

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Rogério Valim da Trindade

**Identificação de alvos moleculares de compostos por meio de proteólise pulsada
e espectrometria de massa**

Porto Alegre

2016

ROGÉRIO VALIM DA TRINDADE

**IDENTIFICAÇÃO DE ALVOS MOLECULARES DE COMPOSTOS POR MEIO DE
PROTEÓLISE PULSADA E ESPECTROMETRIA DE MASSA**

Dissertação apresentada ao Programa de Pós-graduação em Biologia Celular e Molecular da PUCRS como requisito para a obtenção do grau de Mestre em Biologia Celular e Molecular.

Orientador: Prof. Dr. Cristiano Valim Bizarro

Coorientador: Dr. Antônio Frederico Michel Pinto

Porto Alegre

2016

ROGÉRIO VALIM DA TRINDADE

**IDENTIFICAÇÃO DE ALVOS MOLECULARES DE COMPOSTOS POR MEIO DE
PROTEÓLISE PULSADA E ESPECTROMETRIA DE MASSA**

Dissertação apresentada ao Programa de Pós-graduação em Biologia Celular e Molecular da PUCRS como requisito para a obtenção do grau de Mestre em Biologia Celular e Molecular.

BANCA EXAMINADORA:

Carlos Alexandre Sanchez Ferreira – PUCRS

Eduardo Jorge Pilau – UEM

Mário Henrique Bengtson – UNICAMP

AGRADECIMENTOS

Ao meu orientador, Professor Cristiano Valim Bizarro, pela orientação, confiança, paciência, ensinamentos e auxílio, essenciais em todas as etapas de desenvolvimento desse trabalho.

Ao meu co-orientador, Doutor Antônio Frederico Michel Pinto, pela confiança, prestatividade, paciência e auxílio.

Agradeço ao Professor Diógenes Santiago Santos, pela confiança e pela oportunidade de trabalhar no Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF).

Agradeço ao Professor Luiz Augusto Basso e ao Professor Pablo Machado por todo o conhecimento compartilhado durante o período de desenvolvimento desse trabalho.

A minha família, por todas as oportunidades que me proporcionaram e pelo apoio dado durante esses dois anos.

Agradeço aos amigos e colegas do laboratório que eu tive a alegria de conviver durante esses dois anos.

A FAPERGS e demais órgãos de fomento pelo financiamento que viabilizou a realização deste trabalho.

RESUMO

A tuberculose é, junto com a Síndrome da Imunodeficiência Adquirida (SIDA), a doença infecciosa que mais causa mortes no mundo. Grandes esforços foram exercidos nos últimos anos para a busca de novos fármacos antituberculose, porém sem muito sucesso. As plataformas de desenvolvimento de fármacos falham em diferentes etapas do processo de pesquisa e desenvolvimento e um dos aspectos que contribui para isso é a falta de conhecimento dos alvos moleculares de potenciais fármacos. Tanto a elucidação do mecanismo de ação de compostos bioativos quanto a identificação de seus alvos celulares são importantes para o desenvolvimento de novos fármacos. Como alternativa experimental para a identificação de alvos moleculares está a técnica de proteólise pulsada, a qual determina a fração de proteínas diferencialmente degradadas após breve incubação com proteases quando comparadas as amostras tratadas e não tratadas com um ligante. Nesse trabalho, desenvolvemos o método PePTID: um novo método para identificação de alvos moleculares usando a proteólise pulsada e análise por espectrometria de massa. A aplicação do método PePTID resultou em um procedimento reprodutível que minimiza a quantidade necessária de amostra e que apresenta uma metodologia de análise mais robusta em relação a outras abordagens que utilizam a proteólise pulsada.

Palavras-chave: PePTID, proteólise pulsada, identificação de alvos moleculares, desenvolvimento de fármacos anti-TB.

ABSTRACT

Tuberculosis is the infectious disease that causes more deaths worldwide, along with acquired immunodeficiency syndrome (AIDS). Great efforts have been exerted in recent years to the search for new anti-tuberculosis drugs without success. Drug development platforms fail at different stages of the research and development process and one aspect that contributes to this is the lack of knowledge of the potential molecular drug targets. Both the elucidation of the bioactive compounds mechanism of action and the identification of their cellular targets are important for the development of new drugs. As an experimental alternative to the identification of molecular targets is pulse proteolysis technique, which determines the fraction of proteins differentially degraded after a brief incubation with proteases when comparing samples treated and untreated with a ligand. In this work, we developed the PePTID method: a new method for identification of molecular targets using pulse proteolysis and mass spectrometry analysis. The application of PePTID resulted in a reproducible procedure that minimizes the required amount of sample and provides a more robust analysis method compared to other approaches that use pulse proteolysis.

Keywords: PePTID, pulse proteolysis, target discovery, anti-TB drug discovery.

LISTA DE ILUSTRAÇÕES

Figura 1. Representação do envelope celular micobacteriano	9
Figura 2. Estimativa dos novos casos de tuberculose (todas as formas) a cada 100.000 habitantes no ano de 2014.....	10
Figura 3. Metodologia para identificação de novos agentes antimicobacterianos desenvolvida pela GSK	14
Figura 4. Famílias de compostos mais promissores identificados pela GSK.....	16
Figura 5. Procedimento para identificação de alvos moleculares por proteólise pulsada	19
Figura 6. Comparação esquemática dos procedimentos experimentais	24
Figura 7. SDS-PAGE das amostras submetidas a proteólise pulsada	30
Figura 8. Cromatograma, sequência da proteína InhA e contagens espectrais do procedimento realizado utilizando a proteína InhA purificada	31
Figura 9. Cromatograma, sequência da proteína InhA e contagens espectrais do procedimento realizado utilizando o extrato proteico de <i>M. smegmatis</i> em conjunto com a proteína InhA purificada	32
Figura 10. Organização dos complexos multienzimáticos do sistema FAS-II	33
Figura 11. Gráficos das abundâncias relativas para as proteínas KasB e InhA em função da duração do pulso proteolítico.....	34

SUMÁRIO

1	INTRODUÇÃO.....	8
1.1	A tuberculose.....	8
1.2	O desenvolvimento de fármacos antituberculose	12
1.3	GSK e a triagem fenotípica de compostos	14
1.4	Identificação de alvos moleculares	17
1.5	Identificação de alvos moleculares por proteólise pulsada.....	18
2	JUSTIFICATIVA.....	21
3	OBJETIVOS.....	22
3.1	Objetivo geral.....	22
3.2	Objetivos específicos.....	22
4	CAPÍTULO I - DESENVOLVIMENTO DO MÉTODO.....	23
4.1	Métodos.....	25
4.1.1	Preparação do extrato proteico de <i>Mycobacterium smegmatis</i>	25
4.1.2	Amplificação, expressão e purificação da proteína 2-trans-enoil-ACP (CoA) redutase (InhA) de <i>M. tuberculosis</i>	26
4.1.3	Incubações	26
4.1.4	Proteólise pulsada e precipitação.....	27
4.1.5	Preparo da amostra para a análise por Espectrometria de Massa (MS)	27
4.1.6	Análise de dados	29
4.2	Resultados.....	30
5	CAPÍTULO II - APLICAÇÃO DO MÉTODO PePTID.....	35
5.1	Artigo científico	36
6	CONSIDERAÇÕES FINAIS.....	65
	REFERÊNCIAS.....	66

1 INTRODUÇÃO

1.1 A tuberculose

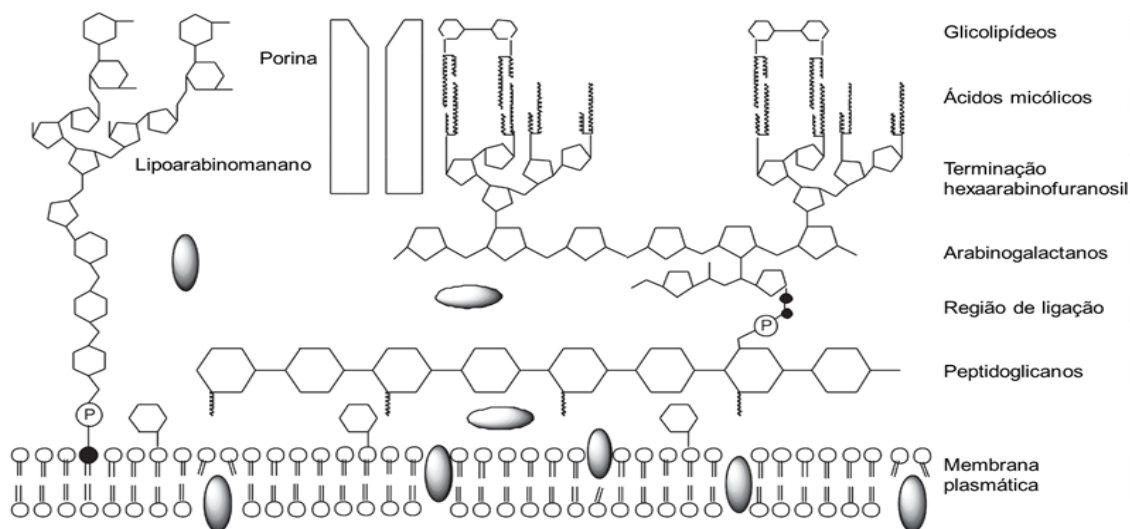
A tuberculose (TB) humana acomete milhões de pessoas em todo o mundo. É considerada a maior causa de mortes por uma doença infecciosa causada por um único patógeno, ao lado da Síndrome da Imunodeficiência Adquirida (SIDA), causada pelo Vírus da Imunodeficiência Humana (HIV) (WHO, 2015). O principal agente causador da TB humana é o bacilo *Mycobacterium tuberculosis* (*Mtb*), porém *M. bovis* e *M. africanum* também são capazes de causar a doença (Ducati *et al.*, 2006).

