasite interface, raising evidence of their involvement in molecular signaling pathways important for parasite survival and development. The EgFBA1–actin binding property and EgEno1 surface association, along with their presence in parasite ES products, are indicative of non-glycolytic functions in the host–parasite interplay. We are currently investigating the EgFBA1 and EgEno1 protein interaction networks, in order to understand how the multifunctionality of these glycolytic enzymes is connected to parasite's biology.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gene.2012.06.046>.

Acknowledgments

We thank Dr. Magdalena Zarowiecki (Parasite Genomics Group, Wellcome Trust Sanger Institute), for information on annotation, structure and expression of *E. multilocularis* FBA and enolase genes; Centro de Microscopia Eletrônica (UFRGS), for technical support with the confocal microscopy and Uniprote-MS (Cbiot/UFRGS), for support in LC-MS/MS analyses. This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Convênio Bilateral de Cooperação Internacional CNPq/CONICYT), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (AUX-PE-PARASITOLOGIA – 1278/2011), in Brazil, and by project DI 56-06/R (Universidad Andrés Bello), project ACT112-CONICYT (to N.G.) and FONDECYT, in Chile. K. R. L. is a recipient of a CNPq doctoral fellowship. K. M. M. is recipient of CAPES postdoctoral fellowship. G. P. P. is recipient of a CNPq graduate fellowship.

References

- Aziz, A., Zhang, W., Li, J., Loukas, A., McManus, D.P., Mulvenna, J., 2011. Proteomic characterisation of *Echinococcus granulosus* hydatid cyst fluid from sheep, cattle and humans. *J. Proteomics* 74, 1560–1572.
- Bernal, D., de La Rubia, J.E., Carrasco-Abad, A.M., Toledo, R., Mas-Coma, S., Marcilla, A., 2004. Identification of enolase as a plasminogen-binding protein in excretory–secretory products of *Fasciola hepatica*. *FEBS Lett.* 563, 203–206.
- Blom, N., Sygusch, J., 1997. Product binding and role of the C-terminal region in class I D-fructose 1, 6-bisphosphate aldolase. *Nat. Struct. Biol.* 4, 36–39.
- Bowles, J., Blair, D., McManus, D.P., 1992. Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Mol. Biochem. Parasitol.* 54, 165–173.
- Breijo, M., Spinelli, P., Sim, R., Ferreira, A.M., 1998. *Echinococcus granulosus*: an intraperitoneal diffusion chamber model of secondary infection in mice. *Exp. Parasitol.* 90, 270–276.
- Budke, C.M., Deplazes, P., Torgerson, P.R., 2006. Global socioeconomic impact of cystic echinococcosis. *Emerg. Infect. Dis.* 12, 296–303.
- Burge, C.B., Karlin, S., 1998. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* 8, 346–354.
- Canback, B., Andersson, S.G.E., Kurland, C.G., 2002. The global phylogeny of glycolytic enzymes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6097–6102.
- Carmena, D., Benito, A., Eraso, E., 2006. Antigens for the immunodiagnosis of *Echinococcus granulosus* infection: an update. *Acta Trop.* 98, 74–86.
- Chemale, G., et al., 2003. Proteomic analysis of the larval stage of the parasite *Echinococcus granulosus*: causative agent of cystic hydatid disease. *Proteomics* 3, 1633–1636.
- Cho, P.Y., Lee, M.J., Kim, T.I., Kang, S.Y., Hong, S.J., 2006. Expressed sequence tag analysis of adult *Clonorchis sinensis*, the Chinese liver fluke. *Parasitol. Res.* 99, 602–608.
- Collingridge, P.W., Brown, R.W.B., Ginger, M.L., 2010. Moonlighting enzymes in parasitic protozoa. *Parasitology* 137, 1467–1475.
- da Fonseca, M.M., Zaha, A., Caffarena, E.R., Vasconcelos, A.T.R., 2011. Structure-based functional inference of hypothetical proteins from *Mycoplasma hyopneumoniae*. *J. Mol. Model.* 17, 1–9.
- Dalby, A., Dauter, Z., Littlechild, J.A., 1999. Crystal structure of human muscle aldolase complexed with fructose 1,6-bisphosphate: mechanistic implications. *Protein Sci.* 8, 291–297.
- de la Torre-Escudero, E., Manzano-Román, R., Pérez-Sánchez, R., Siles-Lucas, M., Oleaga, A., 2010. Cloning and characterization of a plasminogen-binding surface-associated enolase from *Schistosoma bovis*. *Vet. Parasitol.* 173, 76–84.
- Díaz, A., Casaravilla, C., Allen, J.E., Sim, R.B., Ferreira, A.M., 2011. Understanding the laminated layer of larval *Echinococcus* II: immunology. *Trends Parasitol.* 27, 264–273.
- Dzierszinski, F., Mortuaire, M., Dendouga, N., Popescu, O., Tomavo, S., 2001. Differential expression of two plant-like enolases with distinct enzymatic and antigenic properties during stage conversion of the protozoan parasite *Toxoplasma gondii*. *J. Mol. Biol.* 309, 1017–1027.
- Eswar, N., et al., 2007. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* 50, 2.9.1–2.9.31.
- Fernández, C., Gregory, W.F., Loke, P., Maizels, R.M., 2002. Full-length-enriched cDNA libraries from *Echinococcus granulosus* contain separate populations of oligo-capped and trans-spliced transcripts and a high level of predicted signal peptide sequences. *Mol. Biochem. Parasitol.* 122, 171–180.
- Gan, W., et al., 2010. Reverse vaccinology approach identify an *Echinococcus granulosus* tegumental membrane protein enolase as vaccine candidate. *Parasitol. Res.* 106, 873–882.
- Gancedo, C., Flores, C.L., 2008. Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.* 72, 197–210.
- Guillou, F., et al., 2007. Excretory–secretory proteome of larval *Schistosoma mansoni* and *Echinostoma caproni*, two parasites of *Biomphalaria glabrata*. *Mol. Biochem. Parasitol.* 155, 45–56.
- Haile, W.B., Coleman, J.L., Benach, J.L., 2006. Reciprocal upregulation of urokinase plasminogen activator and its inhibitor, PAI-2, by *Borrelia burgdorferi* affects bacterial penetration and host-inflammatory response. *Cell. Microbiol.* 8, 1349–1360.
- Hester, G., et al., 1991. The crystal structure of fructose-1,6-bisphosphate aldolase from *Drosophila melanogaster* at 2.5 Å resolution. *FEBS Lett.* 292, 237–242.
- Huang, J., Huang, Y., Wu, X., Du, W., Yu, X., Hu, X., 2009. Identification, expression, characterization, and immunolocalization of lactate dehydrogenase from *Taenia asiatica*. *Parasitol. Res.* 104, 287–293.
- Huberts, D.H.E.W., Van Der Klei, I.J., 2010. Moonlighting proteins: an intriguing mode of multitasking. *Biochim. Biophys. Acta* 1803, 520–525.
- Inoue, T., et al., 1997. *Caenorhabditis elegans* has two isozymic forms, CE-1 and CE-2, of fructose-1,6-bisphosphate aldolase which are encoded by different genes. *Arch. Biochem. Biophys.* 339, 226–234.
- Jeffery, C., 1999. Moonlighting proteins. *TIBS* 24, 8–11.
- Jewett, T.J., Sibley, L.D., 2003. Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. *Mol. Cell* 11, 885–894.
- Kai, T., Sugimoto, Y., Kusakabe, T., Zhang, R., Koga, K., Hori, K., 1992. Gene structure and multiple mRNA species of *Drosophila melanogaster* aldolase generating three isoforms with different enzymatic properties. *J. Biochem.* 112, 677–688.
- Kim, J.W., Dang, C.V., 2005. Multifaceted roles of glycolytic enzymes. *Trends Biochem. Sci.* 30, 142–150.
- Kukita, A., Mukai, T., Miyata, T., Hori, K., 1988. The structure of brain-specific rat aldolase C mRNA and the evolution of aldolase isozyme genes. *Eur. J. Biochem.* 171, 471–478.
- Labbé, M., Péroval, M., Bourdieu, C., Girard-Misguich, F., Péry, P., 2006. *Eimeria tenella* enolase and pyruvate kinase: a likely role in glycolysis and in others functions. *Int. J. Parasitol.* 36, 1443–1452.
- Larkin, M.A., et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Laskowski, R.A., 2009. PDBsum new things. *Nucleic Acids Res.* 37, D355–D359.
- Liu, Y., et al., 2010. cDNA cloning and expression pattern of two enolase genes from the Chinese oak silkworm, *Antheraea pernyi*. *Acta Biochim. Biophys. Sin.* 42, 816–826.
- Logan-Klumpler, F.J., et al., 2012. GeneDB—an annotation database for pathogens. *Nucleic Acids Res.* 40, D98–D108.
- Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y., Borodovsky, M., 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* 33, 6494–6506.
- Marcilla, A., et al., 2007. *Echinostoma caproni*: identification of enolase in excretory/secretory products, molecular cloning, and functional expression. *Exp. Parasitol.* 117, 57–64.
- McAleese, S.M., Dunbar, B., Fothergill, J.E., Hinks, L.J., Day, I.N., 1988. Complete amino acid sequence of the neurone-specific gamma isozyme of enolase (NSE) from human brain and comparison with the non-neuronal alpha form (NNE). *Eur. J. Biochem.* 178, 413–417.
- McCarthy, J.S., et al., 2002. *Onchocerca volvulus* glycolytic enzyme fructose-1, 6-bisphosphate aldolase as a target for a protective immune response in humans. *Infect. Immun.* 70, 851–858.

- Moitinho-Silva, L., et al., 2012. *Mycoplasma hyopneumoniae* type I signal peptidase: expression and evaluation of its diagnostic potential. *Vet. Microbiol.* 154, 282–291.
- Monteiro, K.M., de Carvalho, M.O., Zaha, A., Ferreira, H.B., 2010. Proteomic analysis of the *Echinococcus granulosus* metacestode during infection of its intermediate host. *Proteomics* 10, 1985–1999.
- Moro, P., Schantz, P.M., 2009. Echinococcosis: a review. *Int. J. Infect. Dis.* 13, 125–133.
- Mulvenna, J., et al., 2010a. Exposed proteins of the *Schistosoma japonicum* tegument. *Int. J. Parasitol.* 40, 543–554.
- Mulvenna, J., et al., 2010b. The secreted and surface proteomes of the adult stage of the carcinogenic human liver fluke *Opisthorchis viverrini*. *Proteomics* 10, 1063–1078.
- Nicholas, K.B., Nicholas, H.B., 1997. GeneDoc: A Tool for Editing and Annotating Multiple Sequence Alignments. Distributed by the Author.
- Nogueira, S.V., et al., 2010. *Paracoccidioides brasiliensis* enolase is a surface protein that binds plasminogen and mediates interaction of yeast forms with host cells. *Infect. Immun.* 78, 4040–4050.
- Pal-Bhowmick, I., Vora, H.K., Jarori, G.K., 2007. Sub-cellular localization and post-translational modifications of the *Plasmodium yoelii* enolase suggest moonlighting functions. *Malar. J.* 6, 45.
- Paredes, R., Jiménez, V., Cabrera, G., Iragüen, D., Galanti, N., 2007. Apoptosis as a possible mechanism of infertility in *Echinococcus granulosus* hydatid cysts. *J. Cell. Biochem.* 100, 1200–1209.
- Parrish, J.R., Limjindaporn, T., Hines, J.A., Liu, J., Liu, G., Finley, R.L., 2004. High-throughput cloning of *Campylobacter jejuni* ORFs by *in vivo* recombination in *Escherichia coli*. *J. Proteome Res.* 3, 582–586.
- Pomel, S., Luk, F.C.Y., Beckers, C.J.M., 2008. Host cell egress and invasion induce marked relocations of glycolytic enzymes in *Toxoplasma gondii* tachyzoites. *PLoS Pathog.* 4, e1000188.
- Ramajo-Hernández, A., Pérez-Sánchez, R., Ramajo-Martín, V., Oleaga, A., 2007. *Schistosoma bovis*: plasminogen binding in adults and the identification of plasminogen-binding proteins from the worm tegument. *Exp. Parasitol.* 115, 83–91.
- Roberts, S., MacGregor, A., Vojvodic, M., Wells, E., Crabtree, J., Wilson, R., 1983. Tegument surface membranes of adult *Schistosoma mansoni*: development of a method for their isolation. *Mol. Biochem. Parasitol.* 9, 105–127.
- Rodrigues, J.J., Ferreira, H.B., Zaha, A., 1993. Molecular cloning and characterization of an *Echinococcus granulosus* cDNA encoding malate dehydrogenase. *Mol. Biochem. Parasitol.* 60, 157–160.
- Salamov, A.A., Solovyev, V.V., 2000. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* 10, 516–522.
- Shiokawa, K., Kajita, E., Hara, H., Yatsuki, H., Hori, K., 2002. A developmental biological study of aldolase gene expression in *Xenopus laevis*. *Cell Res.* 12, 85–96.
- Sievers, F., et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
- Siracusano, A., et al., 2008. Immunomodulatory mechanisms during *Echinococcus granulosus* infection. *Exp. Parasitol.* 119, 483–489.
- Sirover, M.A., 1999. New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. *Biochim. Biophys. Acta* 1432, 159–184.
- Sriram, G., Martinez, J.A., McCabe, E.R.B., Liao, J.C., Dipple, K.M., 2005. Single-gene disorders: what role could moonlighting enzymes play? *Am. J. Hum. Genet.* 76, 911–924.
- Stadelmann, B., et al., 2010. *Echinococcus multilocularis* phosphoglucose isomerase (EmPGI): a glycolytic enzyme involved in metacestode growth and parasite–host cell interactions. *Int. J. Parasitol.* 40, 1563–1574.
- Starnes, G.L., Coincon, M., Sygusch, J., Sibley, L.D., 2009. Aldolase is essential for energy production and bridging adhesion–actin cytoskeletal interactions during parasite invasion of host cells. *Cell Host Microbe* 5, 353–364.
- Tovy, A., Siman Tov, R., Gaentzsch, R., Helm, M., Ankri, S., 2010. A new nuclear function of the *Entamoeba histolytica* glycolytic enzyme enolase: the metabolic regulation of cytosine-5 methyltransferase 2 (Dnmt2) activity. *PLoS Pathog.* 6, e1000775.
- Tunio, S.A., Oldfield, N.J., Berry, A., Ala'Aldeen, D.A.A., Wooldridge, K.G., Turner, D.P.J., 2010. The moonlighting protein fructose-1, 6-bisphosphate aldolase of *Neisseria meningitidis*: surface localization and role in host cell adhesion. *Mol. Microbiol.* 76, 605–615.
- Virginio, V.G., et al., 2012. Excretory/secretory products from *in vitro*-cultured *Echinococcus granulosus* protoscoleces. *Mol. Biochem. Parasitol.* 183, 15–22.
- Vivona, S., et al., 2008. Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. *Trends Biotechnol.* 26, 190–200.
- Waine, G.J., Becker, M., Kalinna, B., Yang, W., McManus, D.P., 1993. Cloning and functional expression of a *Schistosoma japonicum* cDNA homologous to the enolase gene family. *Biochem. Biophys. Res. Commun.* 195, 1211–1217.
- Wang, J., Morris, A.J., Tolan, D.R., Pagliaro, L., 1996. The molecular nature of the F-actin binding activity of aldolase revealed with site-directed mutants. *J. Biol. Chem.* 271, 6861–6865.
- Wang, X., et al., 2011. *Clonorchis sinensis* enolase: identification and biochemical characterization of a glycolytic enzyme from excretory/secretory products. *Mol. Biochem. Parasitol.* 177, 135–142.
- Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, 407–410.
- Zhang, W., McManus, D.P., 2006. Recent advances in the immunology and diagnosis of echinococcosis. *FEMS Immunol. Med. Microbiol.* 47, 24–41.

Prediction_Of_Genes

Table S1. Exon predictions of FBA and enolase genes in the *E. granulosus* and *E. multilocularis* genomes.

Gene (GeneDB ID)	Contig or scaffold ^a	Contig Position	Contig length (bp)	Strand	GeneMark		Fgenesh		GeneScan		Gene DB gene-models ^b	
					Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
<i>EgFBA1</i>	_0001	15406054-15407288	15,989,225	+	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	540	1	540	1	540	1	540
					2	259	2	259	2	259	2	259
<i>EgFBA2</i>	_0002	5697328-5698545	14,522,064	+	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	543	1	543	1	543	1	543
					2	474	2	510	2	510	2	459
<i>EgFBA3</i>	_0002	4379060-4382032	14,522,064	-	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	327	1	327	1	327	1	327
					2	243	2	243	2	243	2	243
<i>EgFBA4</i>	_0002	5703849-5706150	14,522,064	+	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	163	1	163	1	163	1	163
					2	116	2	189	2	206	2	189
<i>EmFBA1</i> (<i>EmW_000905600</i>)	_007768	15142612-15143846	15,716,476 Mb	+	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	540	1	540	1	540	1	540
					2	259	2	259	2	259	2	259
<i>EmFBA2</i> (<i>EmW_000993700</i>)	_007780	5440599-5441816	14,156,623 Mb	+	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	543	1	543	1	543	1	543
					2	474	2	489	2	489	2	459
<i>EmFBA3</i> (<i>EmW_000980100</i>)	_007780	4150072-4152957	14,156,623 Mb	-	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					3	93					3	93

Prediction_Of_Genes

					1	327	1	327	1	327	1	327
					2	243	2	243	2	243	2	243
					3	259	3	259	3	259	3	259
					4	200	4	200	4	200	4	200
					5	93	5	93	5	93	5	93
EmFBA4 (EmW_000993800)	_007780	5447003-5449296	14,156,623 Mb	+	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	163	1	163	1	163	1	163
					2	116	2	189	2	206	2	189
					3	228	3	197	3	259	3	197
					4	259	4	259	4	200	4	259
					5	200	5	200	5	93	5	200
					6	93	6	93	6	93	6	93
EgEno1	_0004	8972402-8973850	9,363,953	-	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	310	1	310	1	310	1	310
					2	635	2	635	2	656	2	656
					3	357	3	357	3	357	3	357
EgEno2	_0008	3200967-3202262	4,383,637	-	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	1296	1	1296	1	1296	1	1296
EgEno3	_0007	2028276-2031582	4,654,179	+	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	79	1	79	1	210	1	85
					2	228	2	228	2	68	2	228
					3	68	3	68	3	289	3	68
					4	289	4	289	4	278	4	289
					5	278	5	278	5	234	5	278
					6	234	6	234	6	135	6	234
					7	135	7	135	7	135	7	135
EmEno1 (EmW_000514200)	_007614	8848514-8849962	8,932,634 Mb	+	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	310	1	310	1	310	1	310
					2	635	2	635	2	635	2	635
					3	357	3	357	3	357	3	357
EmEno2 (EmW_000179700)	62302	3112568-3113863	4,325,953 Mb	+	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length
					1	1296	1	1296	1	1296	1	1296
EmEno3 (EmW_000533500)	_007636	2137626-2140905	4,579,842 Mb	-	Exon	Exon length	Exon	Exon length	Exon	Exon length	Exon	Exon length

Prediction_Of_Genes

1	79	1	85	1	210	1	85
2	228	2	210	2	68	2	228
3	68	3	68	3	289	3	68
4	289	4	289	4	278	4	289
5	278	5	278	5	234	5	278
6	234	6	234	6	135	6	234
7	135	7	135			7	135

^a Contig or scaffold numbers according to Echinococcus genome sequences available at Wellcome Trust Sanger Institute and/or GeneDB in april 25, 2012.

^b *E. multilocularis* gene-models available on GeneDB (which are based on RNA evidence) were used for manual adjustments in some *in silico* predicted exons (in bold).

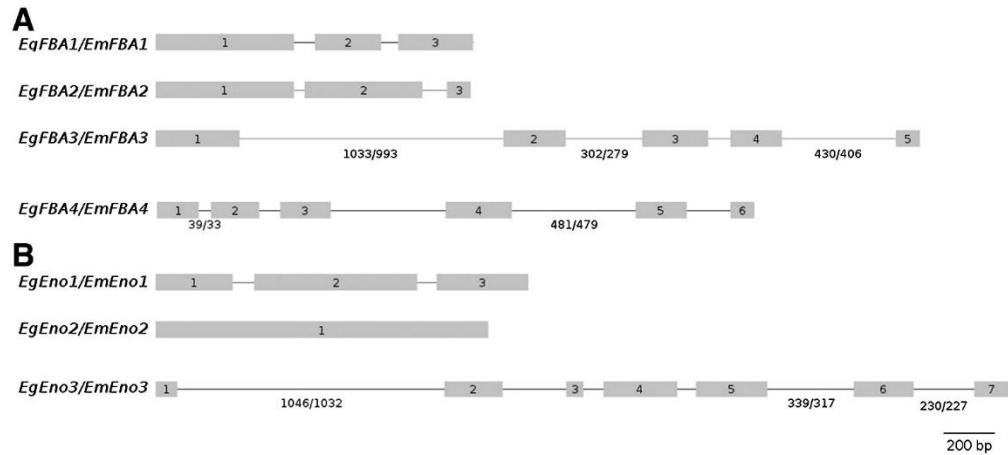


Fig. S1 Overall structure of *E. granulosus* and *E. multilocularis* FBA (A) and enolase (B) genes. Numbered boxes represent exons; lines represent introns. All genes are represented in the same indicated scale. Introns which vary in size between *E. granulosus* and *E. multilocularis* orthologs are indicated with the respective sizes (bp) underneath; in these cases, the lines are representative of *E. granulosus* intron sizes. Physical maps of *E. granulosus* and *E. multilocularis* FBA were all derived from the corresponding genomic sequences found in the sequenced contigs/scaffolds referred in Table S1.

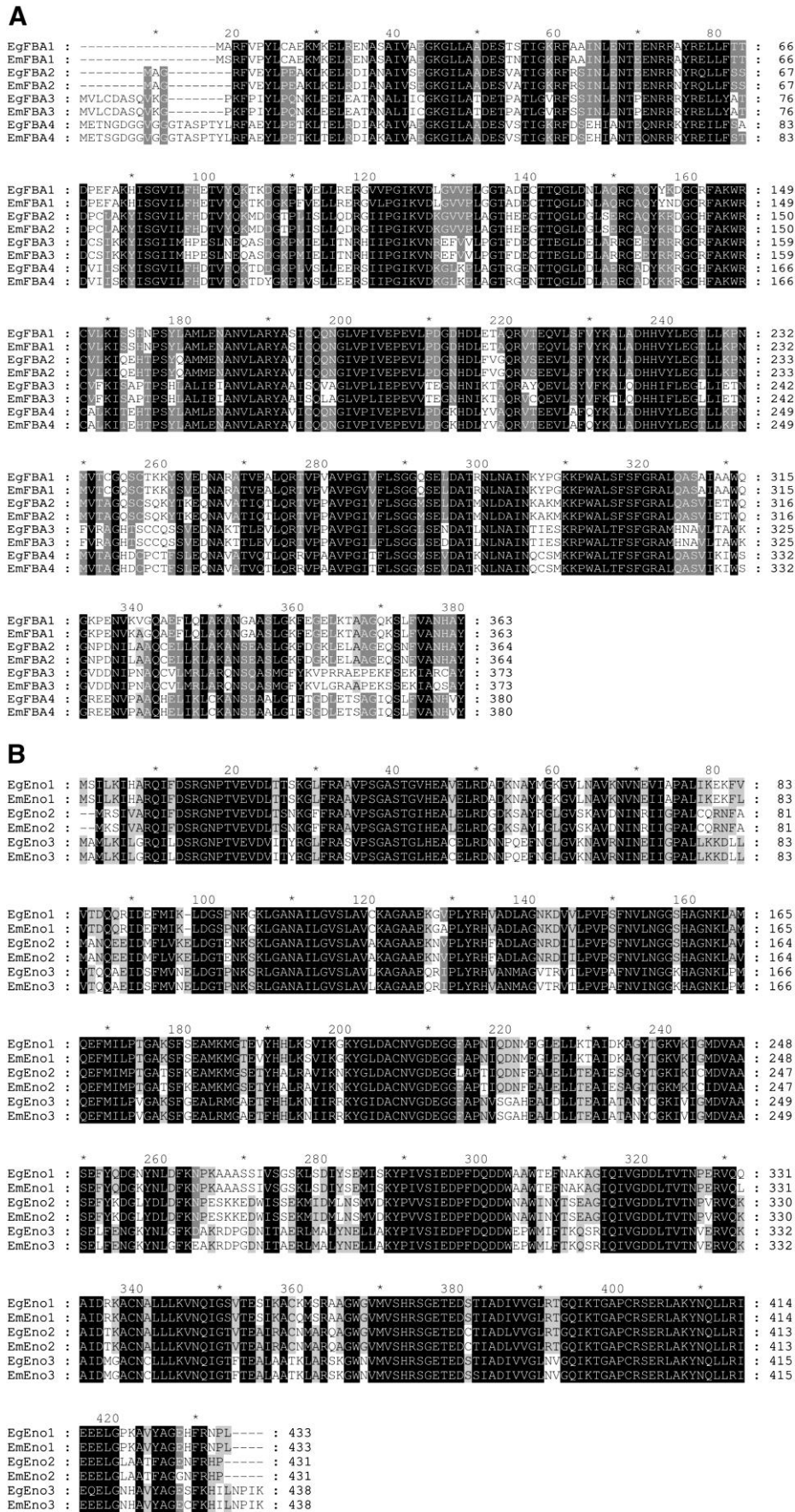


Fig. S4. Alignment of deduced amino acid sequences from *E. granulosus* and *E. multilocularis* FBA (A) and enolase (B) predicted isoforms. For each alignment, amino acids which are conserved in all the aligned sequences are printed in white on a black background; amino acids conserved in at least five of the aligned sequences are printed in white on gray background; and amino acids conserved in four of the aligned sequences are printed in black on gray background. EgFBA and EgEno amino acid sequences were all deduced from the corresponding genomic sequences found in the sequenced contigs/scaffolds referred in Table S1.

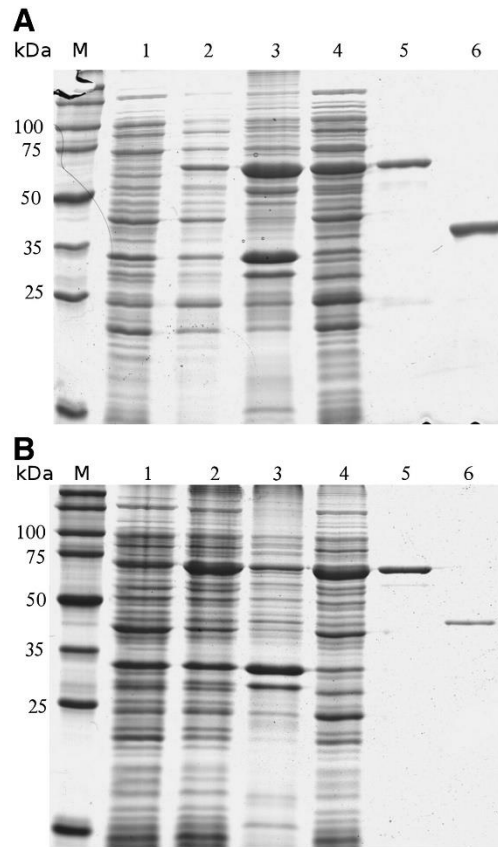


Fig. S5 Expression and purification of the recombinant *E. granulosus* FBA and enolase. Protein samples of rEgFBA1 (A) and rEgEno1 (B) expression and purification steps were resolved by 12% SDS-PAGE and stained with Coomassie R-250. Lane 1: non induced cells; 2: cells induced with IPTG; 3: pellet (insoluble fraction); 4: supernatant (soluble fraction); 5: GST-tagged proteins recovered by affinity purification; 6: recombinant proteins released from GST. Molecular mass markers are indicated (kDa) on the left side of the figures.

4.2 Frutose-bifosfato-aldolase e enolase: potenciais proteínas *moonlighting* de *E. granulosus*

4.2.1 Materiais e métodos

4.2.1.1 Alinhamento de sequências e análises filogenéticas

As sequências aminoacídicas das isoformas de EgFBA e EgEno foram obtidas na etapa de predição dos genes, descrita na seção 4.1. As sequências aminoacídicas das isoformas de FBA e de enolase de outros helmintos e demais organismos foram recuperadas do GeneDB e do UniProtKB, respectivamente. Os alinhamentos foram realizados através do Clustal 2.1 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>), utilizando a matriz de Gonnet. As análises filogenéticas foram realizadas no software MEGA5 (Tamura *et al.*, 2011), através do método de agrupamento de vizinhos (*neighbor joining*), com 1000 replicatas de *bootstrap*.

4.2.1.2 Ensaio enzimáticos

A atividade enzimática da rEgFBA1 e da rEgEno1 (a obtenção das proteínas recombinantes foi descrita na seção 4.1) foi avaliada em espectrofotômetro através do monitoramento da conversão da forma reduzida do NADH para NAD de acordo com (Bergmeyer, 1974), com pequenas alterações. Os ensaios foram realizados a 37°C, em um volume final de 200 µl. Os ensaios com a FBA continham Tris 90 mM, frutose-1,6-bifosfato (Fru-1,6-P2) 1 µM a 1000 µM, β-NADH 0,13 mM, 5 U de gliceraldeído-desidrogenase/triosefosfato-isomerase e 0,5 µg de rEgFBA1. Os ensaios com a enolase continham trietanolamina 81 mM, 2-fosfoglicerato (2-PGA) 0,01 mM a 3,2 mM, β-NADH 0,12 mM, sulfato de magnésio 25 mM, KCl 100 mM, adenosina 5-bifosfato 1,3 mM, 7 U de piruvato quinase, 10 U de lactato desidrogenase e 0,5 µg de rEgEno1. A mudança do NADH foi determinada utilizando o coeficiente de absorção $\epsilon = 6,222 \text{ M}^{-1} \text{ cm}^{-1}$, a 340 nm.

4.2.1.3 Experimentos de *cross-linking*

O ensaio de *cross-linking* utilizando o Sulfo-SBED foi realizado conforme as instruções do fabricante (Thermo Scientific). Brevemente, a rEgFBA1 (1 mg) e a rEgEno1 (1 mg) foram marcadas com um excesso molar de 5 X do Sulfo-SBED em PBS, em gelo

por 2 h e depois estas amostras foram dialisadas 16 h contra PBS para remover o excesso do Sulfo-SBED. A glutationa-S-transferase (GST) recombinante de *Schistosoma japonicum* (do vetor de expressão pGEX-TEV) foi utilizada como controle. A interação proteína-proteína foi realizada por incubação da rEgFBA1 e da rEgEno1 marcadas com Sulfo-SBED com um lisado de protoescólices e o *cross-linking* foi adquirido por irradiação UV durante 30 min, a uma distância de 5 cm. Depois da redução da ponte dissulfeto com DTT, as proteínas marcadas com biotina foram isoladas em uma coluna de avidina e etapas de lavagem foram realizadas com NaCl 100 mM/SDS 0,05% e NaCl 200 mM/SDS 0,05%. Após as lavagens, as proteínas eluídas foram precipitadas durante 16 h em TCA/acetona a -20°C. As proteínas foram ressuspensas em ureia 8 M e digeridas com tripsina a 37°C por 16 h. Os peptídeos foram dessalinizados (Oasis HLB Column, Waters) e ressuspensos em TFA 0,1%. A análise dos peptídeos por LC-MS/MS foi realizada em um Orbitrap-LTQ-Velos. A identificação das proteínas foi realizada através de buscas em bancos de dados, utilizando o software Mascot v. 2.2.1. A classificação funcional das proteínas identificadas foi realizada no BLAST2GO (Conesa *et al.*, 2005) contra o banco de dados NCBI nr. O mapeamento da ontologia gênica (GO) e anotação foram realizados com base nos resultados do Blastp ($E\text{-value} \leq 1.0E-3$).

4.2.2 Resultados

4.2.2.1 Filogenias das isoformas de FBA e de enolase de *E. granulosus*

Análises filogenéticas incluindo as isoformas EgFBA1-4 e EgEno1-3 e isoformas de FBA e de enolase, respectivamente, de outros invertebrados e vertebrados, foram realizadas (Figuras 5 e 6). Com base nas árvores filogenéticas, não foi possível classificar as isoformas destas enzimas conforme a classificação normalmente utilizada para as isoformas de vertebrados (FBA A, B e C e enolase alfa, beta e gama). De modo geral, tanto isoformas de FBA como de enolase de cestódeos e trematódeos estão agrupadas em um ramo separado das isoformas de vertebrados e outros invertebrados. As isoformas de FBA2-4 estão presentes apenas em representantes da família Taeniidae e agrupadas em um ramo separado da isoforma de FBA1 de cestódeos e trematódeos, sendo que o representante de cestódeos pertencente à família Hymenolepididae e as duas espécies representativas de trematódeos apresentam apenas uma isoforma de FBA nos seus

genomas. No caso da enolase, isoformas de Eno2-3 estão presentes apenas em cestódeos e agrupadas em um ramo separado da isoforma de Eno1 de cestódeos e trematódeos, sendo que as espécies representativas de trematódeos novamente apresentam apenas uma isoforma de enolase nos seus genomas.

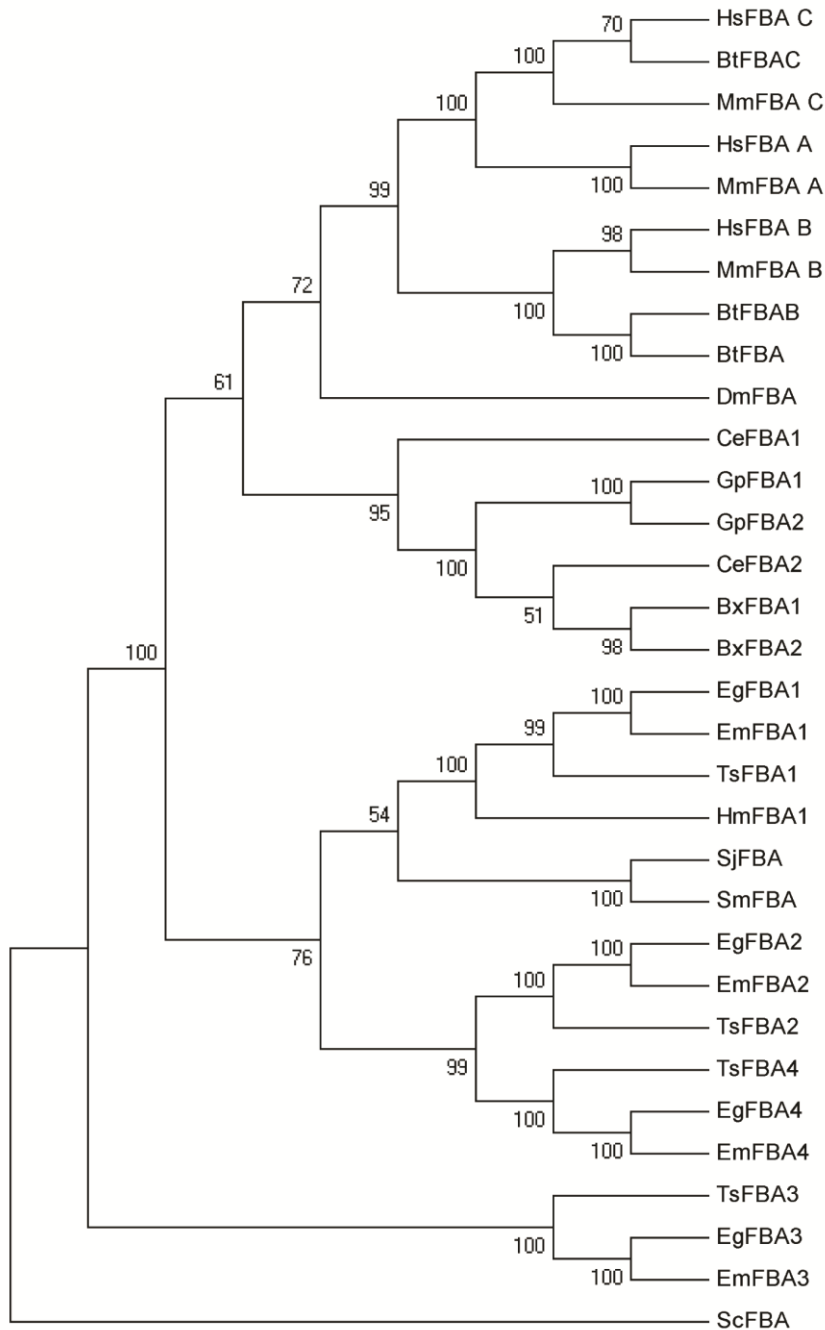


Figure 5. Árvore filogenética de isoformas de FBA de vertebrados e invertebrados. Hs: *Homo sapiens*; Bt: *Bos taurus*; Mm: *Mus musculus*; Dm: *Drosophila melanogaster*; Ce: *Caenorhabditis elegans*; Gp: *Globodera pallida*; Bx: *Bursaphelenchus xylophilus*; Eg: *E. granulosus*; Em: *E. multilocularis*; Ts: *Taenia solium*; Hm: *Hymenolepis microstoma*; Sj: *S. japonicum*; Sm: *Schistosoma mansoni*; Sc: *Saccharomyces cerevisiae*.

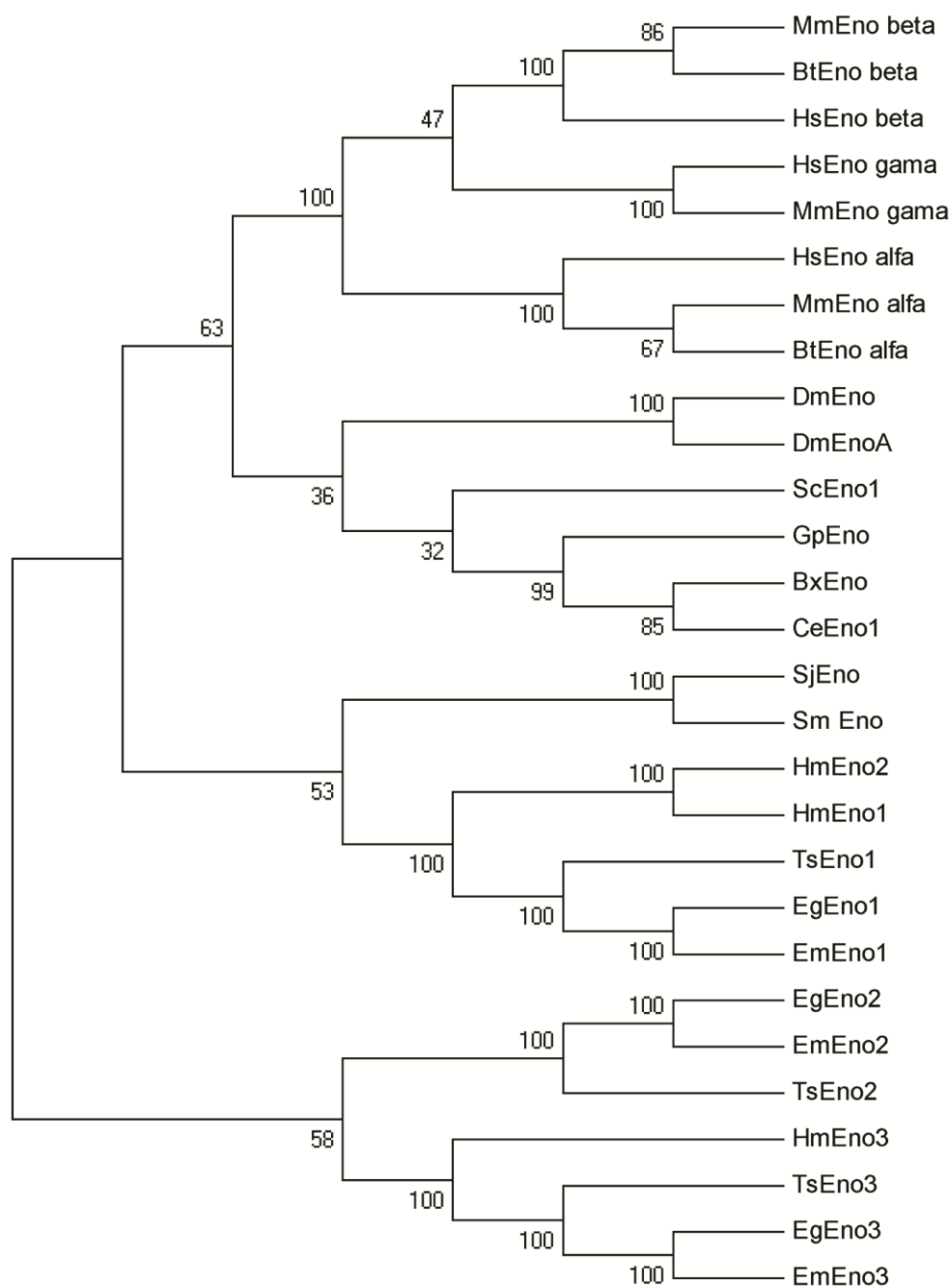


Figure 6. Árvore filogenética de isoformas de enolase de vertebrados e invertebrados. Hs: *Homo sapiens*; Bt: *Bos taurus*; Mm: *Mus musculus*; Dm: *Drosophila melanogaster*; Ce: *Caenorhabditis elegans*; Gp: *Globodera pallida*; Bx: *Bursaphelenchus xylophilus*; Eg: *E. granulosus*; Em: *E. multilocularis*; Ts: *Taenia solium*; Hm: *Hymenolepis microstoma*; Sj: *S. japonicum*; Sm: *Schistosoma mansoni*; Sc: *Saccharomyces cerevisiae*.

4.2.2.2 Atividade enzimática da rEgFBA1 e da rEgEno1

A rEgFBA1 apresentou uma atividade específica de 4,97 U/mg. O K_m foi determinado pela variação da concentração do Fru-1,6 2P (Figura 7A). Como resultado, o $K_{m\text{Fru-1,6 2P}} = 32 \mu\text{M}$ foi determinado quando a Fru-1,6 2P variou de 1 μM a 1000 μM . A velocidade da reação da atividade enzimática variou levemente quando a concentração do substrato Fru-1,6 2P alcançou 75 μM . Já a rEgEno1 apresentou uma atividade específica de 14,12 U/mg. O K_m foi determinado pela variação da concentração do 2-PGA (Figura 7B). Como resultado, o $K_{m2\text{-PGA}} = 0,37 \text{ mM}$ foi determinado quando o 2-PGA variou de 0,01 mM a 3,2 mM. A velocidade da reação da atividade enzimática variou levemente quando a concentração do substrato 2-PGA alcançou 1 mM.

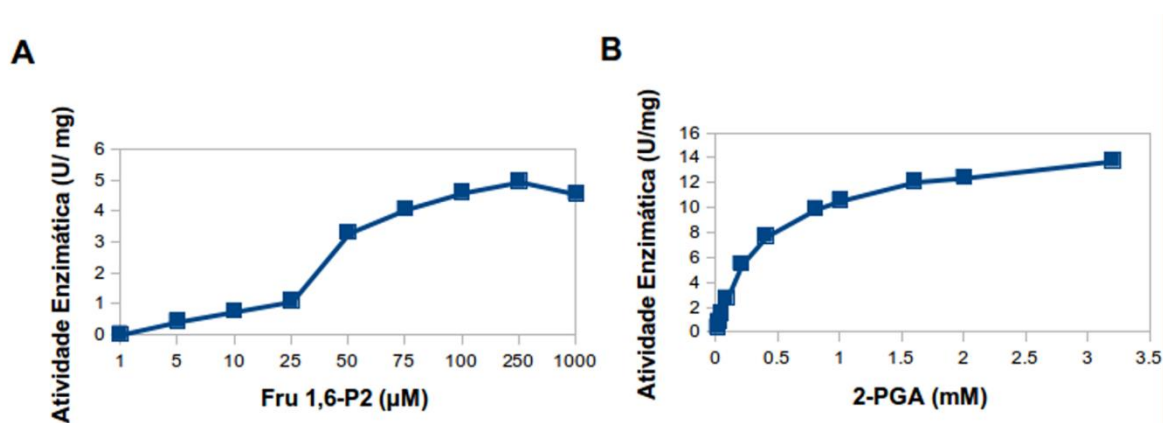


Figure 7. Caracterização cinética das enzimas glicolíticas recombinantes. (A) A atividade enzimática da rEgFBA1 é demonstrada em função da concentração do substrato Fru-1,6-P2. (B) A atividade enzimática da rEgEno1 é demonstrada em função da concentração do substrato 2-PGA.

4.2.2.3 Proteínas de interação com a EgFBA1 e EgEno1

Para obter evidências de possíveis funções *moonlighting* desempenhadas pela EgFBA1 e EgEno1, ensaios *in vitro* de interação proteína-proteína usando a rEgFBA1 e a rEgEno1 marcadas com Sulfo-SBED e extratos de protoescólices de *E. granulosus* foram realizados. A análise das proteínas interagentes, em um espectrômetro de massas mais sensível, permitiu a identificação de 25 e 31 proteínas de interação com a rEgFBA1 e a rEgEno1, respectivamente (Tabela 1).

Tabela 1. Proteínas de interação com a rEgFBA1 e a rEgEno1^a

Número de acesso ^b	Descrição da proteína ^c	rEgFBA1	rEgEno1
EgrG_000008700	Hsp90	•	•
EgrG_000107200	CoA-butirato-ligase	•	•
EgrG_000122100	Profilina		•
EgrG_000142500	Proteína <i>major vault</i>	•	•
EgrG_000190400	Actina citoplasmática 5	•	•
EgrG_000193700	Anexina	•	•
EgrG_000212700	<i>Major egg antigen p40</i>	•	•
EgrG_000254600	Gliceraldeído-3-fosfato-desidrogenase	•	•
EgrG_000292700	Fosfoenolpiruvato	•	•
EgrG_000342600	N,K-ATPase	•	•
EgrG_000364000	14-3-3.3	•	•
EgrG_000372400	Proteína tegumentar		•
EgrG_000389100	Aldeído-desidrogenase		•
EgrG_000465500	Inibidor de fosfatase 2A	•	
EgrG_000514200	Enolase	•	•
EgrG_000575900	Proteína de membrana basal (heparana-sulfato)	•	•
EgrG_000634800	Lactato-desidrogenase	•	•
EgrG_000672200	Tubulina		•
EgrG_000711500	Ubiquitina		•
EgrG_000719700	Calpaína	•	•
EgrG_000775600	Proteína reguladora de quinase dependente de cAMP		•
EgrG_000791700	Tiorredoxina-peroxidase	•	•
EgrG_000905600	Frutose-bifosfato-aldolase	•	•
EgrG_000913300	Peróxido-redutase tiorredoxina dependente	•	•
EgrG_000982200	Fator de alongamento 1 alfa	•	•
EgrG_001008200	Transportador de cálcio atpásico	•	•
EgrG_001028500	Citrato-sintase	•	•
EgrG_001065500	Ndr	•	•
EgrG_001085400	Hsp70	•	•
EgrG_001177600	Fator de ribosilação do ADP		•
EgrG_001185000	Malato-desidrogenase	•	•
EgrG_001192500	14-3-3.1	•	•

^a A detecção de proteínas interagentes é indicada por •; ^b Número de acesso do GeneDB;

^c Descrição da proteína conforme o GeneDB.

A anotação funcional das proteínas identificadas como interagentes da rEgFBA1 e rEgEno1 foi feita com base nas classificações do GO (Figuras 8). Várias proteínas de interação são comuns às duas proteínas e, conseqüentemente, várias categorias funcionais estão sobrepostas. Além das categorias relacionadas com a função primária destas enzimas (processos metabólicos), várias categorias identificadas sugerem que estas enzimas são multifuncionais. Regulação, desenvolvimento e resposta a estímulos estão entre as categorias comuns entre as duas proteínas. Interessantemente, categorias de sinalização e processos do sistema imune foram exclusivamente detectadas nos ensaios com a rEgEno1.

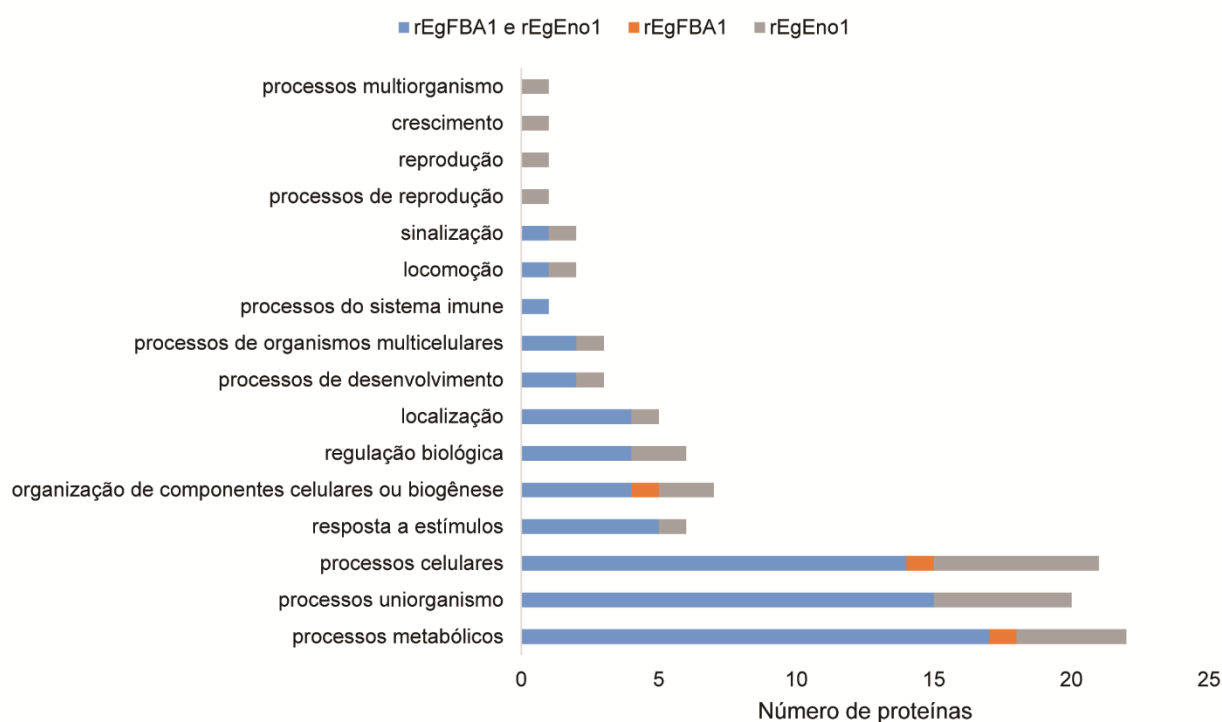


Figure 8. Classificação funcional do GO das proteínas de interação da rEgFBA1 e da rEgEno1. O número total de proteínas identificadas como uma função de todos os termos do GO disponíveis no nível 2 para processos biológicos são mostrados.