O *Mtb* é um patógeno intracelular facultativo, aeróbico, que infecta tipicamente os pulmões, sendo transmitido por intermédio de aerossóis expelidos por indivíduos portadores da forma ativa da doença (Russell *et al.*, 2010; Lawn e Zumla, 2011; Koul *et al.*, 2011). Ao entrar nos pulmões, os bacilos são prontamente fagocitados por macrófagos alveolares, desencadeando uma resposta inflamatória local com a mobilização de células mononucleadas que, na maior parte dos casos, irão conter mas não eliminar a infecção, resultando em um agregado de células compacto denominado granuloma (Lawn e Zumla, 2011). A maioria dos indivíduos infectados permanece com a doença em estado latente, na qual o bacilo permanece no hospedeiro em um estado não-replicativo; apesar de não transmitirem o bacilo durante esta fase, estes indivíduos representam um importante reservatório da doença, uma vez que um terço da população mundial encontra-se nesta situação (Ducati *et al.*, 2006; Koul *et al.*, 2011). Estima-se que de 5 a 15% dos indivíduos infectados irão desenvolver a forma ativa da doença, ou seja, aquela em que o bacilo consegue se replicar, causar lesões no hospedeiro e ser transmitido pelo ar (WHO, 2015).

O *Mtb* possui um envelope celular altamente hidrofóbico que atua como uma barreira de contenção, dificultando a efetividade dos fármacos utilizados no tratamento (Ducati *et al.*, 2006). A composição única desse envelope celular aumenta o tempo de tratamento e dificulta a determinação da susceptibilidade da cepa portada aos medicamentos antituberculose disponíveis. Dentre as estruturas que compõem o

envelope celular, destacam-se três componentes bastante abundantes, os quais estão covalentemente ligados: peptidoglicanos, arabinogalactanos e ácidos micólicos (Figura 1) (Takayama, Wang e Besra, 2005; Ducati *et al.*, 2006; Marrakchi *et al.*, 2014; Bansal-Mutalik e Nikaido, 2014).

Figura 1. Representação do envelope celular micobacteriano

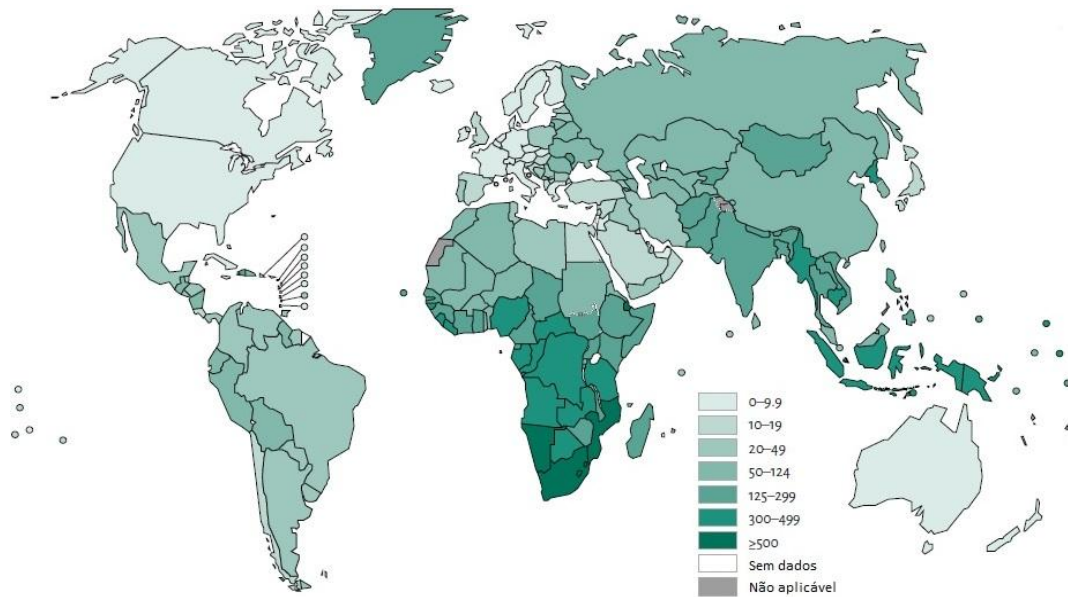


Fonte: Adaptado de Ducati *et al.* (2006).

A Organização Mundial da Saúde (OMS) estimou o surgimento de aproximadamente 9,6 milhões de novos casos de TB no ano de 2014, levando 1,5 milhão de pessoas ao óbito. Estima-se também que 1,2 milhão (12%) de casos e 390.000 mortes estão associadas à coinfeção do *Mtb* e o HIV (Corbett *et al.*, 2003; WHO, 2015).

Dos casos reportados da doença, 58% ocorreram na Ásia e 28% na África; em menores proporções aparecem as regiões do Mediterrâneo oriental (8%), da Europa (3%) e das Américas (3%). Nas Américas, o Brasil tem a maior prevalência, com uma taxa de incidência na ordem de 50 casos a cada 100.000 habitantes (Figura 2), o que o torna parte do grupo dos 22 países que somam 80% dos casos de TB no mundo, denominados países de alta carga (WHO, 2015).

Figura 2. Estimativa dos novos casos de tuberculose (todas as formas) a cada 100.000 habitantes no ano de 2014



Fonte: Adaptado de WHO (2015).

A classificação da TB como um problema emergencial de saúde pública global fez com que esforços fossem voltados para o combate da doença e grandes progressos foram alcançados. A criação do programa de tratamento diretamente supervisionado de curta duração (*DOTS* - do inglês, *directly observed treatment short course*), em 1993, juntamente com o programa que o sucede, *Stop TB Strategy*, proporcionaram o sucesso no tratamento de mais de 56 milhões de pessoas. Estima-se que aproximadamente 43 milhões de vidas foram salvas entre os anos de 2000 e 2014 em virtude das novas estratégias implementadas. No entanto, o número de casos ainda é alarmante, principalmente em países em desenvolvimento onde ainda se observa altas taxas de coinfeção *Mtb*-HIV (WHO, 2015).

O tratamento atualmente recomendado para os portadores de cepas não resistentes tem duração de seis meses e consiste na administração de fármacos de primeira-linha. Para esses pacientes, faz-se uso de isoniazida, rifampicina, etambutol e pirazinamida durante dois meses, seguido por mais quatro meses de isoniazida e rifampicina. Aproximadamente 90% dos casos de portadores de cepas sensíveis a

estes fármacos são curados com a conclusão do tratamento (Koul *et al.*, 2011; WHO, 2015). Contudo, devido aos efeitos adversos, ao longo período e à complexidade do regime, muitos pacientes abandonam ou não aderem corretamente ao tratamento, possibilitando o restabelecimento da infecção e facilitando a disseminação do bacilo. (Lawn e Zumla, 2011; Koul *et al.*, 2011; Cooper, 2013).

A ocorrência de cepas resistentes aos fármacos de primeira-linha representa uma crescente ameaça ao combate à TB. O aumento dos casos de TB multirresistente a fármacos (MDR-TB) foi observado em diversos países, principalmente nos de alta carga (WHO, 2015; Calligaro *et al.*, 2014). Estas cepas apresentam resistência à isoniazida e à rifampicina e requerem um regime de tratamento que envolve a administração de fármacos de segunda-linha (Tabela 1).

Tabela 1. Principais fármacos utilizados no tratamento da TB

Primeira-linha	Isoniazida
	Rifampicina
	Pirazinamida
	Etambutol
Segunda-linha	Estreptomicina
	Canamicina
	Amicacina
	Capreomicina
	Ciclosserina
	Ofloxacina
	Etionamida
Ácido para-amino salicílico	

Em 2006 a OMS reportou a identificação de cepas extensivamente resistentes a fármacos (XDR-TB), as quais não respondem ao tratamento com isoniazida e rifampicina e também são resistentes a três ou mais medicamentos de segunda-linha (Koul *et al.*, 2011; Calligaro *et al.*, 2014; WHO, 2015).

O tratamento da TB resistente a fármacos é muito mais complexo que o administrado para os casos de TB sensível, envolvendo a administração de quimioterápicos menos potentes e mais tóxicos. A duração do tratamento pode ser superior a 24 meses e possui taxas relativamente baixas de sucesso: em 2015, por

exemplo, apenas 50% dos casos de MDR-TB detectados foram tratados com sucesso (Calligaro *et al.*, 2014; Dartois, 2014; WHO, 2015). O tratamento ideal para esses casos envolve um regime planejado com base nos resultados de testes de susceptibilidade a fármacos (*DST*, do inglês, *drug-susceptibility testing*) realizados com isolados individuais de cada paciente. No entanto, devido à infraestrutura laboratorial inadequada em diversos países, a realização de *DSTs* é limitada e atende apenas 5% dos portadores de MDR-TB (Zumla, Nahid e Cole, 2013; WHO, 2015).

Considerando os baixos índices de sucesso no tratamento dos casos de MDR-TB, os sérios efeitos adversos causados pela maioria dos fármacos utilizados no tratamento da MDR e XDR-TB e que poucos fármacos entram com sucesso para a clínica, destaca-se a necessidade de se buscar novos compostos com atividade anti-TB (Koul *et al.*, 2011).

1.2 O desenvolvimento de fármacos antituberculose

A busca por melhores fármacos anti-TB requer, em um primeiro momento, a identificação e caracterização de alvos moleculares essenciais à sobrevivência do *Mtb*. Um fármaco anti-TB ideal deve apresentar uma série de características, como: (1) ser capaz de diminuir a duração do tratamento, (2) ser menos tóxico, (3) não apresentar interações com fármacos de outros regimes de tratamento, permitindo, por exemplo, a coadministração com fármacos anti-HIV, (4) tratar pacientes portadores de MDR e XDR-TB e (5) ser viável financeiramente (Koul *et al.*, 2011; Ballell *et al.*, 2013; Zumla *et al.*, 2014).

Os principais medicamentos estabelecidos para o tratamento da TB são bastante antigos. A estreptomicina, descoberta em 1944, foi o primeiro antibiótico a apresentar atividade inibitória contra o *Mtb*, consistindo na primeira esperança para o combate da doença. Porém, a sua capacidade de curar os pacientes era questionável e percebeu-se que a administração de um único agente para o tratamento da TB favorece a seleção de cepas resistentes (Zumla, Nahid e Cole, 2013). Na década de 1950, novos fármacos anti-TB foram descobertos, o que permitiu o estabelecimento de um regime combinado de tratamento com duração aproximada de 18 meses. Dentre esses

fármacos estão a isoniazida, pirazinamida, ciclosserina, canamicina e o ácido para-amino salicílico. Contudo, o grande avanço no tratamento da TB foi alcançado na década de 1960 com a introdução da rifampicina. O tratamento combinado com a administração de rifampicina e pirazinamida reduziu a duração do tratamento para seis meses, culminando no estabelecimento do regime atual de tratamento com os quatro fármacos de primeira-linha (Zumla *et al.* 2014).

Nos anos subsequentes poucos avanços foram atingidos no desenvolvimento de novos fármacos e regimes de tratamento para o combate à TB. Porém, grandes esforços foram investidos na área na última década e progressos consideráveis foram alcançados (Payne, 2007; Zumla *et al.*, 2014; WHO, 2015). Em 2012, por exemplo, a agência norte-americana reguladora de medicamentos e alimentos (FDA - US Food and Drug Administration) aprovou o uso da bedaquilina (SIRTURO™) no tratamento de pacientes portadores de MDR-TB, sendo o único fármaco anti-TB aprovado em 40 anos de pesquisa e desenvolvimento (WHO, 2015).

A disponibilidade da sequência genômica do *Mtb* em bases de dados proporcionou a identificação de novos alvos para fármacos anti-TB. Esse conhecimento permitiu a exploração de metodologias envolvendo a seleção de compostos, considerando sua atividade contra as proteínas-alvo do bacilo, e a sua otimização, visando elevar sua eficácia como inibidor, seletividade, atividade celular e propriedades farmacocinéticas (Gilbert, 2013). No entanto, a eficiência dessas abordagens vem sendo questionada ao longo dos anos, principalmente no que se refere ao agente causador da TB, no qual o número de bons alvos moleculares é limitado (Koul *et al.*, 2011; Ballell *et al.*, 2013; Gilbert, 2013). Além disso, os dados gerados em nível molecular nem sempre se aplicam aos modelos animais e em humanos, de maneira que a passagem dos ensaios *in vitro* para os estudos *in vivo* representa um salto difícil de ser superado quando se desenvolve um inibidor baseado no seu alvo molecular (Ballell *et al.*, 2013; Cooper, 2013; Futamura, Muroi e Osada, 2013).

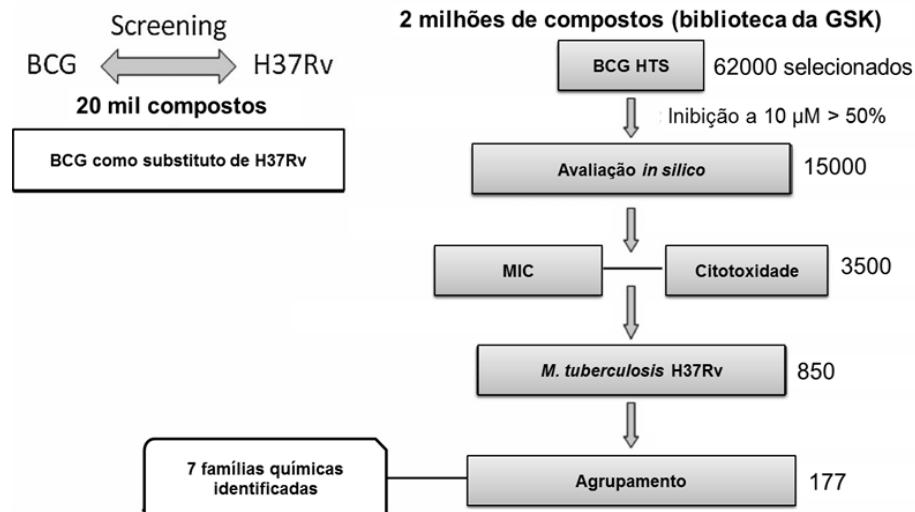
Outra abordagem utilizada para a descoberta de novos fármacos consiste na triagem fenotípica de células. Nessa estratégia, bibliotecas de compostos são testadas para a inibição do crescimento de culturas de células visando a seleção de compostos

que apresentam a atividade de interesse (compostos líder), o que permite a seleção de compostos que superam as limitações relacionadas à permeabilidade e solubilidade (Zumla *et al.*, 2014). Além disso, a realização da triagem de células de mamíferos pode fornecer informações sobre a seletividade, permitindo a seleção de compostos que apresentam baixa citotoxicidade (Gilbert, 2013). A capacidade antimicobacteriana da bedaquilina, por exemplo, foi identificada a partir da triagem fenotípica de diferentes séries de compostos em *M. smegmatis* (Andries *et al.*, 2005).

1.3 GSK e a triagem fenotípica de compostos

Em 2013, pesquisadores da companhia farmacêutica GlaxoSmithKline (GSK) identificaram 177 compostos com potente ação antimicobacteriana e baixa citotoxicidade. Esses compostos foram selecionados a partir de uma biblioteca de mais de dois milhões de estruturas por uma metodologia de triagem fenotípica em larga escala (Figura 3) (Ballell *et al.*, 2013).

Figura 3. Metodologia para identificação de novos agentes antimicobacterianos desenvolvida pela GSK



Fonte: Adaptado de Ballell *et al.* (2013).

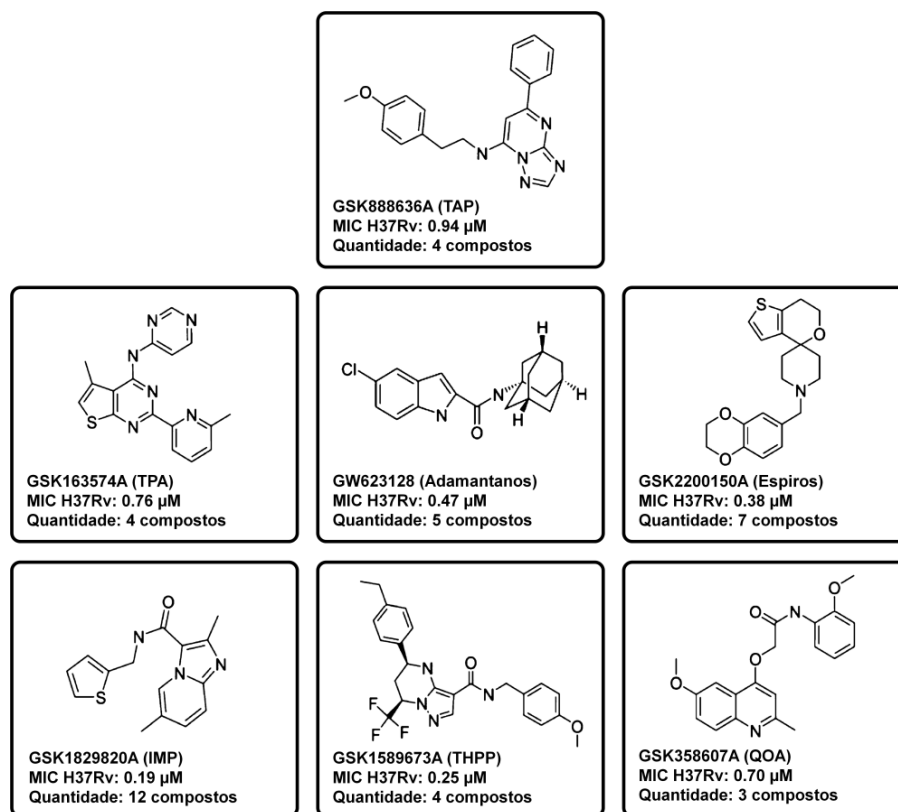
A triagem primária dos compostos foi feita em *M. bovis* BCG, evitando o manuseio de um grande número de material laboratorial infectado por *Mtb*. Para