4.3 Análise proteômica *top down* e *bottom up* de frações subcelulares de protoescólices de *E. granulosus* contendo proteínas de baixa massa

Manuscrito a ser submetido para publicação na revista *Journal of Proteome Research*
Karina Rodrigues Lorenzatto, Kyunggon Kim, Ioanna Ntai, Gabriela Prado Paludo, Jeferson Camargo de Lima, Paul Martin Thomas, Neil L. Kelleher, Henrique Bunselmeyer Ferreira

Top down proteomics reveals mature proteoforms expressed in the larval stage of Echinococcus granulosus

Contribuição dos autores:

K.R.L: delineamento experimental, execução de todos os experimentos e redação do manuscrito; K.K. e I.N.: auxílio nos experimentos de GELFrEE e análise proteômica *top down* e *bottom up*; G.P.P: auxílio na preparação do banco de dados contendo MPTs; J.C.L.: auxílio na preparação das amostras biológicas; P.M.T, N.L.K, H.B.F: delineamento experimental, análise e discussão dos resultados e revisão do manuscrito.

Top down proteomics reveals mature proteoforms expressed in the larval stage of *Echinococcus* *granulosus*

*Karina R. Lorenzatto*¹, *Kyunggon Kim*², *Ioanna Ntai*², *Gabriela P. Paludo*¹, *Jeferson C. de Lima*¹,
*Paul M. Thomas*², *Neil L. Kelleher*², *Henrique B. Ferreira*^{*,1}

¹ Laboratório de Genômica Estrutural e Funcional, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves, 9500, Porto Alegre, Rio Grande do Sul, Brazil

² Departments of Chemistry and Molecular Biosciences, and the Proteomics Center of Excellence, Northwestern University, 2145 N. Sheridan Road, Evanston, Illinois 60208, United States

Keywords: top down proteomics, hydatid disease, low mass proteome, subcellular fractions, post-translational modifications

Abstract

Echinococcus granulosus is the causative agent of cystic hydatid disease, a neglected zoonosis responsible for high morbidity and mortality. Several molecular mechanisms underlying parasite biology remain poorly understood. Here, *E. granulosus* subcellular fractions were analyzed by top down and bottom up proteomics for protein identification and characterization of cotranslational and post-translational modifications (CTMs and PTMs, respectively). Nuclear and cytoplasmic extracts of *E. granulosus* protoscoleces were fractionated by 10% GELFrEE and proteins under 30 kDa were analyzed by LC-MS/MS. By top down analysis, 187 proteins and 207 proteoforms were identified, being 122 and 52 proteoforms exclusively detected in nuclear and cytoplasmic fractions, respectively. CTMs were evident as 71% of the proteoforms had methionine excised and 47% were N-terminal acetylated. In addition, *in silico* internal acetylation prediction coupled with top down MS allowed the characterization of 9 proteins differentially acetylated, including histones. Bottom up analysis increased the overall number of identified proteins in nuclear and cytoplasmic fractions to 154 and 112, respectively. Overall, our results provided the first description of the low mass proteome of *E. granulosus* subcellular fractions and highlighted proteoforms with CTMs and PTMS which might be playing critical roles in regulating cellular processes occurring in this parasite species.

Introduction

The larval stage of *Echinococcus granulosus* (hydatid cyst or metacestode) is the causative agent of cystic hydatid disease or echinococcosis, which has a cosmopolitan distribution¹⁻³ and is one of the 17 neglected tropical diseases prioritized by the World Health Organization (WHO) (http://www.who.int/neglected_diseases/diseases/en/). Molecular mechanisms controlling *E. granulosus* metacestode development, immune evasion and host-parasite interplay are still poorly characterized. The advent of “omics” approaches has greatly impacted the amount of genomic, transcriptomic, and proteomic data for an increasing number of parasites.⁴ The recent descriptions of the *E. granulosus* genome and transcriptome⁵⁻⁷ were undoubtedly a major accomplishment and are expected to boost post-genomic studies, and to unveil a new understanding of the biology of this tapeworm.

E. granulosus has two developmental stages in its life cycle: the adult tapeworm, which is the sexual stage, and the cystic larval metacestode, which reproduces asexually.⁸ Within the metacestode, the cyst germinal layer produces the pre-adult form, called protoscolex. Interestingly, *E. granulosus* protoscoleces present a remarkable developmental plasticity: they can develop into adult worms when ingested by a suitable definitive host (usually dogs), or, upon release from a ruptured cyst in an intermediate host (domestic ungulates or humans), they can redifferentiate into secondary hydatid cysts.^{8, 9} The molecular mechanisms behind such developmental plasticity are still poorly understood, but it is likely that they involve several levels of regulation. “Omics” approaches have identified potential *E. granulosus* regulatory molecules, including microRNAs, found in RNA-Seq studies,^{10, 11} and regulatory proteins, such as transcription factors and signaling proteins, identified in proteomic studies.¹²⁻¹⁵

Protein function regulation in different contexts can be mediated by cotranslational and post-translational modifications (CTMs and PTMs, respectively), which can play critical roles in the regulation of a broad range of cellular processes in both prokaryotes and eukaryotes.¹⁶⁻¹⁹ These

modifications should be regulating several biological processes in *E. granulosus* as well, but they remain largely unknown in this parasite. In fact, CTMs and PTMs have been poorly addressed in parasitic flatworms in general, and their characterization may lead to another level of understanding about molecular mechanisms controlling parasitic flatworm biology.

CTM and PTM characterization is not a trivial task and the availability of tools to achieve complete mapping of all expressed proteins variants (proteoforms) is limited. The term proteoform is currently used to designate all molecular forms in which the protein product of a single gene can be found, including changes due to genetic variation, alternatively spliced RNA transcripts, CTMs and PTMs.²⁰ Mass spectrometry (MS)-based proteomics has evolved dramatically, and is now considered a key technology for identifying and quantifying proteins and PTMs.²¹⁻²⁴ The bottom up workflow allows high throughput protein identification but the assignment of PTM sites is compromised due to the MS dynamic range and the typical low sequence coverage obtained. Peptide enrichment decreases this problem and allows the shotgun approach to be applied to the identification and quantification of PTMs.²³ However, bottom up MS still faces limitations for combinatorial detection of PTMs, unless the modifications are co-located on one peptide. Top down proteomics, on the other hand, is based on the analysis of intact proteins, and is well-suited to the characterization of proteoforms, as it can detect proteins with CTMs and PTMs in a general fashion and can produce a faithful representation of protein variants resulting from maturation within cells.²⁵ Top down MS is not as widely spread as bottom up MS, but recent efforts have enabled its use in a high throughput fashion.²⁶⁻²⁸

Here, for the first time, the low mass proteome from subcellular fractions from *E. granulosus* protoscoleces has been addressed. A top down coupled to a bottom up proteomic approach have been used to not only identify but also characterize the low mass proteome, highlighting the detection of CTMs in proteoforms expressed in *E. granulosus* larval stage. Our findings revealed new molecular aspects of *E. granulosus*, adding another level of understanding in

the biology of this neglected parasite.

Material and Methods

Parasite Material

E. granulosus hydatid cysts were obtained from lungs of cattle slaughtered in a local abattoir (Cooperleo, São Leopoldo, Brazil). Animal slaughtering was conducted according to Brazilian laws and under supervision of the *Serviço de Inspeção Federal* (Brazilian Sanitary Authority) of the Brazilian *Ministério da Agricultura, Pecuária e Abastecimento*. Condemned viscera at post-mortem meat inspection due to the presence of hydatid cysts were collected at the abattoir and dissected in the laboratory, following described protocols¹³, all approved by the Ethical Committee of the *Universidade Federal do Rio Grande do Sul*. Protoscoleces were collected by aspiration, washed several times in PBS and were kept frozen until genotype confirmation. Parasite genotyping was performed for each individual cyst, as previously described²⁹ and all used parasite samples were from the G1 genotype, defined as *E. granulosus* stricto sensu (*E. granulosus* s. s.).³⁰

Protoscolex nuclear and cytosolic proteins were extracted following the protocol described previously,³¹ with some modifications. Briefly, 500 mg of protoscoleces from 9 lung cysts of the same genotype were pooled out and were homogenized with 3 volumes of buffer N (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.5% Nonidet P-40) containing Halt protease and phosphatase inhibitors (Thermo Scientific), on ice, on a Tekmar homogenizer, using 10 bursts of 10 s each and 1 min interval between bursts (at a setting of 60-100). After centrifugation for 5 min at 4000 g, the pellet was suspended, homogenized and centrifuged again under the same conditions. The two supernatants generated in these centrifugation steps were combined, and an aliquot was stained with DAPI and checked on FLoid Cell Imaging Station (Life Technologies), to verify nucleus integrity and enrichment. Then, this preparation was centrifuged using Type 50.2 Ti rotor (Beckman Coulter) at 100,000 g for 1 h at 4°C, for nuclei and cytosolic proteins isolation. The

resulting supernatant containing cytosolic proteins (cytosolic fraction) was saved, and the pellet containing the nuclei (nuclear fraction) was suspended using 1 volume of buffer F (50 mM HEPES pH 7.5, 5 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA, 0.5 mM DTT) containing Halt protease and phosphatase inhibitors. Nuclear and cytosolic fractions were both lyophilized and stored at -20°C until use.

Sample Preparation

Nuclear and cytosolic lyophilized proteins were reconstituted in 4% SDS, acetone precipitated with six volumes of cold acetone and centrifuged at max speed for 10 min. The precipitated pellets were reconstituted in 1% SDS and protein concentration was determined using BCA assay (ThermoFisher, Rockford, IL). For Gel Elution Liquid-based Fractionation Entrapment Electrophoresis (GELFrEE), 250 µg of protein were fractionated in a GELFrEE 8100 Fractionation system (Expedeon, San Diego, CA) using 10% cartridges. Fractions 1-6 were collected and a 6 µL aliquot of each 150 µL fraction was used for conventional SDS-PAGE analysis and silver stain visualization.

For intact protein preparation, fractions were precipitated with MeOH/CH₃Cl/H₂O to remove SDS.³² Prior to MS analysis, the samples were resuspended in 20-30 µL of buffer A (95% H₂O, 5% acetonitrile, 0.2% formic acid). For peptide preparation, fractions were precipitated in acetone and pellets were resuspended in 8 M urea in 100 mM NH₄HCO₃, followed by reduction with 20 mM DTT and cysteine alkylation with 40 mM iodoacetamide. Protein mixtures were trypsin digested (enzyme to protein ratio 1:50) at 37°C for 18 h. Peptides were desalted using Zip-Tip C18 (Millipore), dried in a SpeedVac and suspended in 30 µL of buffer A.

LC-MS/MS of intact proteins

A 6 µL volume of each GELFrEE fraction was injected in triplicate onto a trap column (150 µm ID x 2 cm) and eluted onto an analytical column (75 µm ID x 15 cm) and online separated as

described previously.³³ Both trap and analytical columns were packed in-house³⁴ with polymeric reverse phase (PLRP-S, Phenomenex) media (5 μm , 1000 \AA pore size) and connected to 15 μm nanoelectrospray tips (New Objective, Waltham, MA). Samples were loaded onto the trap using an Ultimate 3000 RPLCnano system (Thermo Scientific Dionex) at a flow rate of 2.5 $\mu\text{L}/\text{min}$, and then separated on the analytical column and eluted into the mass spectrometer using a flow rate of 300nL/min. Mass spectrometry data were obtained on an Orbitrap Elite (Thermo Scientific). The MS method included the following events: (i) FT scan, four microscans, m/z 500-2,000, and resolution 100 000; and (ii) data-dependent MS/MS on the top two peaks in each spectrum from scan event 1 using higher-energy collisional dissociation (HCD) with normalized collision energy of 25, isolation width 50 m/z , four microscans, and detection of ions with resolving power of 60 000. Dynamic exclusion was enabled with a repeat count of 2, a repeat duration of 120 s, and an exclusion duration of 5000 s. Automatic gain control (AGC) was set to 1E6 ions, and maximum injection time was set to 1 s for both MS1 and MS2. Advanced signal processing was turned on, and data were collected in reduced profile mode. A 15 V offset in the source was used over the entire experiment. The capillary temperature was 320°C and the spray voltage was 1.8 kV.

LC-MS/MS of tryptic peptides

A 10 μL volume of each digested GELFrEE fraction was injected in duplicate onto a trap column (150 μm ID x 2 cm) and eluted onto an analytical column (75 μm ID x 15 cm). Both trap and analytical columns were packed in-house with 3 μm Aqua C18 (Phenomenex) and connected to 15 μm nanoelectrospray tips (New Objective, Waltham, MA). Samples were loaded onto the trap using an Ultimate 3000 RPLCnano system at a flow rate of 3 $\mu\text{L}/\text{min}$, and then separated on the analytical column and eluted into the mass spectrometer using a flow rate of 300nL/min. Data acquisition on the 7 T LTQ-FT-ICR (Thermo Scientific) consisted of a full scan event, m/z 400-1,600 and resolution 50 000 and data-dependent MS/MS on the top eight peaks in each spectrum from scan event 1 using collision-induced dissociation (CID) with normalized collision energy of

35 and isolation width 2 m/z. Dynamic exclusion was enabled with a repeat count of 3, a repeat duration of 45 s, and an exclusion duration of 120 s. The capillary temperature was 200°C and the spray voltage was 2.4 kV.

Data processing

Top down analysis was performed essentially as described previously.²⁷ Raw data files were processed in cRAWler, an in-house software which determine intact and fragment masses and compiled them in ProSight Upload Format (*.puf) files. At first, each *.puf file was searched against *E. granulosus* data, available at GeneDB and UniProtKB^{5,6}, using ProSightPC 3.0 (Thermo Scientific, San Jose, CA). An iterative absolute mass search tree was implemented in ProSightPC, starting with a narrow search window of 10 Da of intact mass tolerance. When the first search failed to find matches, it was followed by a second search using a wide search window of 100,000 Da. The wide search detect hits with PTMs that are not annotated in the database as well as those cases with large differences between theoretical and observed intact mass values. In both cases, a 15 ppm mass tolerance was used for the fragment ions. Low confidence proteoform identifications were excluded by considering only those hits arising from an absolute mass search with an E-value below 1×10^{-4} .

The amino acid sequences of proteins identified in the first search were submitted to an *in silico* prediction of lysine acetylation using the web tool PAIL (<http://bdmpail.biocuckoo.org/prediction.php>). For that, a database including the predicted acetylations was built and used in a second round of searches on ProSightPC, applying the same parameters set in the first search.

For bottom up analysis, raw data files were processed on COMPASS.³⁵ Peak lists were searched using the MASCOT search engine (Matrix Science, London, UK; version 2.3.0) against the *E. granulosus* databases (available at GeneDB and UniProt). Database searches were performed with the following parameters: one missed cleavage; fixed modifications: carbamidomethylation of

cysteine; variable modifications: oxidation of methionine; parent ion tolerance: 20 ppm; fragment ion tolerance: 0.6 Da. To validate MS/MS based peptide and protein identifications, MASCOT *.dat files were loaded on Scaffold (Proteome Software, Portland, OR; version 3.6) and parameters to accept an identification were as following: peptide probability > 95.0%, protein probability > 99.0% and ≥ 2 unique peptides.

Functional classification

Online BLASTP searches were performed using BLAST2GO³⁶ version 3.0 against the NCBI nr database. GO mapping, InterProScan analysis and annotation were performed based on BlastP results (E-value $\leq 1.0E-3$). Bar charts were generated using all available GO subcategories of level 3.

RESULTS AND DISCUSSION

Top down MS identification of proteins and proteoforms

To study the mature proteoforms from nuclei and cytoplasm enriched fractions from protoscolecids from *E. granulosus* s.s., those proteins were GELFrEE fractionated and submitted to top down MS. Representative results of LC separation from a GELFrEE fraction are illustrated in **Figure 1A and B**. In total, 186 protein IDs were identified and some of them appear in different proteoforms, resulting in the identification of 207 proteoforms (**Supplementary Table S1A and B**). The distribution of exclusive and shared protein IDs and proteoforms between nuclear and cytosolic fractions are shown in **Figure 2**. There were 122 and 52 proteoforms exclusively detected in nuclear and cytosolic fractions, respectively. Interestingly, the number of proteoforms shared between these fractions is smaller than the number of shared protein IDs. This is explained by the fact that seven proteins IDs shared by both subcellular fractions, have corresponding proteoforms exclusive of the nuclear or the cytoplasmic fraction.

Ribosomal proteins (6-29 kDa) and dyneins (9-23 kDa) were detected in several GELFrEE

fractions. This is a common result with top down due to their high cellular abundance and low molecular weight. A perfect separation is not achievable due to overlaps between fractions and also to the occurrence of proteolytic events, both natural or artifactual.²⁵ In addition, the standard mode of data acquisition used here has also led to overlaps in the identification of highly abundant proteins in subsequent LC-MS injections and the development of new methods is expected to circumvent the overlapping.³⁷ Despite that, less abundant proteins exhibiting confident identifications were also detected; two examples are shown in **Figure 1C**.

To compare the sensitivity of the top down approach, the same nuclear and cytosolic GELFrEE fractions were also analyzed by bottom up proteomics (**Supplementary Tables 2A-D**). In the bottom up analysis, 208 proteins were confidently identified with at least two unique peptides. Of these proteins, 56 were also identified in the top down approach (**Figure 3A**). There were 152 proteins detected only in the bottom up approach, but most of them have molecular masses larger than 30 kDa (**Figure 3B**), which makes them less amenable in our top down analysis. There were 130 proteins uniquely identified in the top down approach and 34 of these proteins have molecular weights below 10 kDa, contrasting with only 4 proteins uniquely identified by bottom up in this same mass range (**Figure 3B**). Characterizing small proteins represents a challenge for the bottom up workflow due to the inability to generate sufficient tryptic peptides for analysis.³⁸ The ability of top down for low mass protein detection enabled the identification of 25 proteins not previously reported, increasing the repertoire of proteins known to be expressed in *E. granulosus* protoscoleces (**Supplementary Table 3**).

Characterization of identified proteoforms

The characterization of CTMs and PTMs is one of the enduring goals of top down proteomics and to accomplish that in a high throughput workflow, large combinations of PTMs can be annotated and considered within databases.²⁶ In the case of *E. granulosus* genome, the currently available annotations were solely based on DNA sequencing and no PTMs have been annotated so

far. Even in cases like this, top down can provide the detection of mature proteoforms. ProSightPC can presume a theoretic N-terminal processing, including methionine excision and N-terminal acetylation, in a way it can potentially occur on all proteins of a given database, even if they are not present in the input, enabling the detection of cotranslationally modified proteoforms.^{39, 40}

The most prevalent types of CTMs identified in nuclear and cytosolic proteoforms from *E. granulosus* in our top down approach were N-terminal methionine excision (NME) and N-terminal acetylation (NTA). NME was evident as 130 out of 182 proteoforms underwent this processing (**Supplementary Table S1C**), corresponding to 71% of the identified proteoforms. It is thought that up to 80% of the proteins of any given proteome undergo this modification.⁴¹ In the case of NTA, 85 out of 182 proteoforms were found to contain this type of processing (**Supplementary Table S1D**), corresponding to 47%, which is close to the 50-80% usually found in other eukaryotes.⁴² The apparent underrepresentation of NMEs and NTAs in the protein repertoires of the analyzed samples may be explained by the fact that proteoforms with no N-terminal fragment ions were not included in these analyses. Moreover, our analyses were essentially focused in proteins under 30 kDa, which might also impact the overall proportion of modified proteoforms.

NME is an essential process that occurs in the cytoplasm of all organisms and in the two organelles displaying protein synthesis.⁴³ NME involves two types of methionine aminopeptidase (MAP), MAP1 (type-I) and MAP2 (type-II), which have similar three-dimensional structures despite having only low levels of sequence identity. Higher eukaryotes have at least one MAP1 and one MAP2 in the cytoplasm and one MAP1 in the organelles.⁴⁴ *E. granulosus* genome has at least two genes containing MAP1 domain (UniProt accessions: U6J5U8 and A0A068X0Z5) and one gene containing MAP2 domain (UniProt accession: A0A068W7Q0). According to the generally accepted cleavage rules for MAP, this enzyme cleaves all proteins with small side chains on the residue in the second position,⁴⁴ but many exceptions are known.⁴⁵ In this work, the penultimate residue was usually consistent with NME rule as Ala, Cys, Gly, Pro, Ser, Thr and Val represent 86% of proteins with NME. Among the rest of potential substrates found here there was Arg, Asp,

Glu, Gln, Ile, Leu, Lys and Trp. This is the first time that NME is described for *E. granulosus* proteins and its detection provides indirect evidence for MAPs activities. Although some substrates are not in agreement with the N-end rules,⁴⁴ the possibility that they represent exceptions for this parasite cannot be ruled out. In these cases, studies for the evaluation of substrate specificity of *E. granulosus* MAPs need to be further explored.

NTA is another cotranslational modification conserved among eukaryotes,⁴⁶ and they are performed by a group of N-terminal acetyltransferases (NATs) complexes, NatA–E. These NATs differ both in subunit composition as well as in substrate specificity, and combined, they act on the majority of eukaryotic proteins.⁴² Most of the NTA events are performed by NatA on MAP newly generated N-termini.⁴² This seems to be the case for *E. granulosus* as well, since most of NTA events occurred in proteins processed by MAP (**Figure 4A**), being Ser, Ala and Thr the top three residues in the second position (**Figure 4B**). N-termini with larger amino acid residues in the second position are not normally cleaved by MAPs, but potentially acetylated directly on the initiator methionine by NatB, C and E, depending on the N-terminal sequence.¹⁸ The NatB, C and E substrate specificities seem to be at least partially conserved in *E. granulosus* (**Figure 4B**). Higher eukaryotes also express NatF, which, due to its distinct substrate preference, is responsible for an overall increase in protein acetylation.⁴⁷ N-acetylated sites detected here indirectly support the activity of all NAT complexes, including NatF. At least 11 genes from *E. granulosus* predicted as NATs contain homolog domains to NAT subunits described for other eukaryotes. *E. granulosus* NATs possible composition and substrates here detected are shown in **Supplementary Table S4**.

Differently of the N-terminal processing, the current unreviewed status of *E. granulosus* database is undoubtedly a bottleneck for detection of internal PTMs. In addition, even though the genotype used to prepare the nuclear and cytosolic fractions was the same used in the genome project (G1), the known variability within *E. granulosus* s. s. requires a non-stringent tolerance to differences between theoretical and experimental masses, hindering the attempts for manual detection of common PTMs based on mass shifts. In an attempt to investigate the effect of the

description of PTMs in the database, acetylation was predicted for all *E. granulosus* nuclear and cytosolic identified proteoforms. In the case of organisms with annotated PTMs (human and yeast, for example), the database of possible proteoforms is generated by combinatorial expansion of all potential proteoforms using a modified UniProtKB flat file.^{39, 40, 48} The size of the database is controlled by setting the maximum features per sequence and 13 is used as default setting, as rarely a protein has more than 13 known modifications.⁴⁰

In the case of prediction data, the number of predicted modifications extrapolates the default setting and its increase would make the storage of all possible proteoforms and also the search time computational prohibitive. Therefore, the database of *E. granulosus* potential proteoforms was generated by including the prediction information in a modified UniProtKB flat file and keeping the default setting for combinatorial expansion. The use of a modified database enabled for the detection of 9 acetylated proteoforms (**Supplementary Table S1E**), highlighting histones among the proteoforms differentially acetylated (**Figure 5**). The inability to consider all predicted modifications in a combinatorial fashion may have limited the detection of modifications. Even though the contribution of our custom database was limited to a few acetylated proteoforms, their detection was database-dependent, anticipating the impact that a more comprehensive database can have in the assignment of modified proteoforms from *E. granulosus*.