comprovar que o *M. bovis* BCG é um bom substituto para o *Mtb* H37Rv, um experimento preliminar foi realizado. Para isso, foram selecionados 20.000 compostos da biblioteca reconhecidos por apresentarem uma boa permeabilidade celular, disponibilidade no estoque e que cumprem requisitos químicos, previamente determinados, que conferem potencial farmacológico. Os compostos foram então testados em *M. bovis* BCG e *Mtb* H37Rv em dois experimentos independentes. O número de inibidores identificados para os dois organismos foi muito semelhante: 88 compostos foram identificados como inibidores do *M. bovis* BCG e 73 como inibidores do *Mtb* H37Rv (Ballell *et al.*, 2013; Cooper *et al.*, 2013). Os compostos identificados como inibidores do *M. bovis* BCG foram então testados contra o *Mtb* H37Rv, e vice-versa. Dos compostos inibidores de *M. bovis* BCG, somente 55% foram capazes de inibir o crescimento do *Mtb* H37Rv na concentração de 10 μ M (48 dos 88 compostos) e 86% na concentração de 25 μ M (76 dos 88 compostos). Por outro lado, a maioria dos compostos identificados como inibidores do *Mtb* H37Rv foram capazes de inibir o crescimento de *M. bovis* BCG, 92% na concentração de 10 μ M (67 dos 73 compostos) e 97% na concentração de 25 μ M (71 dos 73 compostos). Esse experimento permitiu atribuir ao *M. bovis* BCG a posição de modelo plausível para a detecção de compostos com atividade inibitória para o *Mtb* H37Rv (Ballell *et al.*, 2013).

Os compostos da biblioteca foram então testados contra *M. bovis* BCG. Aproximadamente 62.000 compostos foram selecionados por apresentarem uma inibição do crescimento da micobactéria maior que 50% na concentração de 10 μ M. Estes compostos foram submetidos a uma avaliação *in silico*, excluindo antibacterianos conhecidos e agrupando aqueles com características físico-químicas semelhantes, resultando em aproximadamente 15.000 compostos (Ballell *et al.*, 2013; Cooper *et al.*, 2013). As concentrações inibitórias mínimas (MIC, do inglês, *minimum inhibitory concentration*) e o potencial citotóxico foram avaliados para *M. bovis* BCG e para a linhagem de hepatócitos HepG2, respectivamente. Os compostos que apresentaram um valor de MIC menor que 10 μ M e um índice terapêutico (razão entre citotoxicidade e MIC) superior a 50 vezes foram selecionados. Os compostos foram então avaliados em *Mtb*, resultando nos 177 compostos agrupados de acordo com suas características

químicas. As classes mais promissoras apontadas pela empresa estão representadas na figura 4 (Ballell *et al.*, 2013; Cooper *et al.*, 2013). Estudos visando à otimização da atividade inibitória de alguns desses compostos já vem sendo explorados por alguns grupos de pesquisa (ver, por exemplo, Kondreddi *et al.*, 2013; Remuinan *et al.*, 2013; Yokokawa *et al.*, 2013 e Pissinate *et al.*, 2016).

Figura 4. Famílias de compostos mais promissoras identificados pela GSK



Fonte: Adaptado de Ballell *et al.* (2013).

Legenda: TAP, triazolopirimidinas; TPA, tienopirimidinas; IMP, imidazopiridinas; THPP, tetrahidropirazolopirimidinas; QOA, quinoloxiacetamidas.

O trabalho conduzido por Ballell e colaboradores teve como objetivo promover uma série de moléculas que podem servir como pontos de partida para o desenvolvimento de novos candidatos a fármacos anti-TB. No entanto, os mecanismos de ação desses compostos ainda precisam ser elucidados.

1.4 Identificação de alvos moleculares

A falta de compreensão do mecanismo de ação de um fármaco pode comprometer o seu estabelecimento na clínica (Sakamoto *et al.*, 2012). Dessa forma, a elucidação do funcionamento de um determinado composto e a caracterização do seu alvo na célula são fundamentais para o desenvolvimento de um novo candidato a fármaco anti-TB (Cooper, 2013; Gilbert, 2013; Butera, 2013; Zumla *et al.*, 2014). Um bom alvo molecular deve ser caracterizado como "*druggable*", ou seja, deve possuir uma grande capacidade de ser modulado por um possível fármaco, provocando uma reação que pode ser avaliada *in vitro* e *in vivo* (Hughes *et al.*, 2011). O termo alvo pode ser empregado a entidades biológicas, tais como genes, RNA e proteínas. Uma vez que o alvo molecular de um composto é identificado, estudos de relação estrutura-atividade guiados por estrutura (do inglês, *structure-guided structure activity relationship studies*) podem ser desenvolvidos, promovendo subsídios para a otimização da atividade do composto (Futamura, Muroi e Osada, 2013). Além disso, a identificação adequada dos alvos possibilita um melhor conhecimento da relação entre o alvo e a doença e permite explorar se a modulação do alvo pode ocasionar efeitos indesejados ao paciente (Hughes *et al.*, 2011).

Ainda em 2013, Martinez-Jimenez e colaboradores propuseram os possíveis alvos moleculares dos compostos identificados pela GSK usando uma abordagem computacional para integrar dados experimentais, propriedades químicas e comparações estruturais dos compostos selecionados (Martinez-Jimenez *et al.*, 2013). Contudo, esses alvos devem ser confirmados experimentalmente, possibilitando avaliar a sua essencialidade para a sobrevivência do *Mtb* e validar o seu potencial farmacológico.

Diversas abordagens são exploradas para identificação de alvos moleculares. A cromatografia de afinidade é a abordagem mais utilizada. Nessa abordagem, a proteína alvo é isolada por cromatografia utilizando o composto imobilizado em uma matriz como coluna cromatográfica. No entanto, a ligação do composto na matriz pode comprometer sua interação com o alvo, comprometendo a eficácia do método (Titov e Liu, 2012).

Outras alternativas fazem uso de ferramentas de genômica funcional para identificação de componentes celulares funcionalmente relevantes, porém não permitem afirmar que os alvos identificados são aqueles que interagem fisicamente com o ligante e de fato induzem a atividade do composto (Schirle e Jenkins, 2016).

Como alternativa experimental para a identificação de alvos moleculares também está a técnica de proteólise pulsada, a qual permite determinar a fração de proteínas diferencialmente degradadas após breve incubação com proteases quando comparadas as amostras tratadas e não tratadas com um ligante. A técnica possui a vantagem de ser de baixo custo e consiste em uma estratégia com potenciais aplicações nas áreas de engenharia de proteínas e na descoberta de fármacos (Park e Marqusee, 2005).

1.5 Identificação de alvos moleculares por proteólise pulsada

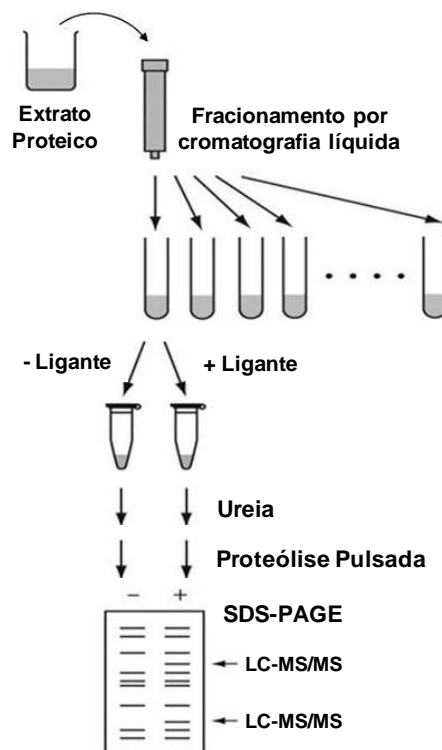
A abordagem da proteólise pulsada para a identificação de alvos busca a identificação de proteínas em um extrato proteico que se ligam a um determinado composto, explorando o efeito dessa ligação na estabilidade conformacional da proteína-alvo (Park e Marqusee, 2005; Liu, Kihara e Park, 2011; Chang *et al.*, 2012). A associação com um ligante é capaz de tornar a proteína mais rígida e menos acessível à ação de proteases, resultando em uma menor suscetibilidade à proteólise. Isso permite comparar a quantidade de proteínas intactas de um extrato proteico bruto após breve digestão proteolítica na presença e na ausência do ligante, sendo possível identificar aquelas em que a presença do ligante desempenhou um efeito protetor (Lomenick *et al.*, 2009; Lomenick, Olsen e Huang, 2011; Liu, Kihara e Park, 2011; Chang *et al.*, 2012).

O procedimento envolve o fracionamento do extrato proteico por cromatografia líquida e a incubação de cada fração com o ligante em estudo. As amostras previamente tratadas com o ligante e seus respectivos controles (incubações sem o ligante) são submetidas a uma breve digestão proteolítica em diferentes concentrações de ureia. Após a interrupção do pulso proteolítico, o conteúdo proteico de cada amostra é separado por SDS-PAGE. As bandas que apresentarem diferenças de intensidade

entre a amostra tratada e controle terão seu conteúdo analisado por cromatografia líquida acoplada à espectrometria de massa (LC-MS) para determinar a proteína que compõe a banda (Figura 5) (Chang *et al.*, 2012).