Functional classification of the low mass proteome of nuclear and cytosolic fractions from *E. granulosus* protoscoleces

The combination of top down and bottom up approaches resulted in an overall identification of 338 proteins, from which 154 and 112 were exclusively identified in nuclear and cytosolic fractions, respectively (**Figure 6A**). We used GO term enrichment analysis to classify the functions of the identified proteins. Sequence annotations associated with 280 proteins (83%) were categorized into the three main GO ontologies: cellular component (**Figure 6B**), molecular function and biological process (**Supplementary Figure S1**). The GO terms were summarized into 72, 84

and 73 subcategories for nuclear, cytosolic and shared proteins, respectively, with GO classifications at level 3.

Functional enrichment analyses revealed that nuclear and cytosolic fractions were enriched in proteins with annotations consistent with nuclear and cytosolic functions, respectively, which points out to the effectivity of the protoscolex subcellular fractionation method. In nuclear fractions, cellular components such as cell part and organelle part (**Figure 6B**) were enriched in nucleosome and spliceosome proteins. Within membrane-bounded organelle, apart from nuclear envelope proteins, there also were membrane proteins of non-nuclear organelles, known as major sources of contamination in nuclear preparations.⁴⁹ On the other hand, glycolytic and oxidative stress proteins were enriched in the cytosolic fractions. Several nuclear proteins were also detected in the cytosolic fraction and cytosolic proteins were detected in the nuclear fractions. This indicates that many proteins in *E. granulosus* shuttle between the nucleus and cytosol and/or perform their tasks at multiple destinations. The unusual locations of these proteins is suggestive that they can be moonlighting, a property previously described for other *E. granulosus* proteins.⁵⁰

There were 58 proteins unassigned to GO terms and most of them remain annotated as uncharacterized proteins. Interestingly, 24 of these proteins were detected exclusively in the nuclear fractions and, 6 of them (UniProt: A0A068WLY9, A0A068WFZ9, U6JNT9, A0A068WT65, A0A068WEA7, A0A068WUA4) appear to be restricted to the *Echinococcus* genus, since only *Echinococcus multilocularis* appear to be sharing these orthologs. In the cytosolic fraction, 1 out of 10 uncharacterized proteins (UniProt: A0A068W6R5) also appear restricted to the *Echinococcus* genus and among the shared proteins, 1 out of 13 (UniProt: A0A068WW55) seems to be exclusive from *E. granulosus* s. s. The group of nuclear, cytosolic and shared uncharacterized proteins also includes 3 (UniProt: U6JF24, A0A068W906, A0A068X009), 1 (UniProt: A0A068WUI4), and 2 (UniProt: U6JBE2, U6JCZ1) orthologs limited to tapeworms, and 2 (UniProt: A0A068X2D5, A0A068WLV4), 1 (UniProt: A0A068WNL1), and 2 (UniProt: A0A068X046, A0A068X0E4) orthologs limited to helminths, respectively. These proteins may carry novel domains likely linked

to unrecognized nuclear and cytosolic functions not only in *Echinococcus* but also in helminths and/or tapeworms, making them intriguing targets to be explored in future studies.

CONCLUSIONS

Top down proteomics gives the first step to fully characterize proteins from *E. granulosus*. The proteoforms detected here support that N-terminal processing is conserved in *E. granulosus* proteins and indirectly suggest the functionality of MAPs and NATs in this parasite. Our modified database also provided the identification of a few internal acetylated sites, adding to detection of mature proteoforms. Moreover, the use of bottom up approach allowed the identification of IDs not detected using top down MS, adding in the coverage of protein profiles expressed in *E. granulosus* subcellular fractions. Future works using complementary approaches, exploring not only how *E. granulosus* proteins are decorated but also addressing their impact in its biological processes and subcellular location, can lead us to unravel further molecular mechanisms of this neglected parasite.

ACKNOWLEDGMENTS

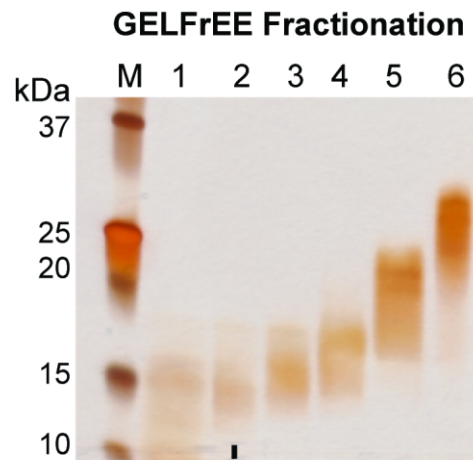
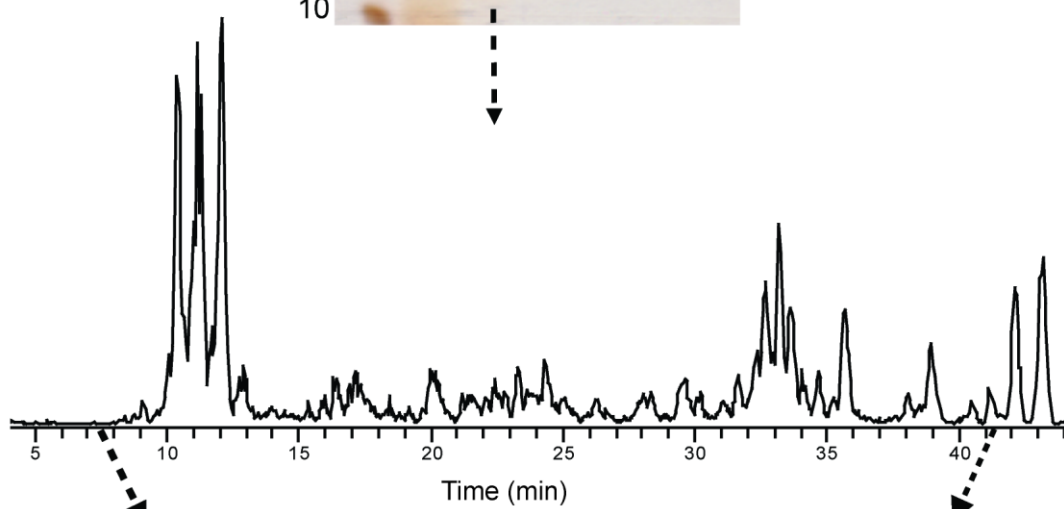
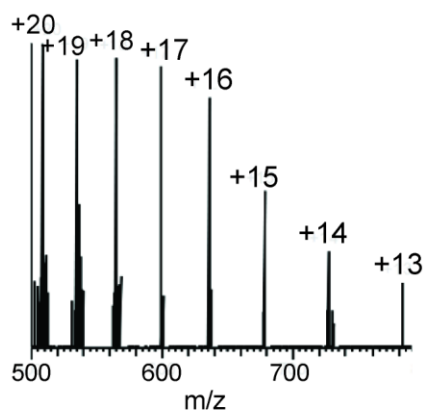
This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Grant 23038.005300/2011-14) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant 477649/2012-2) to H.B.F. and by Award No. GM067193 from the National Institute of General Medical Sciences (N.L.K.). K.R.L. was recipient of CNPq and CAPES PhD fellowships (141812/2011-7 and 12676/13-6). G.P.P. was recipient of a CNPq undergraduate fellowship (164139/2012-5). J.C.L. was recipient of a CNPq technical support fellowship (377742/2013-9).

COMPETING INTERESTS

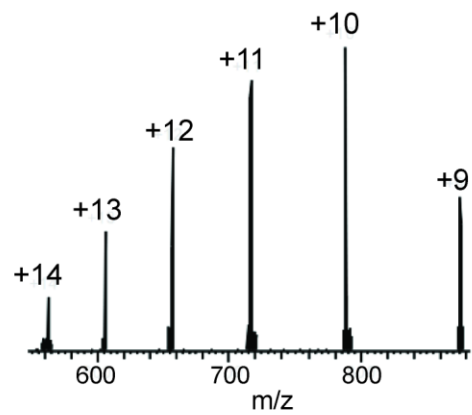
The authors declare the following competing financial interest(s): Qualitative protein analyses were performed in ProSightPC, a product commercialized by the Kelleher Research Group.

ABBREVIATIONS

CID, collisionally induced dissociation; CTM, co-translational modifications; HCD, higher energy collisional dissociation; MS, mass spectrometry; PTM, post-translational modification; PUF, ProSight Upload Format; SDS, sodium dodecyl sulfate

A**B****C**

Histone H1 gamma
(E-value: 1.4×10^{-28})



Guanine nucleotide-binding protein
(E-value: 6.5×10^{-24})

Figure 1. Representative data for LC-MS/MS analysis of a single GELFrEE fraction. (A) GELFrEE fractionation of the nuclear sample is shown as example; (B) LC total ion chromatogram from GELFrEE fraction 2; (C) Two representative intact protein spectra are highlighted.

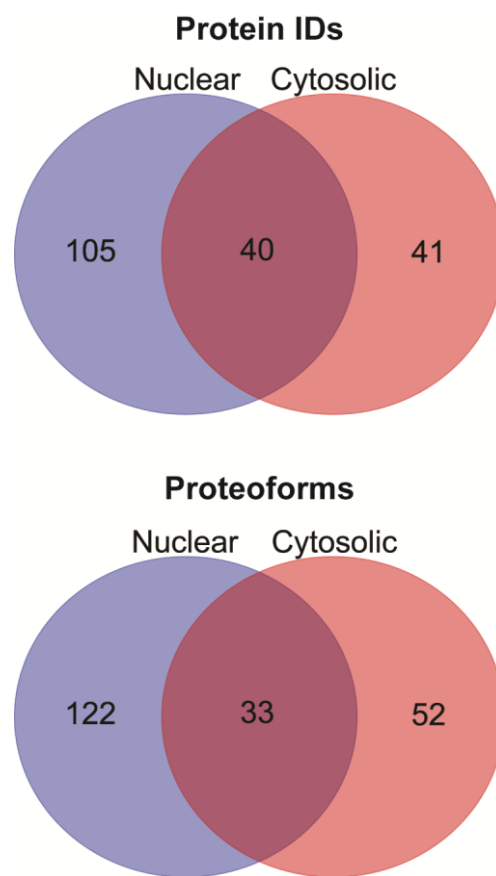


Figure 2. Overlap between nuclear and cytosolic protein IDs and proteoforms identified in our top down approach. The comparison of proteoforms points out that although some IDs seems to be common, they exhibit modifications which make them different proteoforms, decreasing the overlapping between nucleus and cytosol.

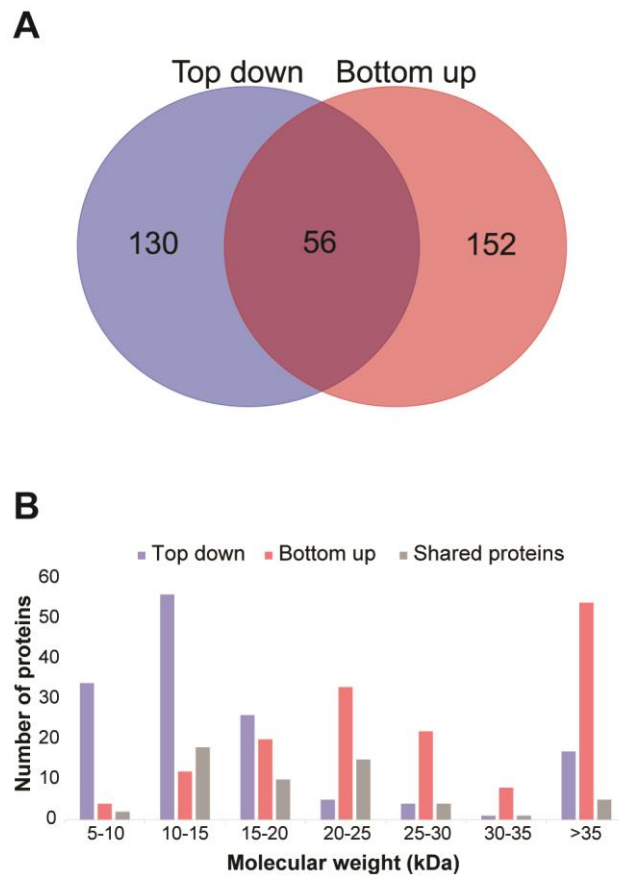


Figure 3. Comparison between proteins identified using top down and bottom up approaches. (A) Venn diagram showing a partial overlap between the approaches; (B) Molecular mass distribution of proteins identified in both approaches.

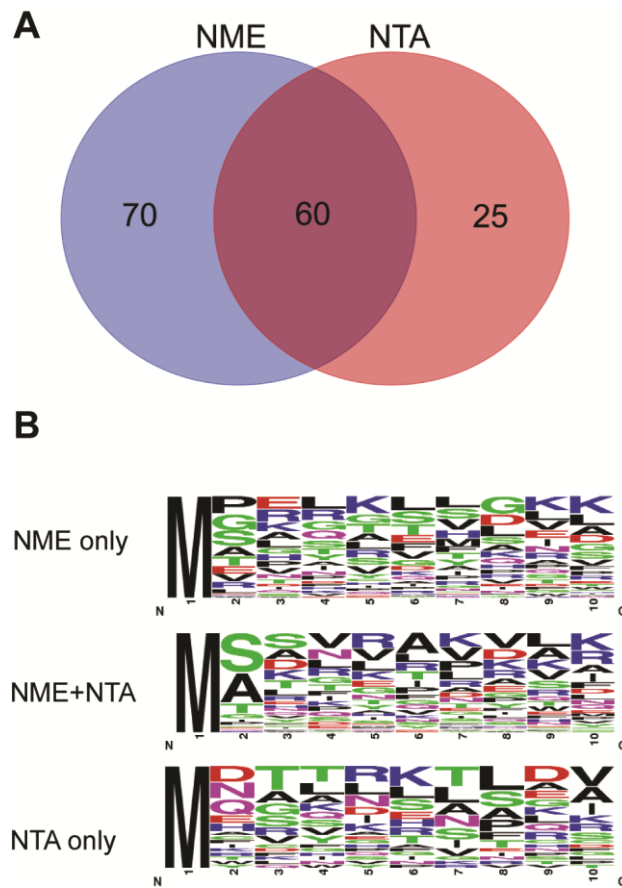


Figure 4. Overview of the CTMs and sequence motifs detected in the identified proteoforms. (A) Venn diagram showing the distribution NME and NTA events. The overlap indicate proteoforms modified by NME+NTA. (B) Sequence motifs for the first ten residues of each proteoform identified displaying NME only, NME+NTA and NTA only.

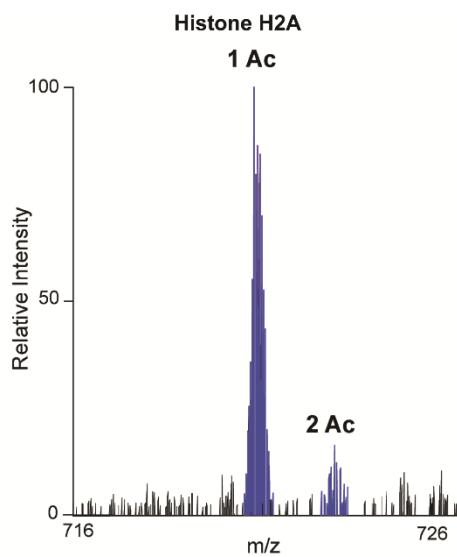
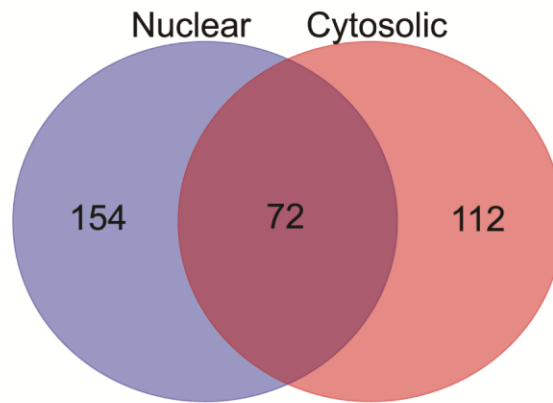


Figure 5. Detection of two proteoforms from histone H2A. These monoacetylated and diacetylated proteoforms were

only detected using our modified database including the acetylation prediction information.

A



B

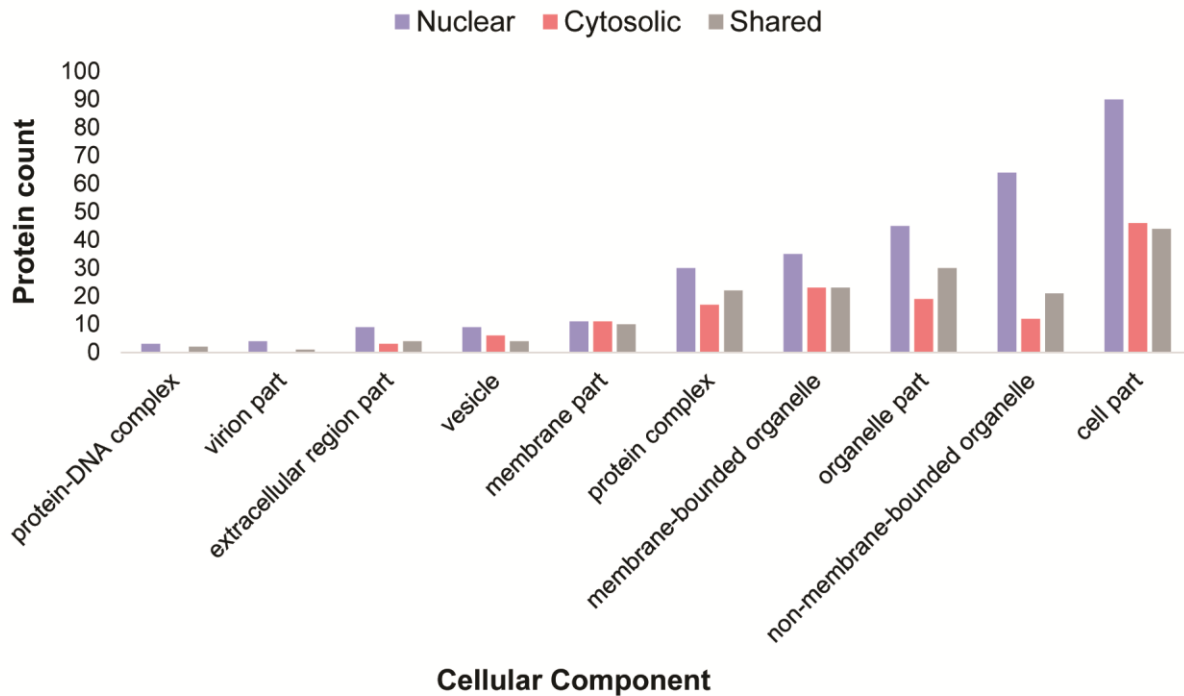


Figure 6. Overview of the low mass proteome from *E. granulosus* protoscoleces subcellular fractions. (A) Distribution of nuclear and cytosolic proteins identified using both top down and bottom up MS; (B) Gene ontology (GO) classification of proteins identified in *E. granulosus* subcellular fractions. The results of BlastP searches against the NCBI nr database were used for GO mapping and annotation. Top 10 most abundant GO terms for cellular component are shown.

REFERENCES

1. Budke, C. M.; Deplazes, P.; Torgerson, P. R., Global socioeconomic impact of cystic echinococcosis. *Emerg Infect Dis* **2006**, 12, (2), 296-303.
2. Moro, P.; Schantz, P. M., Echinococcosis: a review. *Int J Infect Dis* **2009**, 13, (2), 125-33.
3. Carmena, D.; Cardona, G. A., Echinococcosis in wild carnivorous species: epidemiology, genotypic diversity, and implications for veterinary public health. *Vet Parasitol* **2014**, 202, (3-4), 69-94.
4. Thompson, R. C.; Lymbery, A. J., Let's not forget the thinkers. *Trends Parasitol* **2013**, 29, (12), 581-4.
5. Tsai, I. J.; Zarowiecki, M.; Holroyd, N.; Garcarrubio, A.; Sanchez-Flores, A.; Brooks, K. L.; Tracey, A.; Bobes, R. J.; Fragoso, G.; Sciuotto, E.; Aslett, M.; Beasley, H.; Bennett, H. M.; Cai, J.; Camicia, F.; Clark, R.; Cucher, M.; De Silva, N.; Day, T. A.; Deplazes, P.; Estrada, K.; Fernández, C.; Holland, P. W.; Hou, J.; Hu, S.; Huckvale, T.; Hung, S. S.; Kamenetzky, L.; Keane, J. A.; Kiss, F.; Koziol, U.; Lambert, O.; Liu, K.; Luo, X.; Luo, Y.; Macchiaroli, N.; Nichol, S.; Paps, J.; Parkinson, J.; Pouchkina-Stantcheva, N.; Riddiford, N.; Rosenzvit, M.; Salinas, G.; Wasmuth, J. D.; Zamanian, M.; Zheng, Y.; Cai, X.; Soberón, X.; Olson, P. D.; Laclette, J. P.; Brehm, K.; Berriman, M.; Consortium, T. s. G., The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* **2013**, 496, (7443), 57-63.
6. Zheng, H.; Zhang, W.; Zhang, L.; Zhang, Z.; Li, J.; Lu, G.; Zhu, Y.; Wang, Y.; Huang, Y.; Liu, J.; Kang, H.; Chen, J.; Wang, L.; Chen, A.; Yu, S.; Gao, Z.; Jin, L.; Gu, W.; Wang, Z.; Zhao, L.; Shi, B.; Wen, H.; Lin, R.; Jones, M. K.; Brejova, B.; Vinar, T.; Zhao, G.; McManus, D. P.; Chen, Z.; Zhou, Y.; Wang, S., The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nat Genet* **2013**, 45, (10), 1168-75.
7. Parkinson, J.; Wasmuth, J. D.; Salinas, G.; Bizarro, C. V.; Sanford, C.; Berriman, M.; Ferreira, H. B.; Zaha, A.; Blaxter, M. L.; Maizels, R. M.; Fernández, C., A transcriptomic analysis of *Echinococcus granulosus* larval stages: implications for parasite biology and host adaptation. *PLoS Negl Trop Dis* **2012**, 6, (11), e1897.
8. Smyth, J. D., Parasites as biological models. *Parasitology* **1969**, 59, (1), 73-91.
9. Thompson, R. C.; Jenkins, D. J., *Echinococcus* as a model system: biology and epidemiology. *Int J Parasitol* **2014**.
10. Cucher, M.; Prada, L.; Mourglia-Ettlin, G.; Dematteis, S.; Camicia, F.; Asurmendi, S.; Rosenzvit, M., Identification of *Echinococcus granulosus* microRNAs and their expression in different life cycle stages and parasite genotypes. *Int J Parasitol* **2011**, 41, (3-4), 439-48.
11. Bai, Y.; Zhang, Z.; Jin, L.; Kang, H.; Zhu, Y.; Zhang, L.; Li, X.; Ma, F.; Zhao, L.; Shi, B.; Li, J.; McManus, D. P.; Zhang, W.; Wang, S., Genome-wide sequencing of small RNAs reveals a tissue-specific loss of conserved microRNA families in *Echinococcus granulosus*. *BMC Genomics* **2014**, 15, 736.
12. Chemale, G.; van Rossum, A. J.; Jefferies, J. R.; Barrett, J.; Brophy, P. M.; Ferreira, H. B.; Zaha, A., Proteomic analysis of the larval stage of the parasite *Echinococcus granulosus*: causative agent of cystic hydatid disease. *Proteomics* **2003**, 3, (8), 1633-6.
13. Monteiro, K. M.; de Carvalho, M. O.; Zaha, A.; Ferreira, H. B., Proteomic analysis of the *Echinococcus granulosus* metacestode during infection of its intermediate host. *Proteomics* **2010**, 10, (10), 1985-99.
14. Virginio, V. G.; Monteiro, K. M.; Drumond, F.; de Carvalho, M. O.; Vargas, D. M.; Zaha, A.; Ferreira, H. B., Excretory/secretory products from in vitro-cultured *Echinococcus granulosus* protoscoleces. *Mol Biochem Parasitol* **2012**, 183, (1), 15-22.
15. Cui, S. J.; Xu, L. L.; Zhang, T.; Xu, M.; Yao, J.; Fang, C. Y.; Feng, Z.; Yang, P. Y.; Hu, W.; Liu, F., Proteomic characterization of larval and adult developmental stages in *Echinococcus granulosus* reveals novel insight into host-parasite interactions. *J Proteomics* **2013**, 84, 158-75.