A etapa de fracionamento cromatográfico do extrato é necessária pelo fato de que o método de análise por comparação das intensidades das bandas de proteínas em SDS-PAGE exige a diminuição da complexidade do extrato, de modo que seja possível distinguir as bandas de proteínas visualizadas no gel.

Figura 5. Procedimento para identificação de alvos moleculares por proteólise pulsada



Fonte: Adaptado de Chang *et al.* (2012).

Uma vez que a estabilidade conformacional varia de uma proteína para outra, é necessário realizar o experimento em diferentes concentrações de um agente desnaturante, de maneira que seja possível a observação de diferenças significativas que permitam concluir que a interação com o ligante protegeu o alvo de sofrer digestão

proteolítica, aumentando significativamente as chances de identificação do alvo (Park e Marqusee, 2005; Liu, Kihara e Park, 2011; Chang *et al.*, 2012).

O experimento deve ser delineado de modo que as proteínas desnaturadas sejam preferencialmente digeridas pela protease em uma amostra. A proteólise de proteínas enoveladas é mais lenta, e a menos que a proteína tenha regiões desestruturadas, a digestão somente será eficiente após a desnaturação. Com base nisso, o delineamento do experimento requer atenção à duração do pulso proteolítico, que deve ser suficiente para digerir apenas as proteínas desnaturadas (Park e Marqusee, 2005; Liu, Kihara e Park, 2011; Chang *et al.*, 2012).

Embora tanto o fracionamento cromatográfico como a análise por SDS-PAGE sejam procedimentos convencionais e relativamente simples, este desenho experimental apresenta desvantagens. Cada fração deve ser analisada individualmente para todas as condições e cada banda que apresentar diferenças entre a incubação tratada com ligante e controle deverá ser analisada individualmente por espectrometria de massa, o que torna o método laborioso e também exige quantidades elevadas de material. Além disso, a análise por SDS-PAGE só permite acessar proteínas visíveis no gel, desconsiderando proteínas pouco abundantes, o que subestima o conteúdo proteico do extrato em análise.

2 JUSTIFICATIVA

A proteólise pulsada para a identificação de alvos moleculares é um método barato e de fácil aplicação, porém os métodos de análise propostos utilizando SDS-PAGE ou eletroforese em gel bidimensional (2D-GE) são limitados. A análise por 2D-GE apresenta baixa reprodutibilidade e é bastante desfavorável. Já análise por SDS-PAGE apresenta maior reprodutibilidade, porém requer o fracionamento cromatográfico do extrato proteico em estudo, o que aumenta drasticamente a quantidade de géis analisados. Além do mais, a análise visual nos géis subestima o conteúdo proteico das amostras, comprometendo a cobertura proteômica do experimento. Considerando as desvantagens da plataforma experimental utilizando SDS-PAGE para identificação de alvos moleculares por meio da proteólise pulsada e a demanda por metodologias para a identificação de alvos moleculares que sejam aplicáveis em plataformas de descoberta de fármacos em larga escala, propomos neste trabalho um método que permite eliminar as etapas de fracionamento do extrato celular e análise visual por eletroforese em gel no procedimento da proteólise pulsada para identificação de alvos moleculares, utilizando como alternativa a análise direta dos peptídeos gerados pelo pulso proteolítico por espectrometria de massa.

3 OBJETIVOS

3.1 Objetivo geral

Desenvolver e validar PePTID, um novo método de identificação de alvos moleculares baseado em proteólise pulsada, seguida de precipitação e análise por espectrometria de massa.

3.2 Objetivos específicos

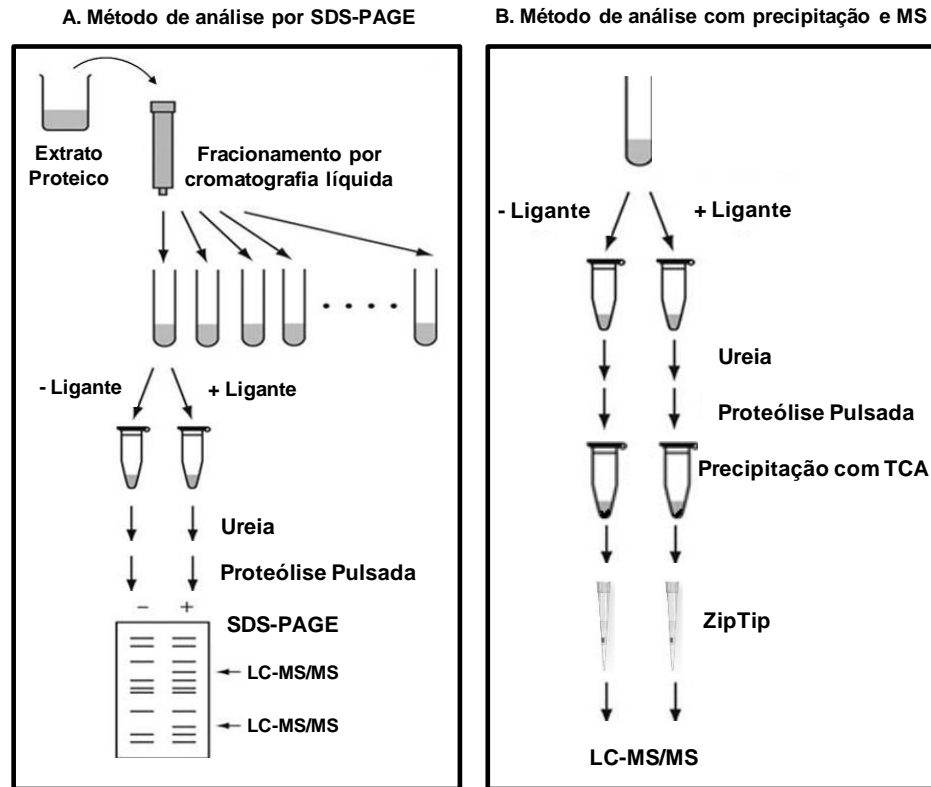
1. Reproduzir o protocolo de identificação de alvos baseado em proteólise pulsada e SDS-PAGE para a identificação de alvos moleculares com base em inibidores com alvos já estabelecidos em *M. smegmatis*
2. Desenvolver o método PePTID e avaliar a eficácia do mesmo para a identificação de alvos moleculares utilizando inibidores com alvos já estabelecidos.
3. Aplicar o método PePTID para identificação de proteínas ligadoras de ATP em *M. smegmatis*.

4 CAPÍTULO I - DESENVOLVIMENTO DO MÉTODO

O protocolo original proposto por Chang e colaboradores (Chang *et al.*, 2012) para a identificação de alvos moleculares utilizando a técnica da proteólise pulsada busca identificar proteínas que foram protegidas da degradação proteolítica devido a interação com o ligante, buscando por SDS-PAGE bandas que representem proteínas que permanecem intactas ou que foram parcialmente degradadas. Como alternativa, propomos nesse trabalho um método simplificado de análise semiquantitativa, que busca identificar a interação entre o ligante e a proteína avaliando as diferenças nos peptídeos gerados por um pulso proteolítico. Os dois métodos são comparados na figura 6.

As etapas de incubação e proteólise pulsada seguem o protocolo descrito por Chang e colaboradores (Chang *et al.*, 2012). Após a interrupção do pulso proteolítico, propomos a realização de uma etapa de precipitação usando ácido tricloroacético (TCA) para eliminar os segmentos das proteínas que permanecem intactas ou que foram parcialmente degradadas, desta forma, o material analisado consiste apenas nos peptídeos gerados pela degradação proteolítica. Feito isso, o sobrenadante é analisado por LC-MS/MS e os conteúdos das incubações são comparados para a identificação das proteínas nas quais o ligante influenciou a degradação pela protease. O procedimento dispensa a realização do fracionamento do extrato.

Figura 6. Comparação esquemática dos procedimentos experimentais



Fonte: Adaptado de Chang *et al.* (2012).

Nota: A. Método proposto por Chang e colaboradores (Chang *et al.*, 2012). O extrato proteico é fracionado por cromatografia líquida e cada fração é incubada com o ligante (amostra tratada) e sem o ligante (amostra controle) em diferentes concentrações de ureia. Para cada uma das incubações, é executado o protocolo da proteólise pulsada e análise por SDS-PAGE. As bandas do gel que apresentarem diferenças na comparação entre a amostra tratada com o ligante e controle terão seu conteúdo analisado por LC-MS/MS. B. Método de análise com precipitação e espectrometria de massa (PePTID) (MS). As incubações com e sem o ligante são realizadas com o extrato bruto em diferentes concentrações de ureia. O pulso proteolítico é executado em cada amostra e o protocolo de precipitação com ácido tricloroacético (TCA) é executado. Os sobrenadantes são submetidos a um protocolo de ziptip para neutralização do tampão e analisadas por MS.

4.1 Métodos

O trabalho foi desenvolvido no Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF) do Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB), o qual está localizado no Parque Tecnológico da Pontifícia Universidade Católica do Rio Grande do Sul (TECNOUC).