16. Cain, J. A.; Solis, N.; Cordwell, S. J., Beyond gene expression: the impact of protein post-translational modifications in bacteria. *J Proteomics* **2014**, *97*, 265-86.
17. Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **2005**, *44*, (45), 7342-72.
18. Arnesen, T.; Van Damme, P.; Plevoda, B.; Helsens, K.; Evjenth, R.; Colaert, N.; Varhaug, J. E.; Vandekerckhove, J.; Lillehaug, J. R.; Sherman, F.; Gevaert, K., Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc Natl Acad Sci U S A* **2009**, *106*, (20), 8157-62.
19. Mischerikow, N.; Heck, A. J., Targeted large-scale analysis of protein acetylation. *Proteomics* **2011**, *11*, (4), 571-89.
20. Smith, L. M.; Kelleher, N. L.; Proteomics, C. f. T. D., Proteoform: a single term describing protein complexity. *Nat Methods* **2013**, *10*, (3), 186-7.
21. Černý, M.; Skalák, J.; Cerna, H.; Brzobohatý, B., Advances in purification and separation of posttranslationally modified proteins. *J Proteomics* **2013**, *92*, 2-27.
22. Olsen, J. V.; Mann, M., Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics* **2013**, *12*, (12), 3444-52.
23. Silva, A. M.; Vitorino, R.; Domingues, M. R.; Spickett, C. M.; Domingues, P., Post-translational modifications and mass spectrometry detection. *Free Radic Biol Med* **2013**, *65*, 925-41.
24. Huang, J.; Wang, F.; Ye, M.; Zou, H., Enrichment and separation techniques for large-scale proteomics analysis of the protein post-translational modifications. *J Chromatogr A* **2014**, *1372C*, 1-17.
25. Kellie, J. F.; Tran, J. C.; Lee, J. E.; Ahlf, D. R.; Thomas, H. M.; Ntai, I.; Catherman, A. D.; Durbin, K. R.; Zamdborg, L.; Vellaichamy, A.; Thomas, P. M.; Kelleher, N. L., The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. *Mol Biosyst* **2010**, *6*, (9), 1532-9.
26. Ahlf, D. R.; Thomas, P. M.; Kelleher, N. L., Developing top down proteomics to maximize proteome and sequence coverage from cells and tissues. *Curr Opin Chem Biol* **2013**, *17*, (5), 787-94.
27. Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L., Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011**, *480*, (7376), 254-8.
28. Catherman, A. D.; Durbin, K. R.; Ahlf, D. R.; Early, B. P.; Fellers, R. T.; Tran, J. C.; Thomas, P. M.; Kelleher, N. L., Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol Cell Proteomics* **2013**, *12*, (12), 3465-73.
29. Bowles, J.; Blair, D.; McManus, D. P., Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Mol Biochem Parasitol* **1992**, *54*, (2), 165-73.
30. Alvarez Rojas, C. A.; Romig, T.; Lightowers, M. W., *Echinococcus granulosus sensu lato* genotypes infecting humans--review of current knowledge. *Int J Parasitol* **2014**, *44*, (1), 9-18.
31. Garat, B.; Esperón, P.; Picón, M.; Ehrlich, R., *Echinococcus granulosus*: preparation of protein extracts from protoscolex nuclei for mobility-shift assays. *Parasitol Res* **1998**, *84*, (7), 598-600.
32. Wessel, D.; Flügge, U. I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **1984**, *138*, (1), 141-3.
33. Ntai, I.; Kim, K.; Fellers, R. T.; Skinner, O. S.; Smith, A. D.; Early, B. P.; Savaryn, J. P.; LeDuc, R. D.; Thomas, P. M.; Kelleher, N. L., Applying label-free quantitation to top down proteomics. *Anal Chem* **2014**, *86*, (10), 4961-8.
34. Vellaichamy, A.; Tran, J. C.; Catherman, A. D.; Lee, J. E.; Kellie, J. F.; Sweet, S. M.; Zamdborg, L.; Thomas, P. M.; Ahlf, D. R.; Durbin, K. R.; Valaskovic, G. A.; Kelleher, N. L., Size-

- sorting combined with improved nanocapillary liquid chromatography-mass spectrometry for identification of intact proteins up to 80 kDa. *Anal Chem* **2010**, 82, (4), 1234-44.
35. Wenger, C. D.; Phanstiel, D. H.; Lee, M. V.; Bailey, D. J.; Coon, J. J., COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* **2011**, 11, (6), 1064-74.
36. Conesa, A.; Götz, S.; García-Gómez, J. M.; Terol, J.; Talón, M.; Robles, M., Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, 21, (18), 3674-6.
37. Durbin, K. R.; Fellers, R. T.; Ntai, I.; Kelleher, N. L.; Compton, P. D., Autopilot: an online data acquisition control system for the enhanced high-throughput characterization of intact proteins. *Anal Chem* **2014**, 86, (3), 1485-92.
38. Wu, S.; Brown, R. N.; Payne, S. H.; Meng, D.; Zhao, R.; Tolić, N.; Cao, L.; Shukla, A.; Monroe, M. E.; Moore, R. J.; Lipton, M. S.; Paša-Tolić, L., Top-Down Characterization of the Post-Translationally Modified Intact Periplasmic Proteome from the Bacterium *Novosphingobium aromaticivorans*. *Int J Proteomics* **2013**, 2013, 279590.
39. Leduc, R. D.; Kelleher, N. L., Using ProSight PTM and related tools for targeted protein identification and characterization with high mass accuracy tandem MS data. *Curr Protoc Bioinformatics* **2007**, Chapter 13, Unit 13.6.
40. Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L., ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res* **2007**, 35, (Web Server issue), W701-6.
41. Brown, J. L., The N-terminal region of soluble proteins from procaryotes and eucaryotes. *Biochim Biophys Acta* **1970**, 221, (3), 480-8.
42. Arnesen, T., Towards a functional understanding of protein N-terminal acetylation. *PLoS Biol* **2011**, 9, (5), e1001074.
43. Giglione, C.; Boularot, A.; Meinnel, T., Protein N-terminal methionine excision. *Cell Mol Life Sci* **2004**, 61, (12), 1455-74.
44. Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinnel, T., The proteomics of N-terminal methionine cleavage. *Mol Cell Proteomics* **2006**, 5, (12), 2336-49.
45. Kim, H. K.; Kim, R. R.; Oh, J. H.; Cho, H.; Varshavsky, A.; Hwang, C. S., The N-terminal methionine of cellular proteins as a degradation signal. *Cell* **2014**, 156, (1-2), 158-69.
46. Starheim, K. K.; Gevaert, K.; Arnesen, T., Protein N-terminal acetyltransferases: when the start matters. *Trends Biochem Sci* **2012**, 37, (4), 152-61.
47. Van Damme, P.; Hole, K.; Pimenta-Marques, A.; Helsens, K.; Vandekerckhove, J.; Martinho, R. G.; Gevaert, K.; Arnesen, T., NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation. *PLoS Genet* **2011**, 7, (7), e1002169.
48. LeDuc, R. D.; Taylor, G. K.; Kim, Y. B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L., ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res* **2004**, 32, (Web Server issue), W340-5.
49. Oehring, S. C.; Woodcroft, B. J.; Moes, S.; Wetzel, J.; Dietz, O.; Pulfer, A.; Dekiwadia, C.; Maeser, P.; Flueck, C.; Witmer, K.; Brancucci, N. M.; Niederwieser, I.; Jenoe, P.; Ralph, S. A.; Voss, T. S., Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite *Plasmodium falciparum*. *Genome Biol* **2012**, 13, (11), R108.
50. Lorenzatto, K. R.; Monteiro, K. M.; Paredes, R.; Paludo, G. P.; da Fonsêca, M. M.; Galanti, N.; Zaha, A.; Ferreira, H. B., Fructose-bisphosphate aldolase and enolase from *Echinococcus granulosus*: genes, expression patterns and protein interactions of two potential moonlighting proteins. *Gene* **2012**, 506, (1), 76-84.

Table S3. Low mass proteins from nuclear and cytosolic fractions reported for the first time in *E. granulosus* protoscolexes using top down approach^a

Accession ^b	Protein description	Fragments	E-value	PTMs	Theoretical mass ^c	Observed mass	Δ mass	ppm error
A0A068WTW8	60S ribosomal protein L39	19	6.5E-25		6071.51	N/D ^d	N/A ^e	N/A
A0A068WTW8	60S ribosomal protein L39	9	7.6E-05	N6-acetyl-L-lysine	6197.55	6170.95	-26.59	-4309.49
A0A068WZK8	expressed conserved protein	28	1E-38		6534.68	N/D	N/A	N/A
A0A068W9Y5	expressed protein	10	4.4E-09		7026.05	N/D	N/A	N/A
U6JI91	ribosome associated membrane protein 4	32	1.8E-45	N-acetyl-L-serine	7217.77	7221.57	3.80	526.28
A0A068WW55	hypothetical protein	6	5.3E-06		7268.44	N/D	N/A	N/A
U6JBR8	40S ribosomal protein S28	13	2.2E-12	N-acetyl-L-methionine	7737.23	7815.01	77.77	9952.52
U6J7W8	Calcium binding protein	72	8E-103		7814.96	7622.18	-192.76	-25290.20
U6J7W8	Calcium binding protein	41	1E-42	N-acetyl-L-methionine	7856.97	7861.17	4.21	535.59
U6JNT9	expressed conserved protein	15	3.9E-10	N-acetyl-L-serine	7848.69	7293.47	-555.21	-76124.7
A0A068WHH3	ribosomal protein mitochondrial	9	3.4E-06		7902.26	7611.98	-290.27	-38134.20
A0A068WIB7	ubiquinol cytochrome c oxidoreductase subunit 6	11	1.5E-09	N-acetyl-L-methionine	8015.96	5815.79	-2200.16	-378308
A0A068WS46	expressed conserved protein	22	2.2E-28		8083.92	7292.87	-791.05	-108469
U6JCT9	60S ribosomal protein L38	9	6.4E-09		8150.7	10049.55	1898.86	188949
A0A068WMB8	expressed conserved protein	26	6.8E-32		8383.39	8387.93	4.54	542.09
A0A068WP65	signal recognition particle 9 kDa protein	15	6.8E-14		8522.58	8923.12	400.54	44888.20
A0A068X2D5	expressed conserved protein	11	7.2E-12		8713.64	N/D	N/A	N/A
A0A068X046	expressed conserved protein	13	5.7E-15		8959.93	17420.22	8460.30	485659
A0A068WA87	sec61 beta subunit	25	7.7E-31		9168.79	9173.79	5.00	545.27
A0A068WEY8	F type H transporting ATPase subunit e	31	2.9E-39		9825.32	9831.04	5.72	582.42
U6JIF4	dynein light chain	26	1.5E-29		9833.06	8922.57	-910.47	-102042
U6JG19	LYR motif containing protein 4	10	7E-06	N-acetyl-L-serine	9843.12	12717.02	2873.90	225989
A0A068WXZ7	hypothetical protein	11	4.9E-07		6093.09	7134.41	1041.33	145959
A0A068W7A8	histone lysine N methyltransferase NSD3	9	7.5E-05		7505.73	7505.77	0.04	5.93
A0A068WEA1	expressed protein	25	7.3E-34		7517.08	7520.57	3.49	464.43
A0A068WNL1	expressed protein	5	5.2E-05	N-acetyl-L-methionine	8103.16	N/D	N/A	N/A
U6JCT9	60S ribosomal protein L38	10	1.1E-09	N-acetyl-L-proline	8192.71	N/D	N/A	N/A
A0A068W6R5	expressed protein	5	4.5E-05		9755.13	27471.35	17716.20	644898

^a Gray rows indicate cytosolic proteins; ^b UniProtKB accession numbers; ^c Masses are given in Da; ^dN/D not detected; ^eN/A not applicable.

Table S4. Substrates and likely composition of *E. granulosus* NAT complexes^a

NAT	Accession number^b	Accession number^c	Subunits^d	Substrates^e
NatA	A0A068W8D6	EgrG_000817600	n alpha acetyltransferase 15 NatA auxiliary	Ser-, Ala-, Gly-, Thr-, Val, Cys- N-termini
	U6IUX7	EgrG_000792500	N alpha acetyltransferase 11	
NatB	A0A068WBF3	EgrG_001014800	peptide alpha N acetyltransferase (Naa25)	Met-Glu-, Met-Asp-, Met-Asn-, Met-Gln- N-termini
	A0A068WPC8	EgrG_000252000	n acetyltransferase (Naa20)	
NatC	A0A068WJN0	EgrG_000247600	n acetyltransferase mak3	Met-Ile-, Met-Trp-, Met-Phe- N-termini
	A0A068X103	EgrG_000574500	n acetyltransferase mak3	
	U6JHP0	EgrG_000575600	n acetyltransferase mak3	
	U6JIR1	EgrG_000117400	n alpha acetyltransferase 35 natc auxiliary	
NatD	U6J0S8	EgrG_001036500	n alpha acetyltransferase 40 NatD catalytic	Ser-Gly-Arg- N-termini
NatE	U6JE19	EgrG_000215400	n alpha acetyltransferase 50 NatE catalytic	Met-Ala-, Met-Lys- N-termini
NatF	A0A068WY05	EgrG_000084400	N alpha acetyltransferase 60	Met-Lys-, Met-Ile-, Met-Trp-, Met-Phe- N-termini

^a Eukaryotic NAT complexes and their substrates were previously described; ^{18, 42} ^b UniProtKB accession numbers; ^c GeneDB accession number; ^d *E. granulosus* subunits were determined according to identification of eukaryotic conserved domains in BLAST analyses; ^e Substrates listed were identified in this work and are also conserved in eukaryotes.

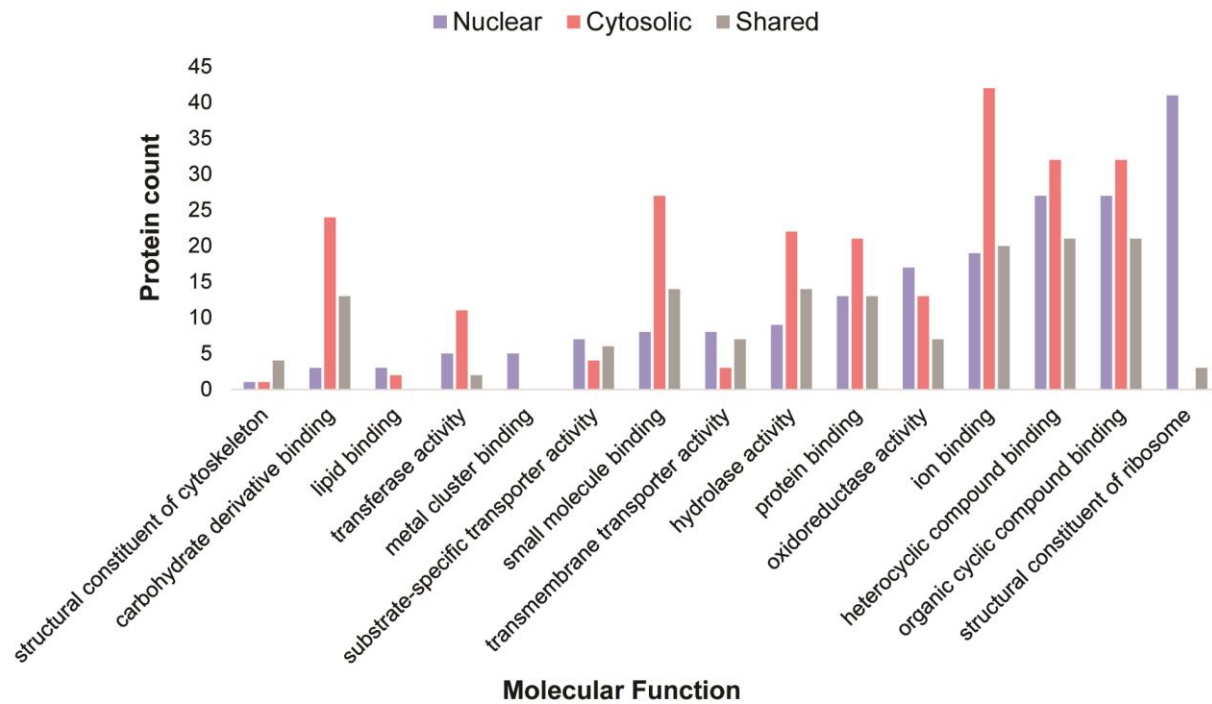
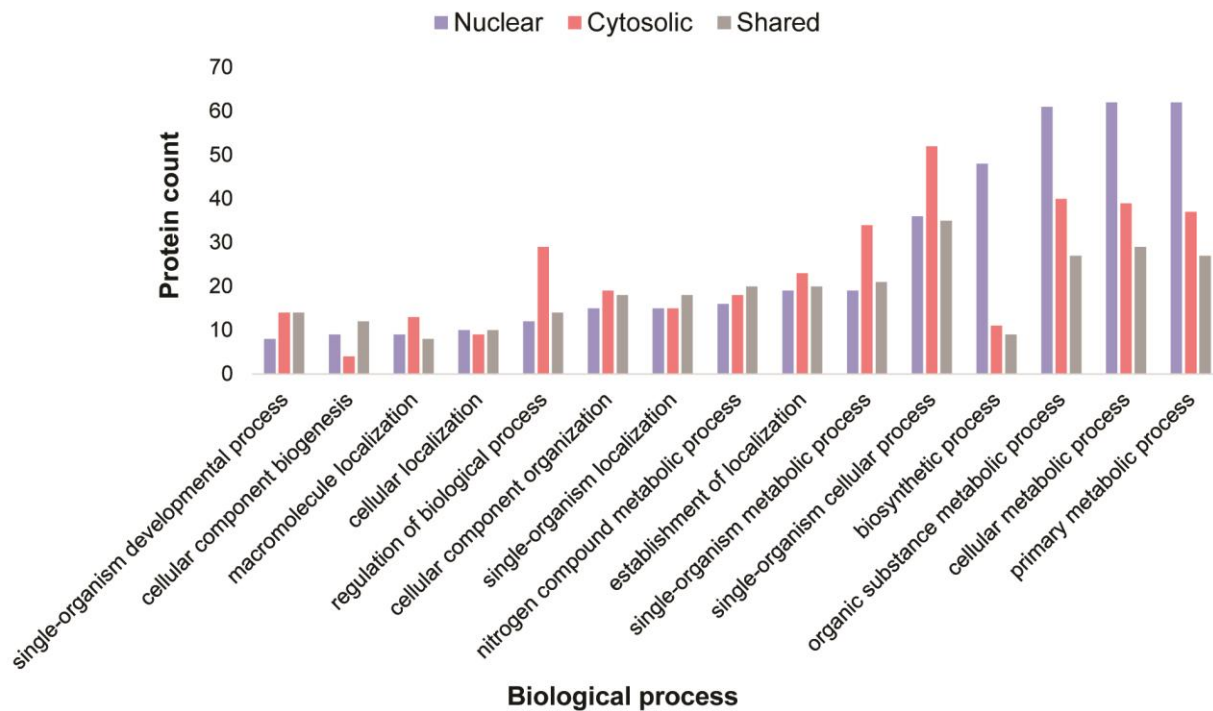
A**B**

Figure S1. Gene ontology (GO) classification of proteins identified in *E. granulosus* subcellular fractions. The results of BLASTP searches against the NCBI nr database were used for GO mapping and annotation. Top 15 most abundant GO terms for molecular function (A) and biological process (B) are shown.

4.4 Análise proteômica *top down* e *bottom up* de frações subcelulares de protoescólices de *E. granulosus* contendo proteínas de maior massa

4.4.1 Materiais e métodos

4.4.1.1 Preparação de frações subcelulares de protoescólices de *E. granulosus*

As proteínas citoplasmáticas e nucleares (obtidas conforme descrito na seção 4.3) foram fracionadas pelo sistema de fracionamento GELFrEE, conforme previamente descrito (Catherman *et al.*, 2013). 250 µg de proteína foram fracionadas no sistema GELFrEE 8100 (Expedeon, San Diego, CA) em cassetes de 10 %. Frações de 1-12 foram coletadas e alíquotas de 6 µL de cada fração de 150 µL foram utilizadas para análise em SDS-PAGE. Para a análise das frações de GELFrEE por *top down*, detergentes e sais foram removidos através da precipitação com clorofórmio/metanol/água, conforme descrito previamente (Wessel & Flügge, 1984). Para a análise por *bottom up*, as frações de GELFrEE foram precipitadas em acetona, preparadas para digestão com tripsina e os peptídeos foram dessalinizados em microcolunas Zip-Tip C18 (Millipore).

4.4.1.2 LC-MS/MS e processamento de dados

Para a análise de *top down*, as proteínas intactas foram analisadas em um Orbitrap Elite (Thermo Fisher Scientific). Os arquivos brutos do espectrômetro de massas foram processados no programa CRAWLER e então analisados no programa ProSightPC 3.0. Dois bancos de dados contendo sequências de aminoácidos deduzidas de sequências genômicas de *E. granulosus*, disponíveis no GeneDB e UniProtKB (Tsai *et al.*, 2013; Zheng *et al.*, 2013), foram utilizados para obtenção das identificações. Para a análise de *bottom up*, os peptídeos foram analisados em um espectrômetro de massas do tipo 12 T LTQ-FT-ICR (Thermo Fisher Scientific). Os arquivos brutos do espectrômetro de massas foram processados no programa COMPASS (Wenger *et al.*, 2011), e então analisados no programa Mascot Server v. 2.3. A classificação funcional das proteínas identificadas foi realizada no BLAST2GO contra o banco de dados NCBI nr. O mapeamento da ontologia gênica (GO) e anotação foram realizados com base nos resultados do Blastp (E-value $\leq 1.0E-3$).

4.4.2 Resultados

4.4.2.1 Análise *top down* e *bottom up* das proteínas de maior massa de frações subcelulares de protoescólicas de *E. granulosus*

As frações nucleares e citosólicas de *E. granulosus* foram submetidas ao fracionamento de GELFREE, gerando 12 frações (3,5 a 100 kDa) de cada amostra (Figura 9). As 12 frações foram submetidas à análise de *bottom up*. A análise das frações 1-6 foi descrita na seção 4.3. A análise das frações 7-12 resultou na identificação de 254 proteínas, sendo 93 exclusivas da frações nucleares e 80 exclusivas das frações citosólicas (Figure 10).

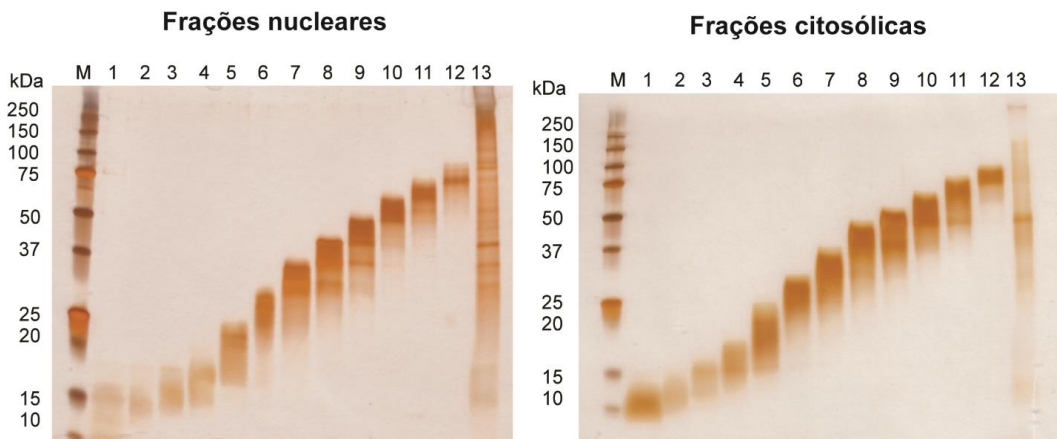


Figure 9. Fracionamento por GELFrEE das amostras nucleares e citosólicas de protoescólicas de *E. granulosus*. M: marcador de massa molecular em kDa; 1-12: frações de GELFrEE; 13: extrato proteico nuclear e citosólico, respectivamente, antes do fracionamento de GELFrEE.

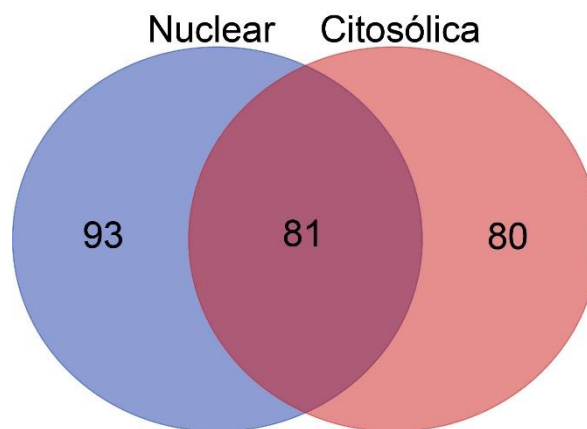


Figure 10. Distribuição das proteínas de maior massa identificadas em frações nucleares e citosólicas de protoescólicas de *E. granulosus* através da abordagem *bottom up*.