4.1.1 Preparação do extrato proteico de *Mycobacterium smegmatis*

Células de *Mycobacterium smegmatis* MC2 155 foram cultivadas em meio Lysogeny Broth (LB) com 0.05% de Tween 80 (Sigma-Aldrich) para evitar a formação de agregados celulares. As culturas foram incubadas a 37 °C sob agitação em shaker (180 RPM) (New Brunswick Scientific Innova 44) durante três dias, coletadas por centrifugação (11800 x g, 30 min, 4 °C) (Hitachi himac CR21G), lavadas com água bidestilada e ressuspensas em tampão de lise (Tris-HCl 20 mM, pH 8,0) contendo ácido etilenodiaminotetracético (EDTA - Sigma-Aldrich) 1 mM e ditioneitol (DTT - Sigma-Aldrich) 1 mM. As células foram rompidas por ultrassom (Sonic VCX-400 Ultrasonic Cell Disruptor) no gelo (15 pulsos de 20 segundos a uma amplitude de 60%). O extrato foi centrifugado (38900 x g, 30 min, 4 °C) para a separação das frações solúvel e insolúvel. A fração solúvel foi coletada e incubada com 50 U/mL de benzonase (Sigma-Aldrich) e cloreto de magnésio (MgCl₂ - Merck) 5 mM a 25 °C durante 1 h e 30 min para a hidrólise dos ácidos nucleicos. EDTA e cloreto de sódio (NaCl - Synth) foram adicionados ao lisado numa concentração final de 10 mM e 25 mM, respectivamente. Os metabólitos no extrato foram removidos por uma coluna de gel filtração HiTrap desalting (GE Healthcare) tendo Tris-HCl (pH 8,0) contendo EDTA 1 mM, DTT 1 mM e NaCl 25 mM como tampão de eluição.

4.1.2 Amplificação, expressão e purificação da proteína 2-trans-enoil-ACP (CoA) redutase (InhA) de *M. tuberculosis*

O gene *inhA* de *M. tuberculosis* H37Rv foi amplificado por PCR, clonado e superexpresso em *E. coli* segundo o protocolo descrito por Quémard e colaboradores (Quemard *et al.*, 1995).

Seis litros de meio de Luria Bertani contendo 50 mg/ml de carbenicilina foram inoculados com um pré-inóculo cultivado durante a noite, cresceram até atingir uma densidade ótica a 600 nm entre 0,8 e 1, induzidas com 1 mM de isopropil- β -D-tiogalactopiranosídeo (IPTG), e cultivadas durante 2 h. As células foram colhidas por centrifugação (11800 x g, 30 min, 4 °C), e a proteína InhA foi purificada. A purificação foi realizada a 4 °C. As células foram rompidas por sonicação e o sobrenadante obtido após centrifugação (38900 x g, 30 min, 4 °C) foi tratada com sulfato de estreptomicina (1% w/v final) para remoção dos ácidos nucleicos. O sobrenadante obtido após centrifugação (38900 x g, 30 min, 4 °C) foi dialisada e aplicada a uma coluna cromatográfica de troca aniônica Q Sepharose de fluxo rápido (Pharmacia). As proteínas foram eluídas usando um gradiente linear de 0-1 M de NaCl, as frações contendo a proteína InhA foram reunidas, dialisadas, concentradas por ultrafiltração e aplicadas a uma coluna de gel filtração Superdex 75 (Pharmacia). As frações contendo a proteína InhA foram reunidas e aplicadas a uma coluna de troca aniônica de alto desempenho Mono Q (Pharmacia) e eluídas com um gradiente não linear de 0-0,6 M de NaCl. As frações contendo a proteína InhA foram reunidas e a análise por SDS-PAGE exibiu uma única banda correspondente ao peso molecular da InhA (28,5 kDa). A solução final contendo a proteína homogênea foi concentrada para aproximadamente 10 mg/mL. A enzima purificada foi estocada em tampão PIPES 100 mM, pH 7,3, contendo 50% de glicerol a -80 °C (Quemard *et al.*, 1995; Basso *et al.*, 1998).

4.1.3 Incubações

Uma quantidade de extrato de proteínas foi incubada com (tratamento) e sem (controle) o inibidor triclosan, para uma concentração final de 100 μ M, em tampão

adequado contendo $MgCl_2$ 5 mM e DTT 1 mM durante uma hora a 25 °C. As quantidades de proteína e o tampão de cada incubação estão descritas na tabela 2.

Tabela 2. Quantidades de proteínas e tampões das incubações

Incubação 1	Solução de proteína <i>mTlnA</i> (totalizando 18 µg de proteínas) em tampão PIPES 100 mM (ph 7,3)
Incubação 2	Extrato de <i>M. smegmatis</i> (totalizando 100 µg de proteínas) e <i>mTlnA</i> (4,5 µg) em tampão PIPES 100 mM (ph 7,3)
Incubação 3	Extrato de <i>M. tuberculosis</i> (totalizando 100 µg de proteínas) em tampão Tris-HCl 20 mM (ph 8)

4.1.4 *Proteólise pulsada e precipitação*

A proteólise pulsada foi iniciada com a adição de termolisina (Sigma-Aldrich) numa concentração de 0,20 mg/mL. Após 1 minuto, a reação foi interrompida com a adição de fosforamidon (Sigma-Aldrich) 60 µM e EDTA 20 mM. Ácido Tricloroacético (Sigma-Aldrich) foi adicionado a ambas as incubações na concentração final de 20% para precipitação de proteínas intactas ou parcialmente degradadas.

4.1.5 *Preparo da amostra para a análise por Espectrometria de Massa (MS)*

Fragmentos de proteínas e peptídeos oriundos da proteólise pulsada foram purificados pelo método de ziptip utilizando resina POROS 50 R2, com uma solução de acetonitrila 90% e ácido fórmico 0,1% como tampão de eluição. As amostras foram secas por SpeedVac (centrífuga a vácuo). Os pellets foram ressuspensos em Tris 500 mM, pH 7,5. Ditiotretitol (DTT) foi adicionado na concentração final de 5 mM durante 20 min a 37 °C para a redução das pontes dissulfeto e, em seguida, iodoacetamida foi adicionada na concentração final de 25 mM durante 20 minutos à temperatura ambiente no escuro para a alquilação das cisteínas. Por fim, as amostras foram digeridas com tripsina (0,5 µg) durante a noite a 37 °C. Ácido fórmico foi adicionado para interromper a reação (5% v/v, concentração final).

As separações cromatográficas das misturas complexas de peptídeos foram realizadas utilizando o sistema nanoLC Ultra (nanoLC Ultra 1D plus, Eksigent, EUA) equipado com um autosampler nanoLC AS-2 (Eksigent, EUA) e conectado a um espectrômetro de massas híbrido LTQ-XL Orbitrap Discovery (Thermo Fischer Scientific, EUA) contendo uma fonte de ionização nano-eletrospray (Thermo Fischer Scientific, EUA). As colunas capilares analíticas (diâmetro interno de 100 μm e 20 cm de comprimento com ponteira de 5 μm de diâmetro interno) e pré-colunas (diâmetro interno de 150 μm e 2 cm de comprimento), foram empacotadas no laboratório com resina de fase reversa (5 μm ODS-AQ C18 Yamamura Chemical Lab, EUA).

As amostras foram carregadas no sistema pelo *autosampler* com fluxo de 1 μL por minuto por 15 minutos. Foram utilizados 10 a 20 % da amostra digerida em um volume final de injeção de 10 μL . O método cromatográfico de 120 minutos empregou o seguinte gradiente descontínuo da fase móvel A (5% de acetonitrila e 0,1% de ácido fórmico em água) à fase móvel B (90% de acetonitrila e 0,1% de ácido fórmico): 0-5% B em 5 min; 5-25% B em 60 min; 25-50% B em 20 min; 50-80% B em 15 min; 80% B isocrático por 5 min; 80-5% B em 1 min; 5% B isocrático por 14 min com fluxo de 400 nL/min.

A eluição e emissão em fonte de nano-eletrospray dos peptídeos ocorreram diretamente da extremidade da coluna analítica. O instrumento foi programado para ciclos de aquisição, em modo de ionização positiva, de um espectro de MS1 no intervalo de razão massa/carga (m/z) de 400-1600 com resolução de 30.000 (em m/z 400) no Orbitrap, seguido de oito espectros de MS2 dos oito íons mais abundantes do espectro de MS1. Dissociação induzida por colisão (CID) foi utilizado como método de fragmentação, com uma ativação Q de 0,250, um tempo de ativação de 30 ms, energia de colisão normalizada de 35%, e uma amplitude de isolamento de 1 Da. Para explorar a complexidade da amostra e adquirir espectros de peptídeos menos abundantes, massas de peptídeos adquiridos foram adicionadas a uma lista de exclusão dinâmica com tamanho de 100 íons e tempo de permanência de 30 segundos durante a aquisição dos espectros MS2.