Com o intuito de detectar proteoformas da EgEno1 e EgFBA1, frações de GELFrEE, contendo estas proteínas, foram submetidas à análise proteômica *top down*. A seleção das frações nas quais estas proteínas estavam mais abundantes foi feita com base nos dados de *bottom up*. Desse modo, foram inicialmente selecionadas as frações citoplasmáticas de 7-9 (30 a 50 kDa) para a análise proteômica *top down*. Nesta análise, 17 proteínas intactas foram identificadas (Tabela 2). A detecção da EgFBA1 não foi possível e, embora a EgEno1 tenha sido detectada, a cobertura obtida para esta proteína foi baixa. Considerando a abundância relativamente alta da EgEno1 e da EgFBA1 demonstrada na análise *bottom up*, o insucesso na detecção destas proteínas por *top down* foi atribuído à fragmentação ineficiente das mesmas.

Tabela 2. Proteínas citosólicas identificadas a partir da análise das frações de GELFrEE 7, 8 e 9 pela abordagem *top down*

Número de acesso ^a	Descrição da proteína ^b	MPTs	Íons b	Íons y	E-value	Massa teórica ^c
EgrG_000084400	N-acetiltransferase 60 alfa		1	3	0,0000177	37064,9
EgrG_000102600	Fator de alongamento Tu		2	2	0,0000502	49096,2
EgrG_000105100	Proteína expressa conservada		3	1	0,0000294	21653,1
EgrG_000212700	<i>Major egg antigen p40</i>		1	14	2,16E-23	35694,1
EgrG_000212700	<i>Major egg antigen p40</i>	N-acetil-L-serina	3	10	2,05E-19	35736,1
EgrG_000212700	<i>Major egg antigen p40</i>	N-acetil-L-metionina	1	16	1,1E-21	35867,1
EgrG_000231000	Galactosylgalactosylxylosylprotein		2	3	3,05E-06	30783
EgrG_000231300	14-3-3.2	N-acetil-L-alanina	3	2	0,0000147	29110,7
EgrG_000303700	Proteína expressa		2	2	0,0000423	45549,1
EgrG_000417100	Malato-desidrogenase citosólica		1	6	2,73E-009	36415,7
EgrG_000514200	Enolase	N-acetil-L-serina	2	2	0,0000357	46442,8
EgrG_000516800	Ubiquitina ribossômica L40		1	5	1,15E-06	25600,8
EgrG_000535600	Ribonucleoproteína		2	2	3,47E-06	7485,74
EgrG_000608500	Lactato-desidrogenase	N-acetil-L-metionina	5	0	8,79E-07	34998,1
EgrG_000610400	Translocador ADP-ATP		4	0	0,0000174	37714,6
EgrG_000703300	Actina	N-acetilglicine	5	1	8,02E-10	41628,8
EgrG_000706000	Proteína hipotética	N-acetil-L-metionina	2	4	8,27E-06	16257,1
EgrG_000993100	Proteína hipotética		1	3	0,0000795	77588,2
EgrG_001074700	Proteína hipotética	N-acetil-L-treonina	2	2	0,0000742	4069,94

^a Número de acesso do GeneDB ; ^b Descrição da proteína conforme o GeneDB; ^c A massa teórica é dada em Da.

4.4.2.1 Classificação funcional do proteoma de frações nucleares e citosólicas de protoescólicas de *E. granulosus*

A combinação dos dados de *top down* (até 30 kDa) e *bottom up* (3,5 a 100 kDa) resultou na identificação de 525 proteínas, das quais 224 e 156 foram exclusivamente identificadas nas frações nucleares e citosólicas, respectivamente (Figura 11A). A análise de enriquecimento de termos do GO foi utilizada para fazer a classificação funcional das proteínas identificadas. Anotações de sequências associadas com 440 proteínas (84%)

foram categorizadas nas três principais ontologias do GO: processos biológicos, funções moleculares e componentes celulares (Figura 11B). Os termos do GO estão resumidos em 79, 81 e 88 subcategorias para as proteínas nucleares, citosólicas e compartilhadas, respectivamente, com classificações de nível 3 do GO. 85 proteínas não foram associadas a termos do GO e a maioria delas permanece anotada como não caracterizada.

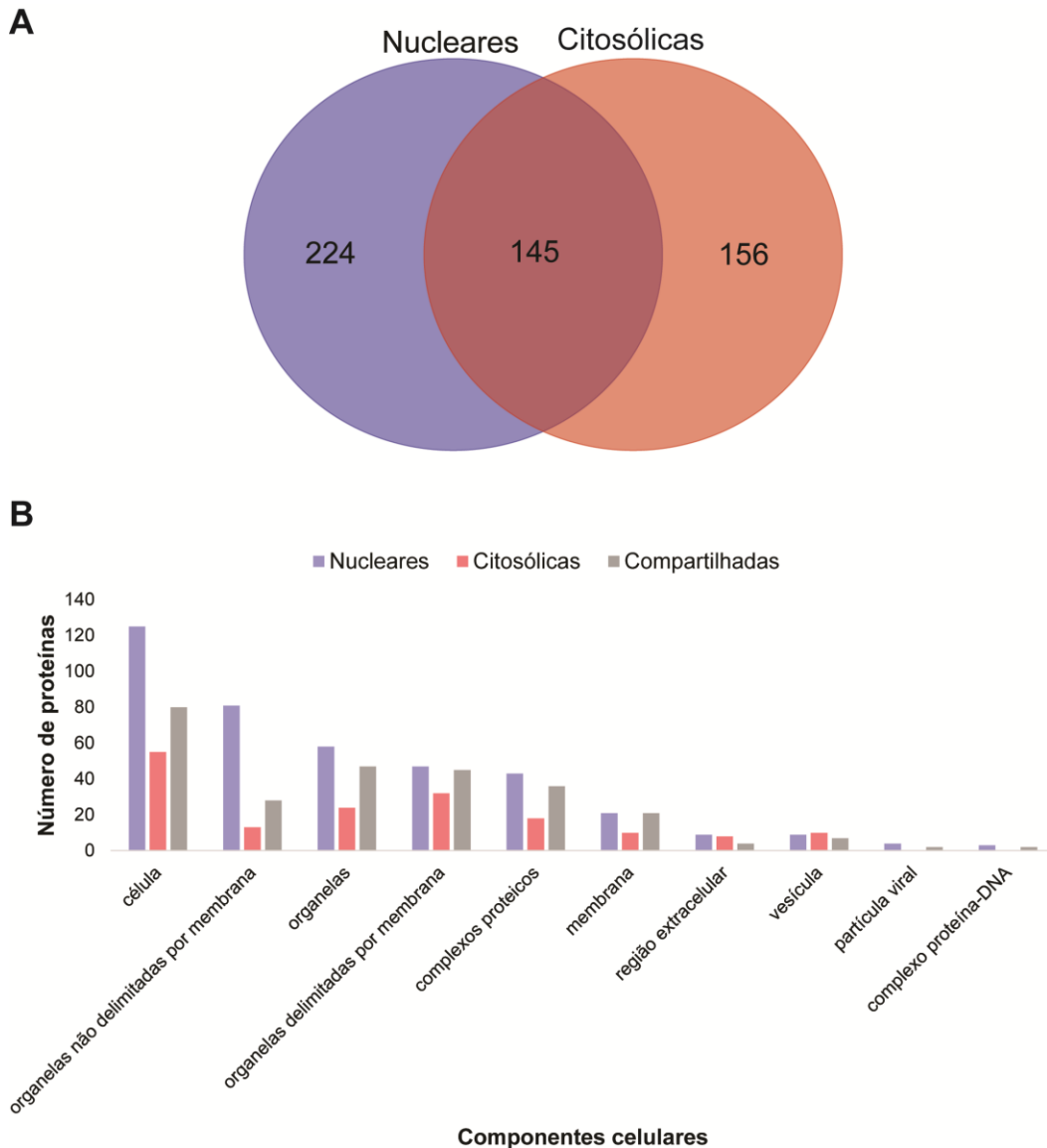


Figure 11. Visão global do proteoma de frações nucleares e citosólicas de protoescólices de *E. granulosus*. (A) Combinação dos dados de proteômica de baixa e maior massa e distribuição das proteínas identificadas em frações nucleares e citosólicas de protoescólices de *E. granulosus*. (B) As proteínas identificadas nas frações subcelulares de *E. granulosus* foram submetidas à classificação funcional do GO. O número de proteínas classificadas de acordo com as 10 subcategorias do GO mais abundantes para componentes celulares é mostrado.

5. DISCUSSÃO

Duas abordagens foram realizadas para a caracterização de proteoformas expressas no estágio larval de *E. granulosus*. Primeiramente, uma abordagem dirigida para duas enzimas glicolíticas potencialmente multifuncionais foi realizada, com o intuito de evidenciar tanto mecanismos associados à multifuncionalidade como potenciais funções não glicolíticas destas enzimas. Em segundo lugar, uma abordagem abrangente para a caracterização do repertório de proteoformas expressas em frações subcelulares de protoescolices de *E. granulosus* também foi explorada, possibilitando não somente a identificação de proteínas nucleares e citosólicas, mas também a caracterização de MCTs e MPTs nas proteínas identificadas.

A via glicolítica está entre uma das mais antigas redes metabólicas (Canback *et al.*, 2002). Vários estudos têm mostrado que além da via glicolítica, algumas destas enzimas são proteínas multifuncionais, desempenhando funções não glicolíticas importantes em vários processos biológicos e patofisiológicos (Díaz-Ramos *et al.*, 2012). A FBA e a enolase são exemplos de proteínas multifuncionais, tendo sido descritas com funções não glicolíticas em diferentes contextos biológicos, incluindo o de parasitismo (Marcos *et al.*, 2014; Ramajo-Hernández *et al.*, 2007; Tunio *et al.*, 2010).

Em vertebrados, existem três isoformas de aldolase (A, B e C) e três isoformas de enolase (alfa, beta e gama), sendo cada uma delas codificada por genes diferentes (Steinke *et al.*, 2006). Em *E. granulosus*, a análise de predição de genes indicou a presença de quatro genes de FBA, *EgFBA1-4*, e três genes de enolase, *EgEno1-3*. De acordo com estudos transcricionais e proteômicos de amostras de cistos hidáticos, a *EgFBA1* e *EgEno1* são isoformas predominantemente expressas em todos os estágios do ciclo de vida do parasito (Cui *et al.*, 2013; Monteiro *et al.*, 2010; Parkinson *et al.*, 2012; Zheng *et al.*, 2013). Estudos de RNA-seq detectaram *EgFBA2-4* somente no estágio adulto, enquanto *EgEno2-3* foram detectadas nos estágios larval e adulto (Zheng *et al.*, 2013) (Figura 12).

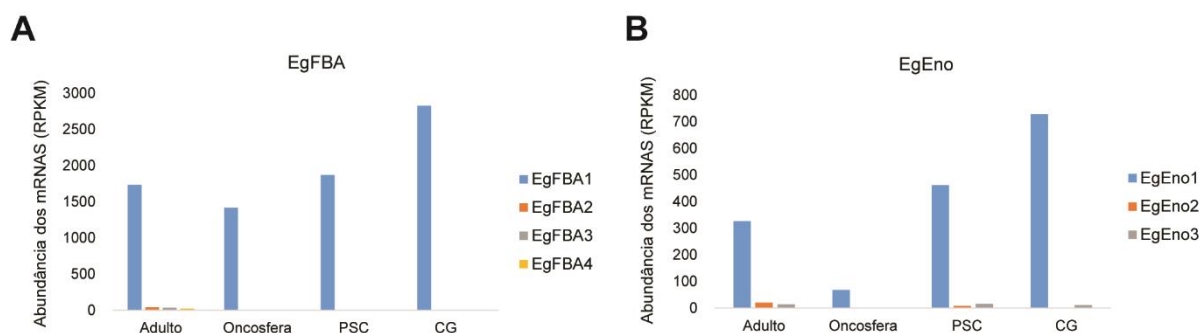


Figure 12. Perfis de transcrição das isoformas de EgFBA e EgEno nos diferentes estágios do ciclo de vida do *E. granulosus*. A abundância dos transcritos é dada em *reads per kilobase per million mapped reads* (RPKM). PSC: protoescólicas; CG: camada germinativa. Modificado de (Zheng *et al.*, 2013).

As filogenias das isoformas de FBA e enolase sugerem que a evolução das EgFBA1-4 e de EgEno1-3 é complexa. As diferentes isoformas de FBA e enolase de helmintos são divergentes das isoformas de vertebrados, de modo que não foi possível enquadrá-las na mesma classificação utilizada para as FBAs e enolases de vertebrados. Além disso, a presença de múltiplas isoformas de FBA e enolase não foi observada em trematódeos. A aparente ausência de ortólogos de EgFBA2-4 e EgEno2-3 não somente em vertebrados, mas também em trematódeos, poderia ser explicada por eventos de duplicação específicos da classe Cestoda. Estudos adicionais são necessários para entender não somente a evolução destes genes, mas também a função dos seus produtos proteicos nestes parasitos.

As atividades catalíticas da FBA e enolase na via glicolítica consistem na atividade primária destas enzimas. A FBA está envolvida na quarta reação desta via, catalisando a clivagem reversível da fru-1,6-P2 em diidroxiacetona fosfato e gliceraldeído 3-fosfato. A nona reação dessa via é realizada pela enolase, a qual catalisa a transformação reversível entre 2-PGA e fosfoenolpiruvato. Com o intuito de verificar se a EgFBA1 e EgEno1 são cataliticamente ativas, ensaios de atividade enzimática foram realizados com a rEgFBA1 e a rEgEno1. As duas enzimas recombinantes foram capazes de hidrolisar seus substratos, o que indica que as versões nativas também sejam enzimaticamente ativas.

Além de participar da glicólise, a FBA foi descrita como ligante a bombas de H⁺-ATPase vacuolares em *S. cerevisiae* (Lu *et al.*, 2007). Ela está envolvida na endocitose (Rangarajan *et al.*, 2010), e na adesão de *Neisseria meningitis* a células humanas (Tunio *et*

al., 2010). Além disso, a FBA foi descrita como intermediadora da ligação de filamentos de actina tanto a proteínas TRAP em *Plasmodium falciparum* (Heiss *et al.*, 2008) como a adesina microneme (MIC2) em *Toxoplasma gondii* (Sheiner *et al.*, 2010; Starnes *et al.*, 2009). No caso da enolase, o seu envolvimento na regulação da transcrição gênica em uma ampla variedade de organismos foi descrito. Além disso, ela foi identificada como uma proteína estrutural do cristalino de vertebrados (Wistow *et al.*, 1988), está envolvida na fusão da membrana de vacúolos e tráfego de proteínas para o vacúolo em *S. cerevisiae* (Decker & Wickner, 2006), e desempenha função de adesina em *Streptococcus pneumoniae* (Bergmann *et al.*, 2013). A habilidade de ligação ao plasminogênio da FBA e da enolase tem sido cada vez mais descrita em diversas bactérias, leveduras e helmintos parasitas (Ginger, 2014; Gómez-Arreaza *et al.*, 2014; Henderson, 2014; Karkowska-Kuleta & Kozik, 2014).

Na tentativa de encontrar evidências de possíveis funções *moonlighting* para a EgFBA1 e a EgEno1, estas proteínas foram imunolocalizadas em componentes do estágio larval do *E. granulosus*. A utilização de soros específicos permitiu detectar a presença da EgFBA1 e EgEno1 em diferentes compartimentos do cisto hidático, incluindo frações tegumentares e produtos de ES. Na parede do cisto, a distribuição da EgEno1 chama atenção devido a sua presença não somente na camada germinativa mas também na camada laminar. A detecção da EgEno1 na camada laminar é indicativa do seu envolvimento direto com mecanismos de sobrevivência na interface parasito-hospedeiro, uma vez que a camada laminar protege as células da camada germinativa da resposta imune do hospedeiro (Díaz *et al.*, 2011).

Igualmente notável é a presença da EgFBA1 e da EgEno1 em produtos de ES. Ambas foram detectadas em estudos proteômicos de líquido hidático (Ahn *et al.*, 2015; Aziz *et al.*, 2011; Monteiro *et al.*, 2010) e em produtos de ES *in vitro* de protoescólices (Virginio *et al.*, 2012). Aqui, a presença destas proteínas no líquido hidático e em produtos de ES de cultivos *in vitro* de protoescólices também foi confirmada pelos experimentos de imunoblot. A detecção da FBA e da enolase em produtos de ES também já foi descrita em outros parasitos é sugestiva de funções diretamente relacionadas com a interação parasito-hospedeiro (Ditgen *et al.*, 2014; Harnett, 2014).

A identificação de proteínas citosólicas em produtos de ES tem se demonstrado frequente e, por isso, tem despertado interesse a respeito dos mecanismos associados com

o direcionamento extracelular destas proteínas. Entre os mecanismos que poderiam explicar a localização destas proteínas em produtos de ES estão o *turnover* do tegumento (Liu *et al.*, 2009; Marcilla *et al.*, 2007) e/ou vesículas extracelulares (VEs) (Maizels *et al.*, 2012; Marcilla *et al.*, 2012; Nono *et al.*, 2012). As VEs são vesículas que contêm uma mistura complexa de proteínas, lipídeos, glicanos e informação genética (Twu & Johnson, 2014). Elas têm sido descritas em bactérias, protozoários e helmintos parasitas (Kim *et al.*, 2015; Montaner *et al.*, 2014). A maioria das proteínas identificadas nas VEs correspondem aquelas descritas previamente nos secretomas destes parasitos (Maizels *et al.*, 2012; Marcilla *et al.*, 2012). Este mecanismo explica a presença de proteínas atípicas nos produtos de ES, como a FBA e a enolase, nas quais peptídeos-sinal de secreção estão ausentes. Desse modo, as VEs também devem constituir um importante mecanismo de secreção de proteínas em *E. granulosus*. O isolamento e caracterização de VEs deste parasito ainda não foram realizados mas estão em andamento no nosso grupo de pesquisa.

Um outro mecanismo descrito para que uma proteína *moonlighting* altere suas funções é através da formação de diferentes complexos proteicos (Jeffery, 2004). No caso da EgFBA1 e EgEno1, tanto características estruturais como ligantes proteicos suportam a multifuncionalidade destas proteínas. O sítio de ligação à actina conservado na EgFBA1 é sugestivo de interações com o citoesqueleto, o qual é consistente com os resultados de interação proteína-proteína nos quais a actina foi detectada como uma proteína interagente da EgFBA1. A capacidade de ligação à actina já foi demonstrada em FBAs tanto de organismos não parasitas (Wang *et al.*, 1996) como em organismos parasitas (Starnes *et al.*, 2009), e em parasitas protozoários, a aldolase foi descrita como membro de um complexo multiproteico conectando adesinas ao citoesqueleto, fornecendo um modelo que liga adesão à motilidade (Jewett & Sibley, 2003; Starnes *et al.*, 2009). No caso da EgEno1, domínios transmembrana foram preditos em um estudo estrutural (Gan *et al.*, 2010), cuja funcionalidade é corroborada por nossos ensaios de imunolocalização e também pela detecção de uma proteína tegumentar como ligante da rEgEno1.

Além das proteínas acima mencionadas, várias outras, não relacionadas ao metabolismo, foram recuperadas como interagentes da rEgFBA1 e da rEgEno1. O repertório de proteínas de interação da rEgFBA1 é totalmente compartilhado com a rEgEno1, com exceção da proteína inibidora da fosfatase 2A, a qual é sugestiva do envolvimento da EgFBA1 em processos de sinalização. No caso da rEgEno1, o repertório

de proteínas interagentes foi maior, de modo que 7 proteínas foram detectadas somente quando essa proteína foi utilizada como isca. Além da proteína tegumentar já mencionada, a qual pode estar envolvida em processos do sistema imune, se destacam uma profilina, uma proteína reguladora de quinase e a ubiquitina, as quais são sugestivas do envolvimento da EgEno1 em processos de sinalização. De modo geral, a capacidade de interação da EgFBA1 e, especialmente da EgEno1 com várias proteínas não glicolíticas é sugestiva de funções *moonlighting*.

Na tentativa de identificar interações proteicas da enolase conservadas, redes de interação proteína-proteína (IPP) foram projetadas para *H. sapiens*, *D. melanogaster*, *C. elegans*, e *S. cerevisiae*. Esta análise permitiu a identificação de algumas interações conservadas da enolase em eucariotos (Paludo *et al.*, 2015). Proteínas como actina, fator de alongamento 1 alfa, Hsp70, tioredoxina peroxidase e ubiquitina são pelo menos parcialmente conservadas nas redes de IPP dos organismos modelo e, estas proteínas também foram detectadas como ligantes da EgEno1. Estas semelhanças nas proteínas de interação sugerem que pelo menos uma parcela das interações da EgEno1 aparecem evolutivamente conservadas, podendo estas estarem relacionadas a funções *moonlighting* conservadas da enolase. Além disso, a identificação de interações da EgEno1 com proteínas não relacionadas à glicólise e não identificadas nas redes de IPP, podem possuir relevância em aspectos espécie-específicos e/ou relacionadas ao parasitismo.

Duas isoformas de 14-3-3 (14-3-3.1 e 14-3-3.3), por exemplo, foram detectadas como ligantes da EgFBA1 e da EgEno1 tanto aqui (ver seção 4.2.2.3), como no recente estudo de caracterização de proteínas 14-3-3 de *E. granulosus* (Teichmann *et al.*, 2015). Proteínas 14-3-3 desempenham um papel central na transdução de sinal intracelular, especialmente devido ao seu amplo interatoma e à habilidade específica para reconhecer sítios fosforilados (Aitken, 2006; Mackintosh, 2004). Além disso, as 14-3-3 desempenham atividades *moonlighting* como chaperonas (Sluchanko *et al.*, 2014; Sluchanko & Uversky, 2015). O fato desta proteína não ter sido detectada como interação conservada nas redes de IPP da enolase é sugestivo de que estas interações da EgEno1 com 14-3-3 estejam relacionadas com funções *moonlighting* relevantes no contexto parasito-hospedeiro.

A presença de MPTs também têm sido descrita como um dos fatores determinantes da multifuncionalidade de proteínas (Pal-Bhowmick *et al.*, 2007). Para isso, foram realizados experimentos de proteômica *bottom up* e *top down* visando à identificação e caracterização

de proteoformas das enzimas EgEno1 (47 kDa) e EgFBA1 (39 kDa). Pela análise de *top down*, a detecção da EgFBA1 não foi possível. Uma proteoforma de EgEno1 foi detectada com poucos fragmentos, os quais permitiram pelo menos a detecção de CTMs. O insucesso na detecção destas proteínas por *top down* foi atribuído à fragmentação ineficiente das mesmas, o que deve estar relacionado a suas massas moleculares relativamente elevadas, uma vez que proteínas intactas acima de 30 kDa ainda representam um desafio para identificação em amostras complexas neste tipo de abordagem. Modos alternativos de fragmentação deverão ser testados com as versões recombinantes da EgEno1 e a EgFBA1 para definição daquele(s) que favoreça(m) a detecção de proteoformas destas enzimas em abordagens de *top down*.

Além da caracterização da EgFBA1 e EgEno1, a caracterização de outras proteoformas expressas em frações subcelulares de protozoários de *E. granulosus* através de diferentes abordagens proteômicas também foi explorada. Tendo em vista as dificuldades de detecção e caracterização de proteoformas de maior massa molecular, a abordagem de proteômica *top down* foi aplicada para proteínas com massas moleculares inferiores a 30 kDa. Já a abordagem *bottom up* foi aplicada tanto para proteínas com massas moleculares inferiores e superiores a 30 kDa. O uso destas abordagens proteômicas possibilitou não somente a identificação dos repertórios de proteínas expressas no núcleo e citoplasma, mas também a caracterização de MCTs e MPTs.

Um dos tipos de MCTs mais abundantes identificados nas proteoformas nucleares e citosólicas de *E. granulosus* em nossa abordagem de *top down* foi a excisão da metionina N-terminal (EMN). A metionina é utilizada para iniciar a síntese proteica em quase todas as proteínas, e é subsequentemente removida na maioria dos casos, seja através da clivagem de um peptídeo sinal N-terminal ou pela ação específica de metionina aminopeptidases (MAPs) (Bonissone *et al.*, 2013). A EMN é um processo necessário para o adequado funcionamento das células, de modo que a deleção da MAP1 é letal em bactérias e a deleção da MAP1 e da MAP2 é letal em leveduras (Hutchison *et al.*, 1999; Li & Chang, 1995).

As MAPs estão presentes em todos os organismos e, cerca de 80% das proteínas de qualquer proteoma são substratos potenciais para a EMN (Giglione *et al.*, 2004). O segundo amino ácido nos substratos proteicos, ou P2, é crucial para a EMN, pois a especificidade das MAPs depende principalmente da natureza desse resíduo, sendo que

esta seletividade é conservada ao longo da escala evolutiva (Frottin *et al.*, 2006; Meinnel *et al.*, 2006). Estas enzimas geralmente realizam a EMN quando o segundo resíduo é alanina, serina, treonina, glicina, cisteína, prolina ou valina, os quais são amino ácidos com menor tamanho (Frottin *et al.*, 2006; Meinnel & Giglione, 2008). De acordo com os resultados obtidos aqui, o resíduo P2 foi geralmente consistente com a regra de EMN, visto que esses resíduos representam 86% das proteínas com EMN. Entre os demais substratos encontrados estão arginina, asparagina, ácido glutâmico, glutamina, isoleucina, leucina, lisina e triptofano. Esta é a primeira vez que a EMN é descrita em proteínas de *E. granulosus* e a detecção dela fornece evidências indiretas de atividade das MAPs. Embora alguns substratos fogem à regra de EMN (Frottin *et al.*, 2006), a possibilidade de eles representarem exceções deste parasito não pode ser descartada. A especificidade de substratos das MAPs de *E. granulosus* precisa ser adicionalmente estudada.

A acetilação N-terminal (AcN) foi outra modificação abundantemente encontrada nas proteínas de *E. granulosus*. Ela é uma modificação irreversível, que ocorre durante a síntese de proteínas, sendo catalisada por N-acetiltransferases (NATs) associadas aos ribossomos (Polevoda *et al.*, 2008). Existem várias NATs em eucariotos (NatA-NatF), sendo cada uma delas composta por uma ou mais subunidades, e a acetilação da extremidade N-terminal também depende da sequência dos primeiros aminoácidos (Polevoda *et al.*, 2009). De modo geral, os padrões de AcN e os complexos de NATs também parecem ser conservados em *E. granulosus*. Sítios N-acetilados detectados nas proteínas deste parasito sugerem a atividade de todos os complexos de NATs, incluindo da NatF, pensada estar presente somente em eucariotos superiores (Arnesen, 2011). Além da evidência indireta a partir dos substratos, uma NatF foi detectada na análise *top down* de proteínas de maior massa.

Além das CTMs, a abordagem de proteômica *top down* utilizada possibilitou a detecção de acetilação de lisinas. Esta MPT pode afetar várias funções proteicas, através da regulação de interações de proteínas com ácidos nucleicos ou com outras proteínas, da atividade enzimática e da localização subcelular (Choudhary *et al.*, 2014). A acetilação foi detectada em 9 proteoformas, destacando as histonas como proteoformas diferencialmente acetiladas. A caracterização de MPTs através da proteômica *top down* é enormemente influenciada pela sua descrição no banco de dados, de modo que a detecção destas proteoformas só foi possível quando um banco de dados modificado, contendo

informações de predição de acetilação, foi utilizado. Mesmo que a contribuição do nosso banco de dados customizado tenha sido limitada a poucas proteoformas, a detecção delas foi dependente do banco de dados, antecipando o impacto que um banco mais completo pode ter na determinação de proteoformas modificadas de *E. granulosus*. Dada a importância não somente da acetilação, mas de várias outras MPTs na regulação de processos biológicos, estudos futuros explorando a caracterização delas certamente contribuirão para a elucidação de mecanismos moleculares deste parasito.

Para a obtenção do perfil geral de proteínas expressas nas frações subcelulares de *E. granulosus*, os dados de *top down* e *bottom up* foram combinados e as identificações de frações nucleares e citosólicas foram submetidas à classificação funcional através do GO. Nas frações nucleares, componentes celulares como célula e organelas estão enriquecidos com proteínas do nucleossomo e spliceossomo. Na subcategoria organelas delimitadas por membrana, além de proteínas do envelope nuclear, várias proteínas associadas a membranas de outras organelas foram detectadas, as quais são reconhecidas como principais fontes de contaminação em preparações nucleares (Oehring *et al.*, 2012). Por outro lado, enzimas glicolíticas e envolvidas em estresse oxidativo estão enriquecidas na fração citosólica.

Várias proteínas nucleares foram detectadas na fração citosólica e várias proteínas citosólicas foram detectadas na fração nuclear, indicando que muitas proteínas de *E. granulosus* transitam entre núcleo e citosol e/ou realizam suas funções em múltiplas localizações. Algumas enzimas glicolíticas, como enolase e GAPDH foram as que apresentaram maior contagem espectral na fração nuclear. Considerando que algumas enzimas glicolíticas foram exclusivamente detectadas no citosol, a presença destas enzimas no núcleo não deve ser efeito unicamente da contaminação de proteínas citosólicas na fração nuclear. Em outros organismos, tanto a localização nuclear quanto funções nucleares para estas enzimas foram descritas. Em parasitos, a enolase já foi descrita como reguladora transcricional e moduladora da atividade de uma metiltransferase de DNA e RNA (Mouveaux *et al.*, 2014; Ruan *et al.*, 2015; Tovy *et al.*, 2010). A GAPDH também possui funções nucleares, incluindo controle transcricional, reconhecimento de nucleotídeos incorporados erroneamente ao DNA e manutenção da estrutura telomérica (He *et al.*, 2013; Pariona-Llanos *et al.*, 2015; Tristan *et al.*, 2011).

Várias proteínas detectadas nas frações nucleares e citosólicas não foram associadas

a termos do GO e a maioria delas permanece anotada como não caracterizada. Tanto no núcleo quanto no citosol, foram detectadas proteínas com ortólogos restritos a helmintos, cestódeos, *Echinococcus* spp. e, até mesmo, restritos à espécie *E. granulosus* s.s. Estas proteínas podem conter novos domínios associados a funções nucleares e citosólicas desconhecidas não somente em *Echinococcus* spp. mas também em cestódeos e/ou helmintos, tornando elas alvos interessantes para estudos futuros. A caracterização funcional de proteínas como estas deve contribuir não somente para a elucidação de novos aspectos da biologia destes parasitos mas também podem revelar novos alvos para utilização em vacinação e/ou desenvolvimento de novas drogas antihelmínticas.

6. CONCLUSÕES E PERSPECTIVAS

A caracterização de proteoformas expressas no estágio larval de *E. granulosus* revelou aspectos importantes para a produção de variabilidade proteica neste parasito, incluindo evidências da presença de proteínas *moonlighting* bem como a presença de proteínas contendo MCTs. As localizações não usuais da EgFBA1 e da EgEno1 no tegumento de protoescólices e em produtos de ES *in vivo* e *in vitro* e, a capacidade destas proteínas de interagir com várias proteínas não relacionadas à glicólise são sugestivas de que estas proteínas sejam *moonlighting*. A caracterização de MCTs em proteoformas expressas em frações nucleares e citosólicas do estágio larval de *E. granulosus* foi realizada pela primeira vez e, a proporção de proteoformas contendo MCTs está de acordo com a proporção observada em outros eucariotos, indicando que este tipo de modificação é conservado em *E. granulosus*. Embora o efeito das MCTs sobre as proteoformas ainda seja pouco compreendido, a localização subcelular diferencial de algumas proteoformas contendo diferentes MCTs é indicativa de que as MCTs podem estar envolvidas na localização subcelular de algumas proteoformas. Além disso, as proteoformas identificadas em frações nucleares e citosólicas de protoescólices de *E. granulosus* demonstram que além de proteínas exclusivamente expressas nestes compartimentos, algumas proteínas são detectadas em ambos compartimentos, indicando que elas teriam capacidade de transitar entre os dois compartimentos. Estudos futuros explorando outras proteínas potencialmente multifuncionais e também proteoformas resultantes de outras MPTs devem revelar novos mecanismos moleculares deste parasito. As perspectivas deste trabalho incluem: (i) a elucidação dos papéis desempenhados pela EgFBA1 e EgEno1 na interação direta com o hospedeiro, através de ensaios de ligação a macromoléculas relevantes para o processo de infecção, como o plasminogênio; (ii) a caracterização do papel das CTMs na localização subcelular de proteoformas do estágio larval de *Echinococcus granulosus*; (iii) a caracterização adicional de MPTs nas proteínas de *Echinococcus granulosus*, incluindo fosforilação, acetilação e metilação, as quais podem estar envolvidas na regulação de diferentes mecanismos moleculares do parasito.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- AHLF, D.R., THOMAS, P.M., & KELLEHER, N.L. (2013). Developing top down proteomics to maximize proteome and sequence coverage from cells and tissues. *Curr Opin Chem Biol* 17, 787-794.
- AHN, C.S., HAN, X., BAE, Y.A., MA, X., KIM, J.T., CAI, H., YANG, H.J., KANG, I., WANG, H., & KONG, Y. (2015). Alteration of immunoproteome profile of *Echinococcus granulosus* hydatid fluid with progression of cystic echinococcosis. *Parasit Vectors* 8, 10.
- AITKEN, A. (2006). 14-3-3 proteins: a historic overview. *Semin Cancer Biol* 16, 162-172.
- ALVAREZ ROJAS, C.A., GAUCI, C.G., & LIGHTOWLERS, M.W. (2013). Antigenic differences between the EG95-related proteins from *Echinococcus granulosus* G1 and G6 genotypes: implications for vaccination. *Parasite Immunol* 35, 99-102.
- ALVAREZ ROJAS, C.A., ROMIG, T., & LIGHTOWLERS, M.W. (2014). *Echinococcus granulosus sensu lato* genotypes infecting humans--review of current knowledge. *Int J Parasitol* 44, 9-18.
- ARNESEN, T. (2011). Towards a functional understanding of protein N-terminal acetylation. *PLoS Biol* 9, e1001074.
- AZIZ, A., ZHANG, W., LI, J., LOUKAS, A., MCMANUS, D.P., & MULVENNA, J. (2011). Proteomic characterisation of *Echinococcus granulosus* hydatid cyst fluid from sheep, cattle and humans. *J Proteomics* 74, 1560-1572.
- BERGMANN, S., SCHOENEN, H., & HAMMERSCHMIDT, S. (2013). The interaction between bacterial enolase and plasminogen promotes adherence of *Streptococcus pneumoniae* to epithelial and endothelial cells. *Int J Med Microbiol* 303, 452-462.
- BERGMEYER, H.U. (1974). *Methods in Enzymatic Analysis*, 2 edn (Academic Press).
- BERNAL, D., DE LA RUBIA, J.E., CARRASCO-ABAD, A.M., TOLEDO, R., MAS-COMA, S., & MARCILLA, A. (2004). Identification of enolase as a plasminogen-binding protein in excretory-secretory products of *Fasciola hepatica*. *FEBS Lett* 563, 203-206.
- BONISSONE, S., GUPTA, N., ROMINE, M., BRADSHAW, R.A., & PEVZNER, P.A. (2013). N-terminal protein processing: a comparative proteogenomic analysis. *Mol Cell Proteomics* 12, 14-28.
- BRUNETTI, E., & WHITE, A.C. (2012). Cestode infestations: hydatid disease and cysticercosis. *Infect Dis Clin North Am* 26, 421-435.
- BUDKE, C.M., DEPLAZES, P., & TORGERSON, P.R. (2006). Global socioeconomic impact of cystic echinococcosis. *Emerg Infect Dis* 12, 296-303.
- CANBACK, B., ANDERSSON, S.G., & KURLAND, C.G. (2002). The global phylogeny of glycolytic enzymes. *Proc Natl Acad Sci U S A* 99, 6097-6102.
- CATHERMAN, A.D., DURBIN, K.R., AHLF, D.R., EARLY, B.P., FELLERS, R.T., TRAN, J.C., THOMAS, P.M., & KELLEHER, N.L. (2013). Large-scale top-down proteomics of the human

- proteome: membrane proteins, mitochondria, and senescence. *Mol Cell Proteomics* 12, 3465-3473.
- CHAVES, E.G., WEBER, S.S., BÁO, S.N., PEREIRA, L.A., BAILÃO, A.M., BORGES, C.L., & SOARES, C.M. (2015). Analysis of *Paracoccidioides* secreted proteins reveals fructose 1,6-bisphosphate aldolase as a plasminogen-binding protein. *BMC Microbiol* 15, 393.
- CHEMALE, G., VAN ROSSUM, A.J., JEFFERIES, J.R., BARRETT, J., BROPHY, P.M., FERREIRA, H.B., & ZAHA, A. (2003). Proteomic analysis of the larval stage of the parasite *Echinococcus granulosus*: causative agent of cystic hydatid disease. *Proteomics* 3, 1633-1636.
- CHOUDHARY, C., WEINERT, B.T., NISHIDA, Y., VERDIN, E., & MANN, M. (2014). The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol* 15, 536-550.
- COLLINGRIDGE, P.W., BROWN, R.W., & GINGER, M.L. (2010). Moonlighting enzymes in parasitic protozoa. *Parasitology* 137, 1467-1475.
- CONESA, A., GÖTZ, S., GARCÍA-GÓMEZ, J.M., TEROL, J., TALÓN, M., & ROBLES, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.
- CUI, S.J., XU, L.L., ZHANG, T., XU, M., YAO, J., FANG, C.Y., FENG, Z., YANG, P.Y., HU, W., & LIU, F. (2013). Proteomic characterization of larval and adult developmental stages in *Echinococcus granulosus* reveals novel insight into host-parasite interactions. *J Proteomics* 84, 158-175.
- DE LA TORRE-ESCUADERO, E., MANZANO-ROMÁN, R., PÉREZ-SÁNCHEZ, R., SILES-LUCAS, M., & OLEAGA, A. (2010). Cloning and characterization of a plasminogen-binding surface-associated enolase from *Schistosoma bovis*. *Vet Parasitol* 173, 76-84.
- DECKER, B.L., & WICKNER, W.T. (2006). Enolase activates homotypic vacuole fusion and protein transport to the vacuole in yeast. *J Biol Chem* 281, 14523-14528.
- DITGEN, D., ANANDARAJAH, E.M., MEISSNER, K.A., BRATTIG, N., WRENGER, C., & LIEBAU, E. (2014). Harnessing the helminth secretome for therapeutic immunomodulators. *Biomed Res Int* 2014, 964350.
- DÍAZ, A., CASARAVILLA, C., IRIGOÍN, F., LIN, G., PREVIATO, J.O., & FERREIRA, F. (2011). Understanding the laminated layer of larval *Echinococcus* I: structure. *Trends Parasitol* 27, 204-213.
- DÍAZ-MARTÍN, V., MANZANO-ROMÁN, R., OLEAGA, A., ENCINAS-GRANDES, A., & PÉREZ-SÁNCHEZ, R. (2013). Cloning and characterization of a plasminogen-binding enolase from the saliva of the argasid tick *Ornithodoros moubata*. *Vet Parasitol* 191, 301-314.
- DÍAZ-RAMOS, A., ROIG-BORRELLAS, A., GARCÍA-MELERO, A., & LÓPEZ-ALEMANY, R. (2012). α -Enolase, a multifunctional protein: its role on pathophysiological situations. *J Biomed Biotechnol* 2012, 156795.
- ECKERT, J., & DEPLAZES, P. (2004). Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern. *Clin Microbiol Rev* 17, 107-135.
- FROTTIN, F., MARTINEZ, A., PEYNOT, P., MITRA, S., HOLZ, R.C., GIGLIONE, C., &

- MEINNEL, T. (2006). The proteomics of N-terminal methionine cleavage. *Mol Cell Proteomics* 5, 2336-2349.
- GALINDO, M., GONZALEZ, M.J., & GALANTI, N. (2002). *Echinococcus granulosus* protoscolex formation in natural infections. *Biol Res* 35, 365-371.
- GALINDO, M., SCHADEBRODT, G., & GALANTI, N. (2008). *Echinococcus granulosus*: cellular territories and morphological regions in mature protoscoleces. *Exp Parasitol* 119, 524-533.
- GAN, W., ZHAO, G., XU, H., WU, W., DU, W., HUANG, J., YU, X., & HU, X. (2010). Reverse vaccinology approach identify an *Echinococcus granulosus* tegumental membrane protein enolase as vaccine candidate. *Parasitol Res* 106, 873-882.
- GANCEDO, C., & FLORES, C.L. (2008). Moonlighting proteins in yeasts. *Microbiol Mol Biol Rev* 72, 197-210, table of contents.
- GIGLIONE, C., BOULAROT, A., & MEINNEL, T. (2004). Protein N-terminal methionine excision. *Cell Mol Life Sci* 61, 1455-1474.
- GINGER, M.L. (2014). Protein moonlighting in parasitic protists. *Biochem Soc Trans* 42, 1734-1739.
- GÓMEZ-ARREAZA, A., ACOSTA, H., QUIÑONES, W., CONCEPCIÓN, J.L., MICHELS, P.A., & AVILÁN, L. (2014). Extracellular functions of glycolytic enzymes of parasites: unpredicted use of ancient proteins. *Mol Biochem Parasitol* 193, 75-81.
- HAILE, W.B., COLEMAN, J.L., & BENACH, J.L. (2006). Reciprocal upregulation of urokinase plasminogen activator and its inhibitor, PAI-2, by *Borrelia burgdorferi* affects bacterial penetration and host-inflammatory response. *Cell Microbiol* 8, 1349-1360.
- HARNETT, W. (2014). Secretory products of helminth parasites as immunomodulators. *Mol Biochem Parasitol* 195, 130-136.
- HE, H., LEE, M.C., ZHENG, L.L., ZHENG, L., & LUO, Y. (2013). Integration of the metabolic/redox state, histone gene switching, DNA replication and S-phase progression by moonlighting metabolic enzymes. *Biosci Rep* 33, e00018.
- HEISS, K., NIE, H., KUMAR, S., DALY, T.M., BERGMAN, L.W., & MATUSCHEWSKI, K. (2008). Functional characterization of a redundant *Plasmodium* TRAP family invasin, TRAP-like protein, by aldolase binding and a genetic complementation test. *Eukaryot Cell* 7, 1062-1070.
- HENDERSON, B. (2014). An overview of protein moonlighting in bacterial infection. *Biochem Soc Trans* 42, 1720-1727.
- HENDERSON, B., & MARTIN, A. (2011). Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect Immun* 79, 3476-3491.
- HENDERSON, B., & MARTIN, A.C. (2014). Protein moonlighting: a new factor in biology and medicine. *Biochem Soc Trans* 42, 1671-1678.
- HOLCMAN, B., & HEATH, D.D. (1997). The early stages of *Echinococcus granulosus* development. *Acta Trop* 64, 5-17.

- HUBERTS, D.H., & VAN DER KLEI, I.J. (2010). Moonlighting proteins: an intriguing mode of multitasking. *Biochim Biophys Acta* 1803, 520-525.
- HUTCHISON, C.A., PETERSON, S.N., GILL, S.R., CLINE, R.T., WHITE, O., FRASER, C.M., SMITH, H.O., & VENTER, J.C. (1999). Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* 286, 2165-2169.
- JEFFERY, C.J. (1999). Moonlighting proteins. *Trends Biochem Sci* 24, 8-11.
- JEFFERY, C.J. (2004). Molecular mechanisms for multitasking: recent crystal structures of moonlighting proteins. *Curr Opin Struct Biol* 14, 663-668.
- JEFFERY, C.J. (2009). Moonlighting proteins--an update. *Mol Biosyst* 5, 345-350.
- JEWETT, T.J., & SIBLEY, L.D. (2003). Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. *Mol Cell* 11, 885-894.
- JIANG, L., ZHANG, Y.G., LIU, M.X., & FENG, Z. (2012). Analysis on the reactivity of five subunits of antigen B family in serodiagnosis of echinococcosis. *Exp Parasitol* 131, 85-91.
- JUYI, L., YAN, J., XIUFANG, W., ZHAOQING, Z., JUNLIANG, L., MINGXING, Z., & WEI, Z. (2013). Analysis of the chemical components of hydatid fluid from *Echinococcus granulosus*. *Rev Soc Bras Med Trop* 46, 605-610.
- KARKOWSKA-KULETA, J., & KOZIK, A. (2014). Moonlighting proteins as virulence factors of pathogenic fungi, parasitic protozoa and multicellular parasites. *Mol Oral Microbiol* 29, 270-283.
- KELLEHER, N.L., THOMAS, P.M., NTAI, I., COMPTON, P.D., & LEDUC, R.D. (2014). Deep and quantitative top-down proteomics in clinical and translational research. *Expert Rev Proteomics* 11, 649-651.
- KELLIE, J.F., TRAN, J.C., LEE, J.E., AHLF, D.R., THOMAS, H.M., NTAI, I., CATHERMAN, A.D., DURBIN, K.R., ZAMDBORG, L., VELLAICHAMY, A., *et al.* (2010). The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. *Mol Biosyst* 6, 1532-1539.
- KHAN, I., CHEN, Y., DONG, T., HONG, X., TAKEUCHI, R., MORI, H., & KIHARA, D. (2014). Genome-scale identification and characterization of moonlighting proteins. *Biol Direct* 9, 30.
- KIM, J.H., LEE, J., PARK, J., & GHO, Y.S. (2015). Gram-negative and Gram-positive bacterial extracellular vesicles. *Semin Cell Dev Biol*.
- KUNERT, A., LOSSE, J., GRUSZIN, C., HÜHN, M., KAENDLER, K., MIKKAT, S., VOLKE, D., HOFFMANN, R., JOKIRANTA, T.S., SEEBERGER, H., *et al.* (2007). Immune evasion of the human pathogen *Pseudomonas aeruginosa*: elongation factor Tuf is a factor H and plasminogen binding protein. *J Immunol* 179, 2979-2988.
- LARRIEU, E., HERRERO, E., MUJICA, G., LABANCHI, J.L., ARAYA, D., GRIZMADO, C., CALABRO, A., TALMON, G., RUESTA, G., PEREZ, A., *et al.* (2013). Pilot field trial of the EG95 vaccine against ovine cystic echinococcosis in Rio Negro, Argentina: early impact and preliminary data. *Acta Trop* 127, 143-151.
- LI, X., & CHANG, Y.H. (1995). Amino-terminal protein processing in *Saccharomyces cerevisiae* is

an essential function that requires two distinct methionine aminopeptidases. *Proc Natl Acad Sci U S A* 92, 12357-12361.

LI, Y., XU, H., CHEN, J., GAN, W., WU, W., & HU, X. (2012). Gene cloning, expression, and localization of antigen 5 in the life cycle of *Echinococcus granulosus*. *Parasitol Res* 110, 2315-2323.

LIU, F., CUI, S.J., HU, W., FENG, Z., WANG, Z.Q., & HAN, Z.G. (2009). Excretory/secretory proteome of the adult developmental stage of human blood fluke, *Schistosoma japonicum*. *Mol Cell Proteomics* 8, 1236-1251.

LIU, H., ZENG, H., YAO, Q., YUAN, J., ZHANG, Y., QIU, D., YANG, X., YANG, H., & LIU, Z. (2012). *Steinernema glaseri* surface enolase: molecular cloning, biological characterization, and role in host immune suppression. *Mol Biochem Parasitol* 185, 89-98.

LU, M., AMMAR, D., IVES, H., ALBRECHT, F., & GLUCK, S.L. (2007). Physical interaction between aldolase and vacuolar H⁺-ATPase is essential for the assembly and activity of the proton pump. *J Biol Chem* 282, 24495-24503.

LUO, S., POLTERMANN, S., KUNERT, A., RUPP, S., & ZIPFEL, P.F. (2009). Immune evasion of the human pathogenic yeast *Candida albicans*: Pra1 is a Factor H, FHL-1 and plasminogen binding surface protein. *Mol Immunol* 47, 541-550.

MACKINTOSH, C. (2004). Dynamic interactions between 14-3-3 proteins and phosphoproteins regulate diverse cellular processes. *Biochem J* 381, 329-342.

MAIZELS, R.M., HEWITSON, J.P., MURRAY, J., HARCUS, Y.M., DAYER, B., FILBEY, K.J., GRAINGER, J.R., MCSORLEY, H.J., REYNOLDS, L.A., & SMITH, K.A. (2012). Immune modulation and modulators in *Heligmosomoides polygyrus* infection. *Exp Parasitol* 132, 76-89.

MANDAL, S., & MANDAL, M.D. (2012). Human cystic echinococcosis: epidemiologic, zoonotic, clinical, diagnostic and therapeutic aspects. *Asian Pac J Trop Med* 5, 253-260.

MANI, M., CHEN, C., AMBLEE, V., LIU, H., MATHUR, T., ZWICKE, G., ZABAD, S., PATEL, B., THAKKAR, J., & JEFFERY, C.J. (2015). MoonProt: a database for proteins that are known to moonlight. *Nucleic Acids Res* 43, D277-282.

MANN, M., & KELLEHER, N.L. (2008). Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci U S A* 105, 18132-18138.

MARCILLA, A., PÉREZ-GARCÍA, A., ESPERT, A., BERNAL, D., MUÑOZ-ANTOLÍ, C., ESTEBAN, J.G., & TOLEDO, R. (2007). *Echinostoma caproni*: identification of enolase in excretory/secretory products, molecular cloning, and functional expression. *Exp Parasitol* 117, 57-64.