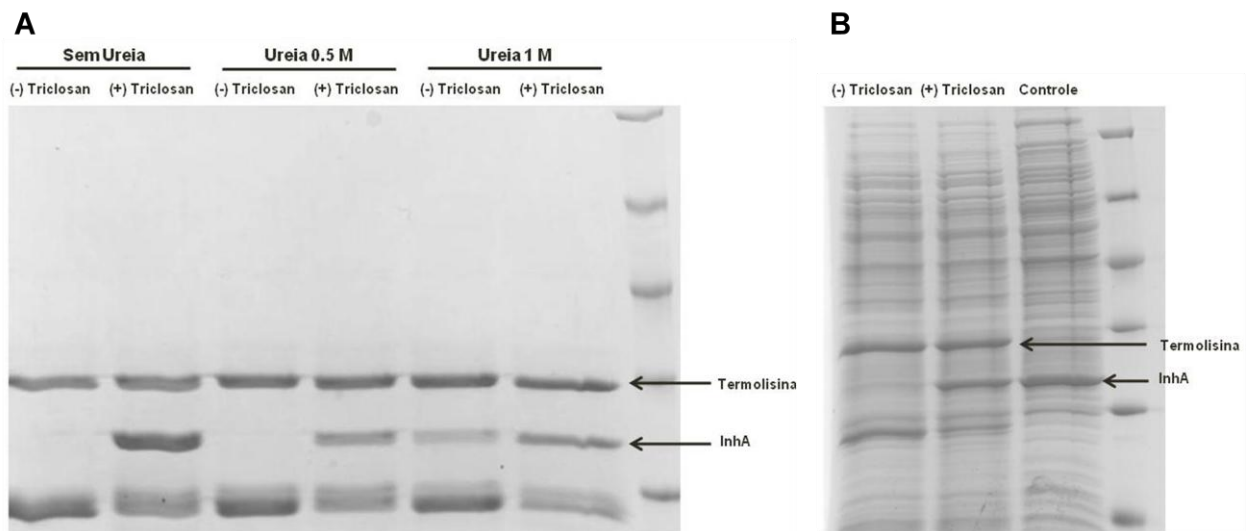
4.1.6 Análise de dados

A identificação de peptídeos e proteínas foi realizada com software Patternlab for Proteomics (Carvalho *et al.*, 2016). Espectros de MS2 foram buscados diretamente dos arquivos RAW pelo software de busca Comet (Eng, Jahane Hoopmann, 2013) utilizando banco de dados não-redundante do proteoma referência de *M. smegmatis* (Proteome ID UP000000757, www.uniprot.org) contendo também a sequência da proteína InhA de *Mtb* (P9WGR1, www.uniprot.org). Os dados foram buscados com tolerância de 50 ppm para íons precursores e 1 Da para íons fragmento. Os espectros identificados foram filtrados usando o software Search Engine Processor (Carvalho *et al.*, 2016). Os peptídeos identificados foram agrupados de acordo com o estado triptico e carga. Os parâmetros Xcorr, DeltaCN, DeltaMass, Peaks Matched e Spec Count Score foram usados para gerar um escore bayesiano. Foi estabelecido um valor de corte para aceitar 1% de falso positivo baseado nas identificações do banco de dados reverso. Adicionalmente, um comprimento de sequência mínimo de 6 resíduos foi usado. Os resultados foram pós-processados para acertar identificações com menos de 10 ppm de variação de massa.

4.2 Resultados

As incubações 1 e 2 (ver Tabela 2) foram realizadas com (+) e sem (-) o triclosan nas concentrações de 100 μ M e 1 mM, respectivamente. O triclosan é um inibidor conhecido da proteína 2-trans-enoil-ACP (CoA) redutase (InhA). Cada amostra foi submetida ao procedimento de proteólise pulsada e analisada por SDS-PAGE, conforme o descrito por Chang e colaboradores (Chang *et al.*, 2012). Nas incubações sem o triclosan, a proteína InhA foi completamente digerida nas concentrações de 0 e 0.5 M de ureia, enquanto que na presença do triclosan ainda é possível observar a banda correspondente a proteína InhA (Figura 7A). O efeito protetor exercido pelo triclosan também foi observado em uma amostra complexa, contendo extrato proteico de *M. smegmatis* (Figura 7B).

Figura 7. SDS-PAGE das amostras submetidas a proteólise pulsada



Nota: A. SDS-PAGE das incubações contendo a proteína InhA com (+) e sem (-) triclosan 100 μ M (Incubação 1, ver Tabela 2) nas concentrações de 0, 0.5 e 1 M de ureia. B. SDS-PAGE das incubações contendo o extrato proteico de *M. smegmatis* e a proteína InhA (Incubação 2, ver Tabela 2) com (+) e sem (-) triclosan 1 mM. A amostra controle não foi submetida ao protocolo de proteólise pulsada.

Um experimento inicial foi realizado para avaliar a possibilidade de observar a proteção da degradação proteolítica por intermédio dos peptídeos gerados durante o pulso proteolítico por LC-MS. Para isso, foram executados os protocolos da proteólise pulsada e de precipitação com ácido tricloroacético (TCA) em incubações contendo a proteína InhA e triclosan (Incubação 1, ver Tabela 2). A proteína foi incubada com e sem o ligante durante 60 minutos e submetida a um pulso proteolítico de 1 minuto.

Os dados da cobertura de sequência da proteína e contagens espectrais e de peptídeos, quando comparadas as duas amostras, mostraram diferenças que permitem concluir que o ligante exerceu efeito protetor na proteína. As reduções da abundância relativa dos peptídeos e da cobertura de sequência na presença do ligante apontam uma redução significativa na degradação proteolítica exercida pela protease. Na ausência do ligante, são observados 57 peptídeos, 256 contagens espectrais e uma cobertura de sequência de aproximadamente 78% enquanto que na presença do ligante, são observados 15 peptídeos, 46 contagens espectrais e uma cobertura de sequência de 36% (Figura 8). A proteção também foi observada utilizando o mesmo ligante em incubações contendo extrato de *M. smegmatis* e a proteína InhA de *Mtb* (Incubação 2, Figura 9), atestando a fiabilidade da análise mesmo em amostras complexas.

Figura 8. Cromatograma, sequência da proteína InhA e contagens espectrais do procedimento realizado utilizando a proteína InhA purificada

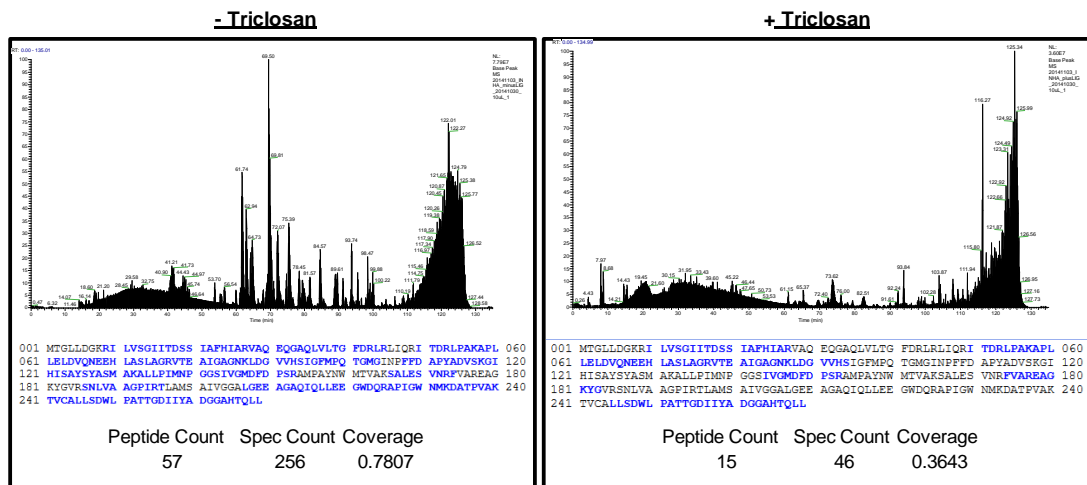
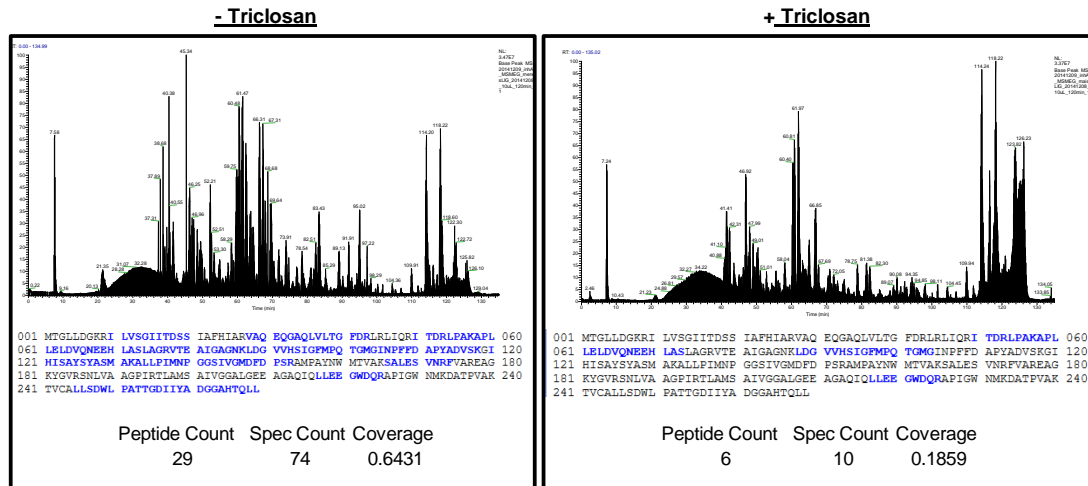