MARCILLA, A., TRELIS, M., CORTÉS, A., SOTILLO, J., CANTALAPIEDRA, F., MINGUEZ, M.T., VALERO, M.L., SÁNCHEZ DEL PINO, M.M., MUÑOZ-ANTOLÍ, C., TOLEDO, R., & BERNAL, D. (2012). Extracellular vesicles from parasitic helminths contain specific excretory/secretory proteins and are internalized in intestinal host cells. *PLoS One* 7, e45974.

MARCOS, C.M., DE OLIVEIRA, H.C., DA SILVA, J.E.F., ASSATO, P.A., FUSCO-ALMEIDA, A.M., & MENDES-GIANNINI, M.J. (2014). The multifaceted roles of metabolic enzymes in the *Paracoccidioides* species complex. *Front Microbiol* 5, 719.

- MCMANUS, D.P. (2009). Reflections on the biochemistry of *Echinococcus*: past, present and future. *Parasitology* 136, 1643-1652.
- MCMANUS, D.P. (2013). Current status of the genetics and molecular taxonomy of *Echinococcus* species. *Parasitology* 140, 1617-1623.
- MCMANUS, D.P., & THOMPSON, R.C. (2003). Molecular epidemiology of cystic echinococcosis. *Parasitology* 127 Suppl, S37-51.
- MEINNEL, T., & GIGLIONE, C. (2008). Tools for analyzing and predicting N-terminal protein modifications. *Proteomics* 8, 626-649.
- MEINNEL, T., SERERO, A., & GIGLIONE, C. (2006). Impact of the N-terminal amino acid on targeted protein degradation. *Biol Chem* 387, 839-851.
- MODUN, B., MORRISSEY, J., & WILLIAMS, P. (2000). The staphylococcal transferrin receptor: a glycolytic enzyme with novel functions. *Trends Microbiol* 8, 231-237.
- MONTANER, S., GALIANO, A., TRELIS, M., MARTIN-JAULAR, L., DEL PORTILLO, H.A., BERNAL, D., & MARCILLA, A. (2014). The Role of Extracellular Vesicles in Modulating the Host Immune Response during Parasitic Infections. *Front Immunol* 5, 433.
- MONTEIRO, K.M., CARDOSO, M.B., FOLLMER, C., DA SILVEIRA, N.P., VARGAS, D.M., KITAJIMA, E.W., ZAHA, A., & FERREIRA, H.B. (2012). *Echinococcus granulosus* antigen B structure: subunit composition and oligomeric states. *PLoS Negl Trop Dis* 6, e1551.
- MONTEIRO, K.M., DE CARVALHO, M.O., ZAHA, A., & FERREIRA, H.B. (2010). Proteomic analysis of the *Echinococcus granulosus* metacestode during infection of its intermediate host. *Proteomics* 10, 1985-1999.
- MORO, P., & SCHANTZ, P.M. (2009). Echinococcosis: a review. *Int J Infect Dis* 13, 125-133.
- MORO, P.L., BUDKE, C.M., SCHANTZ, P.M., VASQUEZ, J., SANTIVANEZ, S.J., & VILLAVICENCIO, J. (2011). Economic impact of cystic echinococcosis in peru. *PLoS Negl Trop Dis* 5, e1179.
- MOUVEAUX, T., ORIA, G., WERKMEISTER, E., SLOMIANNY, C., FOX, B.A., BZIK, D.J., & TOMAVO, S. (2014). Nuclear glycolytic enzyme enolase of *Toxoplasma gondii* functions as a transcriptional regulator. *PLoS One* 9, e105820.
- NIINAKA, Y., PAKU, S., HAGA, A., WATANABE, H., & RAZ, A. (1998). Expression and secretion of neuroleukin/phosphohexose isomerase/maturation factor as autocrine motility factor by tumor cells. *Cancer Res* 58, 2667-2674.
- NOGUEIRA, S.V., FONSECA, F.L., RODRIGUES, M.L., MUNDODI, V., ABI-CHACRA, E.A., WINTERS, M.S., ALDERETE, J.F., & DE ALMEIDA SOARES, C.M. (2010). *Paracoccidioides brasiliensis* enolase is a surface protein that binds plasminogen and mediates interaction of yeast forms with host cells. *Infect Immun* 78, 4040-4050.
- NONO, J.K., PLETINCKX, K., LUTZ, M.B., & BREHM, K. (2012). Excretory/secretory-products of *Echinococcus multilocularis* larvae induce apoptosis and tolerogenic properties in dendritic cells in vitro. *PLoS Negl Trop Dis* 6, e1516.

OEHRING, S.C., WOODCROFT, B.J., MOES, S., WETZEL, J., DIETZ, O., PULFER, A., DEKIWADIA, C., MAESER, P., FLUECK, C., WITMER, K., *et al.* (2012). Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite *Plasmodium falciparum*. *Genome Biol* 13, R108.

OLSEN, J.V., & MANN, M. (2013). Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics* 12, 3444-3452.

PAGNOZZI, D., BIOSA, G., ADDIS, M.F., MASTRANDREA, S., MASALA, G., & UZZAU, S. (2014). An easy and efficient method for native and immunoreactive *Echinococcus granulosus* antigen 5 enrichment from hydatid cyst fluid. *PLoS One* 9, e104962.

PAL-BHOWMICK, I., VORA, H.K., & JARORI, G.K. (2007). Sub-cellular localization and post-translational modifications of the *Plasmodium yoelii* enolase suggest moonlighting functions. *Malar J* 6, 45.

PALUDO, G.P., LORENZATTO, K.R., BONATTO, D., & FERREIRA, H.B. (2015). Systems biology approach reveals possible evolutionarily conserved moonlighting functions for enolase. *Computational Biology and Chemistry. In Press.*

PAN, W., SHEN, Y., HAN, X., WANG, Y., LIU, H., JIANG, Y., ZHANG, Y., XU, Y., & CAO, J. (2014). Transcriptome profiles of the protoscoleces of *Echinococcus granulosus* reveal that excretory-secretory products are essential to metabolic adaptation. *PLoS Negl Trop Dis* 8, e3392.

PANCHOLI, V. (2001). Multifunctional alpha-enolase: its role in diseases. *Cell Mol Life Sci* 58, 902-920.

PARIONA-LLANOS, R., PAVANI, R.S., REIS, M., NOËL, V., SILBER, A.M., ARMELIN, H.A., CANO, M.I., & ELIAS, M.C. (2015). Glyceraldehyde 3-Phosphate Dehydrogenase-Telomere Association Correlates with Redox Status in *Trypanosoma cruzi*. *PLoS One* 10, e0120896.

PARKINSON, J., WASMUTH, J.D., SALINAS, G., BIZARRO, C.V., SANFORD, C., BERRIMAN, M., FERREIRA, H.B., ZAHA, A., BLAXTER, M.L., MAIZELS, R.M., & FERNÁNDEZ, C. (2012). A transcriptomic analysis of *Echinococcus granulosus* larval stages: implications for parasite biology and host adaptation. *PLoS Negl Trop Dis* 6, e1897.

PATRUSHEV, L.I., & MINKEVICH, I.G. (2008). The problem of the eukaryotic genome size. *Biochemistry (Mosc)* 73, 1519-1552.

POLEVODA, B., ARNESEN, T., & SHERMAN, F. (2009). A synopsis of eukaryotic Nalpha-terminal acetyltransferases: nomenclature, subunits and substrates. *BMC Proc* 3 *Suppl* 6, S2.

POLEVODA, B., BROWN, S., CARDILLO, T.S., RIGBY, S., & SHERMAN, F. (2008). Yeast N(alpha)-terminal acetyltransferases are associated with ribosomes. *J Cell Biochem* 103, 492-508.

POMEL, S., LUK, F.C., & BECKERS, C.J. (2008). Host cell egress and invasion induce marked relocations of glycolytic enzymes in *Toxoplasma gondii* tachyzoites. *PLoS Pathog* 4, e1000188.

PRABAKARAN, S., LIPPENS, G., STEEN, H., & GUNAWARDENA, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev Syst Biol Med* 4, 565-583.

PRUITT, K.D., TATUSOVA, T., & MAGLOTT, D.R. (2007). NCBI reference sequences (RefSeq):

a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35, D61-65.

RAMAJO-HERNÁNDEZ, A., PÉREZ-SÁNCHEZ, R., RAMAJO-MARTÍN, V., & OLEAGA, A. (2007). *Schistosoma bovis*: plasminogen binding in adults and the identification of plasminogen-binding proteins from the worm tegument. *Exp Parasitol* 115, 83-91.

RANGARAJAN, E.S., PARK, H., FORTIN, E., SYGUSCH, J., & IZARD, T. (2010). Mechanism of aldolase control of sorting nexin 9 function in endocytosis. *J Biol Chem* 285, 11983-11990.

ROGAN, M.T., HAI, W.Y., RICHARDSON, R., ZEYHLE, E., & CRAIG, P.S. (2006). Hydatid cysts: does every picture tell a story? *Trends Parasitol* 22, 431-438.

ROMIG, T. (2003). Epidemiology of echinococcosis. *Langenbecks Arch Surg* 388, 209-217.

RUAN, J., MOUVEAUX, T., LIGHT, S.H., MINASOV, G., ANDERSON, W.F., TOMAVO, S., & NGÔ, H.M. (2015). The structure of bradyzoite-specific enolase from *Toxoplasma gondii* reveals insights into its dual cytoplasmic and nuclear functions. *Acta Crystallogr D Biol Crystallogr* 71, 417-426.

SCHLÜTER, H., APWEILER, R., HOLZHÜTTER, H.G., & JUNGBLUT, P.R. (2009). Finding one's way in proteomics: a protein species nomenclature. *Chem Cent J* 3, 11.

SHEINER, L., SANTOS, J.M., KLAGES, N., PARUSSINI, F., JEMMELY, N., FRIEDRICH, N., WARD, G.E., & SOLDATI-FAVRE, D. (2010). *Toxoplasma gondii* transmembrane microneme proteins and their modular design. *Mol Microbiol*.

SIRACUSANO, A., DELUNARDO, F., TEGGI, A., & ORTONA, E. (2012a). Cystic echinococcosis: aspects of immune response, immunopathogenesis and immune evasion from the human host. *Endocr Metab Immune Disord Drug Targets* 12, 16-23.

SIRACUSANO, A., DELUNARDO, F., TEGGI, A., & ORTONA, E. (2012b). Host-parasite relationship in cystic echinococcosis: an evolving story. *Clin Dev Immunol* 2012, 639362.

SIRACUSANO, A., RIGANÒ, R., ORTONA, E., PROFUMO, E., MARGUTTI, P., BUTTARI, B., DELUNARDO, F., & TEGGI, A. (2008). Immunomodulatory mechanisms during *Echinococcus granulosus* infection. *Exp Parasitol* 119, 483-489.

SLUCHANKO, N.N., ROMAN, S.G., CHEBOTAREVA, N.A., & GUSEV, N.B. (2014). Chaperone-like activity of monomeric human 14-3-3 ζ on different protein substrates. *Arch Biochem Biophys* 549, 32-39.

SLUCHANKO, N.N., & UVERSKY, V.N. (2015). Hidden disorder propensity of the N-terminal segment of universal adapter protein 14-3-3 is manifested in its monomeric form: Novel insights into protein dimerization and multifunctionality. *Biochim Biophys Acta*.

SMITH, L.M., KELLEHER, N.L., & PROTEOMICS, C.F.T.D. (2013). Proteoform: a single term describing protein complexity. *Nat Methods* 10, 186-187.

SRIRAM, G., MARTINEZ, J.A., MCCABE, E.R., LIAO, J.C., & DIPPLE, K.M. (2005). Single-gene disorders: what role could moonlighting enzymes play? *Am J Hum Genet* 76, 911-924.

STARNES, G.L., COINCON, M., SYGUSCH, J., & SIBLEY, L.D. (2009). Aldolase is essential for

energy production and bridging adhesin-actin cytoskeletal interactions during parasite invasion of host cells. *Cell Host Microbe* 5, 353-364.

STEINKE, D., HOEGG, S., BRINKMANN, H., & MEYER, A. (2006). Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates. *BMC Biol* 4, 16.

TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M., & KUMAR, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731-2739.

TEICHMANN, A., VARGAS, D.M., MONTEIRO, K.M., MENEGHETTI, B.V., DUTRA, C.S., PAREDES, R., GALANTI, N., ZAHA, A., & FERREIRA, H.B. (2015). Characterization of 14-3-3 isoforms expressed in the *Echinococcus granulosus* pathogenic larval stage. *J Proteome Res*.

THOMPSON, R.C., & JENKINS, D.J. (2014). *Echinococcus* as a model system: biology and epidemiology. *Int J Parasitol*.

THOMPSON, R.C., & LYMBERY, A.J. (2013). Let's not forget the thinkers. *Trends Parasitol* 29, 581-584.

THOMPSON, R.C., & MCMANUS, D.P. (2002). Towards a taxonomic revision of the genus *Echinococcus*. *Trends Parasitol* 18, 452-457.

TORRES-MACHORRO, A.L., ARIS, J.P., & PILLUS, L. (2015). A moonlighting metabolic protein influences repair at DNA double-stranded breaks. *Nucleic Acids Res* 43, 1646-1658.

TOVY, A., SIMAN TOV, R., GAENTZSCH, R., HELM, M., & ANKRI, S. (2010). A new nuclear function of the *Entamoeba histolytica* glycolytic enzyme enolase: the metabolic regulation of cytosine-5 methyltransferase 2 (Dnmt2) activity. *PLoS Pathog* 6, e1000775.

TRAN, J.C., ZAMDBORG, L., AHLF, D.R., LEE, J.E., CATHERMAN, A.D., DURBIN, K.R., TIPTON, J.D., VELLAICHAMY, A., KELLIE, J.F., LI, M., *et al.* (2011). Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 480, 254-258.

TRISTAN, C., SHAHANI, N., SEDLAK, T.W., & SAWA, A. (2011). The diverse functions of GAPDH: views from different subcellular compartments. *Cell Signal* 23, 317-323.

TSAI, I.J., ZAROWIECKI, M., HOLROYD, N., GARCIARRUBIO, A., SANCHEZ-FLORES, A., BROOKS, K.L., TRACEY, A., BOBES, R.J., FRAGOSO, G., SCIUTTO, E., *et al.* (2013). The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496, 57-63.

TUNIO, S.A., OLDFIELD, N.J., BERRY, A., ALA'ALDEEN, D.A., WOOLDRIDGE, K.G., & TURNER, D.P. (2010). The moonlighting protein fructose-1, 6-bisphosphate aldolase of *Neisseria meningitidis*: surface localization and role in host cell adhesion. *Mol Microbiol* 76, 605-615.

TWU, O., & JOHNSON, P.J. (2014). Parasite extracellular vesicles: mediators of intercellular communication. *PLoS Pathog* 10, e1004289.

VIRGINIO, V.G., MONTEIRO, K.M., DRUMOND, F., DE CARVALHO, M.O., VARGAS, D.M., ZAHA, A., & FERREIRA, H.B. (2012). Excretory/secretory products from in vitro-cultured *Echinococcus granulosus* protoscoleces. *Mol Biochem Parasitol* 183, 15-22.

WANG, J., MORRIS, A.J., TOLAN, D.R., & PAGLIARO, L. (1996). The molecular nature of the

- F-actin binding activity of aldolase revealed with site-directed mutants. *J Biol Chem* *271*, 6861-6865.
- WENGER, C.D., PHANSTIEL, D.H., LEE, M.V., BAILEY, D.J., & COON, J.J. (2011). COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* *11*, 1064-1074.
- WESSEL, D., & FLÜGGE, U.I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* *138*, 141-143.
- WISTOW, G.J., LIETMAN, T., WILLIAMS, L.A., STAPEL, S.O., DE JONG, W.W., HORWITZ, J., & PIATIGORSKY, J. (1988). Tau-crystallin/alpha-enolase: one gene encodes both an enzyme and a lens structural protein. *J Cell Biol* *107*, 2729-2736.
- ZHANG, J., YU, C., WANG, Y., FANG, W., & LUO, D. (2014). Enolase of *Angiostrongylus cantonensis*: more likely a structural component? *Parasitol Res* *113*, 3927-3934.
- ZHANG, W., WEN, H., LI, J., LIN, R., & MCMANUS, D.P. (2012). Immunology and immunodiagnosis of cystic echinococcosis: an update. *Clin Dev Immunol* *2012*, 101895.
- ZHANG, W.B., JONES, M.K., LI, J., & MCMANUS, D.P. (2005). *Echinococcus granulosus*: pre-culture of protoscoleces in vitro significantly increases development and viability of secondary hydatid cysts in mice. *Exp Parasitol* *110*, 88-90.
- ZHENG, H., ZHANG, W., ZHANG, L., ZHANG, Z., LI, J., LU, G., ZHU, Y., WANG, Y., HUANG, Y., LIU, J., *et al.* (2013). The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nat Genet* *45*, 1168-1175.
- ZHENG, L., ROEDER, R.G., & LUO, Y. (2003). S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell* *114*, 255-266.

8. CURRICULUM VITAE

LORENZATTO, K. R.

(karilorenzatto@gmail.com)

Formação acadêmica

- 2009 - 2011** Mestrado em Biologia Celular e Molecular
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Título: Estudo de três enzimas da via glicolítica de *Echinococcus granulosus* com possíveis funções *moonlighting* na interação da forma larval com o hospedeiro intermediário, Ano de obtenção: 2011
Orientador: Henrique Bunselmeyer Ferreira
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- 2004 - 2008** Graduação em Ciências Biológicas - Bacharelado.
Universidade Federal de Santa Maria, UFSM, Brasil

Formação complementar

- 2012** Análise de dados proteômicos.
Instituto Carlos Chagas, (FIOCRUZ Paraná), Curitiba, PR, Brasil
- 2010** Metodologias de histologia e imunofluorescência.
Universidad Andrés Bello, UAB, Chile
- 2008** Biologia Molecular de Parasitas.
Universidad de la Republica Uruguay, UDELAR, Montevideo, Uruguay
- 2007** Otimização e Automação de PCR em tempo real.
Universidade Federal de Santa Maria, UFSM, Santa Maria, RS, Brasil

Estágios realizados

- Fev-Jul/ 2014** Programa de Doutorado Sanduíche no Exterior (PDSE)/ CAPES
Análise proteômica *top down* de frações subcelulares de protoescolíces de *Echinococcus granulosus*
Proteomics Center of Excellence, Northwestern University, Evanston, Estados Unidos
- Out/ 2013** Missão de curta duração no exterior (PROPG/UFRGS)
Introdução à proteômica *top down*
Proteomics Center of Excellence, Northwestern University, Evanston, Estados Unidos

Produção bibliográfica

Artigos completos publicados em periódicos

1. MIYAMOTO, KENDI NISHINO, MONTEIRO, KARINA MARIANTE, DA SILVA CAUMO, KARIN, **LORENZATTO, KARINA RODRIGUES**, FERREIRA, HENRIQUE BUNSELMEYER, BRANDELLI, ADRIANO

Comparative proteomic analysis of *Listeria monocytogenes* ATCC 7644 exposed to a sublethal concentration of nisin. *Journal of Proteomics (Print)*. , v.119, p.230 - 237, 2015.

2. COSTA, CAROLINE B., MONTEIRO, KARINA M., TEICHMANN, ALINE, DA SILVA, EDILEUZA D., **LORENZATTO, KARINA R.**, CANCELA, MARTÍN, PAES, JÉSSICA A., BENITZ, ANDRÉ DE N. D., CASTILLO, ESTELA, MARGIS, ROGÉRIO, Zaha, Arnaldo, FERREIRA, HENRIQUE B.

Expression of the histone chaperone SET/TAF-Iβ during the strobilation process of *Mesocostoides corti* (Platyhelminthes, Cestoda). *Parasitology (London. Print)*. , v.xx, p.1 - 12, 2015.

3. **LORENZATTO, KARINA RODRIGUES**, MONTEIRO, KARINA MARIANTE, PAREDES, RODOLFO, PALUDO, GABRIELA PRADO, DA FONSÊCA, MARBELLA MARIA, GALANTI, NORBEL, ZAHA, ARNALDO, FERREIRA, HENRIQUE BUNSELMEYER

Fructose-bisphosphate aldolase and enolase from *Echinococcus granulosus*: Genes, expression patterns and protein interactions of two potential moonlighting proteins. *Gene (Amsterdam)*. , v.506, p.76 - 84, 2012.

4. FREITAS ORTIZ, MAURO, **LORENZATTO, KARINA RODRIGUES**, CORRÊA, BRUNA RENATA SILVA, LORETO, ELGION LUCIO SILVA

hAT transposable elements and their derivatives: an analysis in the 12 *Drosophila* genomes. *Genetica ('s-Gravenhage)*. , p.x - x, 2010.

5. KOCHHANN, DAIANI, BENADUCE, ANA PAULA S., COPATTI, CARLOS E., **LORENZATTO, K. R.**, MESKO, MÁRCIA F., FLORES, ÉRICO M. M., DRESSLER, VALDERI L., BALDISSEROTTO, BERNARDO

Protective Effect of High Alkalinity Against the Deleterious Effects of Chronic Waterborne Cadmium Exposure on the Detection of Alarm Cues by Juvenile Silver Catfish (*Rhamdia quelen*). *Archives of Environmental Contamination and Toxicology*. , v.56, p.770 - 775, 2009.

Artigo aceito para publicação

PALUDO, G. P., **LORENZATTO, K. R.**, BONATTO, D., FERREIRA, H. B.

Systems biology approach reveals possible evolutionarily conserved moonlighting functions for enolase. *Computational Biology and Chemistry*. 2015. *In Press*

Trabalhos publicados em anais de eventos (resumo)

1. **LORENZATTO, K. R.**, KIM, K., NTAI, I., THOMAS, P. M., KELLEHER, N. L., FERREIRA, H. B.

Top down mass spectrometry of up to 30 kDa proteins from *E. granulosus*, 2nd Proteomics Meeting of the Brazilian Proteomics Society jointly with the 2nd Pan American HUPO Meeting, 2014.

2. **LORENZATTO, K. R.**, PALUDO, GABRIELA PRADO, MONTEIRO, KARINA MARIANTE, ZAHA, ARNALDO, FERREIRA, H. B. **Frutose-bifosfato-aldolase e enolase de *Echinococcus granulosus*: duas potenciais proteínas moonlighting**. XXIII Congresso Brasileiro de Parasitologia – III Encontro de Parasitologia do Mercosul, 2013.

3. **LORENZATTO, K. R.**, PAREDES, R., PALUDO, G. P., MONTEIRO, K. M., ZAHA, A., FERREIRA, H. B.

Estudo de duas enzimas da via glicolítica de *Echinococcus granulosus* com possíveis funções moonlighting na interação da forma larval com o hospedeiro intermediário. XXII Congresso Brasileiro de Parasitologia, 2011.

4. **LORENZATTO, K. R., ZAHA, A., FERREIRA, H. B.**
Investigation of enolase roles in *Echinococcus granulosus*-host interactions. XXXIX Annual Meeting of the Brazilian Biochemistry and Molecular Biology Society, 2010, Foz do Iguaçu.
5. **LORENZATTO, K. R., ZAHA, A., FERREIRA, H. B.**
Functional studies of the *Echinococcus granulosus* fructose-biphosphate aldolase. XXIII International Congress of Hydatidology, 2009, Colonia del Sacramento.
6. **ARBILDI, P., LORENZATTO, K. R., MARQUES, J. M., LA ROCCA, S., FERNANDEZ, V.**
***In vitro* analysis of glutathione transferases expression in *Echinococcus granulosus* protoescolex.** XXIII International Congress of Hydatidology, 2009, Colonia del Sacramento.
7. **LORENZATTO, K. R., Ortiz, M. F., Sepel, L. M. N., Loreto, E. L. S.**
Identificação de elementos relacionados à hATs nos genomas de *Drosophila*. Congresso Brasileiro de Genética, 2007, Águas de Lindóia.
8. **LORENZATTO, K. R., Loreto, E. L. S., Sepel, L. M. N.**
Uso de Meio Essencial Mínimo-Eagle (MEM) para cultivo de células de embriões de *Zaprionus indianus*.
52° Congresso Brasileiro de Genética, 2006, Foz do Iguaçu.

Apresentação de trabalho

1. **LORENZATTO, K. R.,** Monteiro, Karina Mariante, Paredes, Rodolfo, Paludo, Gabriela Prado, da Fonsêca, Marbella Maria, Galanti, Norbel, Zaha, Arnaldo, FERREIRA, H. B.
***Echinococcus granulosus* aldolase and enolase: two potential moonlighting functions.** XVIII International Congress for Tropical Medicine and Malaria and XLVIII Congress of the Brazilian Society of Tropical Medicine, 2012.