Figura 9. Cromatograma, sequência da proteína InhA e contagens espectrais do procedimento realizado utilizando o extrato proteico de *M. smegmatis* em conjunto com a proteína InhA purificada



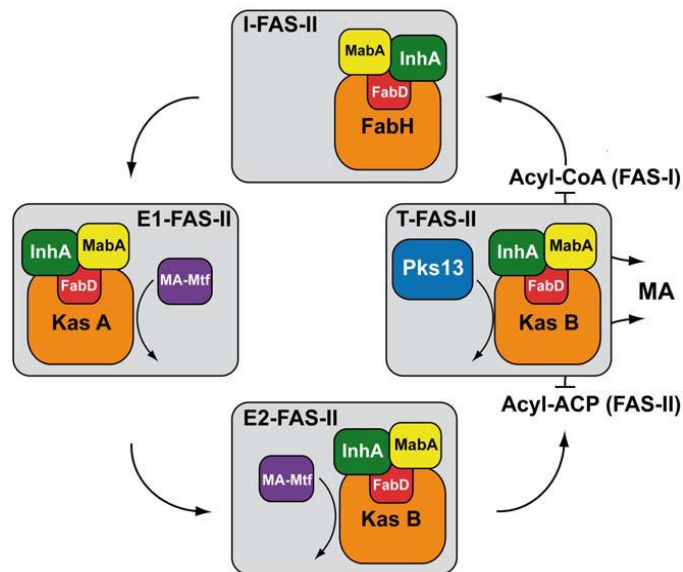
Como os experimentos anteriores utilizam quantidades excessivas da proteína alvo, o procedimento foi realizado com o extrato bruto, sem a adição da proteína alvo purificada, para constatar a capacidade do método de acessar proteínas pouco abundantes. O procedimento foi então executado incubando apenas o triclosan com o extrato proteico de *M. tuberculosis*, sem adição da proteína InhA purificada. Peptídeos correspondentes a proteína InhA foram identificados somente na incubação que não continha o ligante, porém em baixa abundância. Curiosamente, diferenças significativas também foram identificadas para a proteína β -cetoacil-ACP sintase B (KasB) (Tabela 3).

Tabela 3. Valores da cobertura de sequência e contagens espectrais e de peptídeos para as proteínas InhA e KasB nas incubações tratadas com o ligante e controle

	+ Triclosan			- Triclosan		
	Contagem de peptídeos	Contagem espectral	Cobertura	Contagem de peptídeos	Contagem espectral	Cobertura
InhA	-	-	-	2	3	0.0706
KasB	2	4	0.0935	13	18	0.3118

Assim como a *InhA*, a *KasB* também faz parte do sistema FAS-II (sistema de síntese de ácidos graxos – do inglês, *Fatty Acid Synthase*). Em conjunto com o sistema FAS-I, o sistema multienzimático FAS-II tem como função sintetizar os ácidos micólicos, que são os principais componentes do envelope celular micobacteriano. Estudos envolvendo sistemas duplo e triplo híbrido (Y2H e Y3H) e ensaios de coimunoprecipitação mostraram a organização das enzimas do sistema FAS-II em complexos multienzimáticos. Neste modelo, três tipos de complexos são interconectados: o complexo FAS-II de iniciação (I-FAS-II), composto pelas proteínas *MabA*, *InhA* e *FabD* (que representam o cerne do complexo) em conjunto com a *FabH*; dois complexos FAS-II de alongamento, compostos pelo cerne em conjunto com as proteínas *KasA* (E1-FAS-II) ou *KasB* (E2-FAS-II); e o complexo de terminação (T-FAS-II), composto pelo cerne e pela proteína *Pks3* associada com a *KasB* (Figura 10) (Cantaloube *et al.*, 2011).

Figura 10. Organização dos complexos multienzimáticos do sistema FAS-II



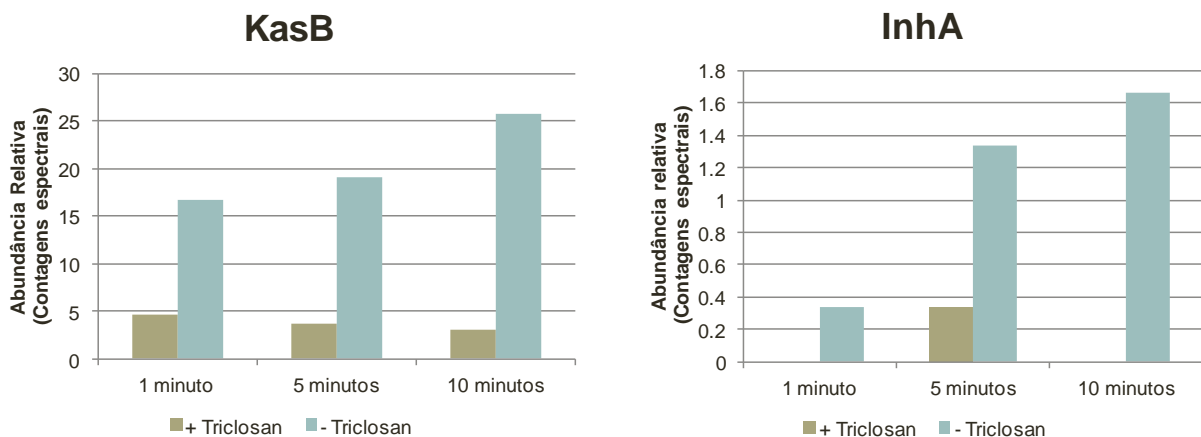
Fonte: Cantaloube *et al.* (2011).

Nota: I-FAS-II, complexo FAS-II de iniciação; E1-FAS-II e E2-FAS-II, complexos FAS-II de alongamento 1 e 2; T-FAS-II, complexo FAS-II de terminação; *InhA*, 2-trans-enoil-ACP (CoA) redutase; *MabA*, 3-oxoacil-ACP redutase *FabG1*; *FabD*, malonil CoA-ACP transacilase *FabD*;

FabH, 3-oxoacil-ACP sintase III FabH; KasA, 3-oxoacil-ACP sintase 1; KasB, 3-oxoacil-ACP sintase 2; Pks13, policetídeo sintase Pks13; MA, ácido micólico.

A organização do sistema FAS-II em complexos multienzimáticos justifica a identificação da proteína KasB utilizando o triclosan como ligante, pois a susceptibilidade proteolítica pode se manifestar na degradação da proteína KasB uma vez que a interação do ligante com a InhA pode elevar a estabilidade conformacional do complexo de maneira global. Como complemento para esse resultado, o procedimento foi realizado com pulsos proteolíticos de 1, 5 e 10 minutos e observa-se que a abundância relativa dos peptídeos que correspondem as proteínas InhA e KasB permanecem constantes nas incubações que contém o ligante, porém na ausência do ligante a abundância relativa cresce proporcionalmente com o aumento da duração do pulso (Figura 11).

Figura 11. Gráficos das abundâncias relativas para as proteínas KasB e InhA em função da duração do pulso proteolítico



Os dados permitem atribuir que o método analítico proposto é confiável para a identificação de alvos moleculares de compostos. É importante ressaltar que a metodologia pode fornecer uma lista de proteínas nas quais o ligante exerce uma possível atividade protetora, de modo que estudos posteriores devem ser realizados para validar a interação da proteína proposta com o ligante.

5 CAPÍTULO II - APLICAÇÃO DO MÉTODO PePTID

Para validação do método, o procedimento utilizando precipitação e análise direta por MS foi realizado para a identificação de proteínas que ligam ATP utilizando como ligante um análogo não hidrolisável do ATP (ATP γ S). O ATP foi escolhido como ligante para a validação do método por possuir interação documentada com diversas proteínas. Outros trabalhos descrevendo métodos para identificação de alvos, fazendo uso da proteólise pulsada (Chang *et al.*, 2012; Liu *et al.*, 2011) ou até mesmo outras abordagens experimentais (Adachi *et al.*, 2014), também utilizam o ATP como ligante para validação dos métodos e podem contribuir com a fundamentação deste trabalho. O método foi denominado PePTID, *Pulse Proteolysis and Precipitation for Target Identification* (em português, Proteólise pulsada e precipitação para a identificação de alvos). Os dados obtidos utilizando o ATP γ S apontam interação do ligante com 32 proteínas em apenas uma concentração de ureia, sendo o maior número de proteínas acessadas em um único experimento visando a identificação de alvos moleculares. Os dados estão descritos na forma de artigo científico e serão submetidos para o periódico *Journal of Proteome Research* (IF 4,245, JCR 2014).

5.1 Artigo científico

PePTID – Pulse Proteolysis and Precipitation for Target Identification

*Rogério V. Trindade, Antônio F. M. Pinto, Diógenes S. Santos, Cristiano V. Bizarro**

Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF), Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), 6681/92-A. Av. Ipiranga – TECNOPUC – Prédio 92A, 90619-900, Porto Alegre, RS, Brazil.

KEYWORDS: PePTID, Pulse Proteolysis, Target Discovery, Energetics-based target discovery approach.

ABSTRACT: In recent years, phenotypic screening is assuming a leading role in drug discovery efforts. However, the development of new drugs from bioactive compounds obtained in screening campaigns requires the identification of the cellular targets responsible for their biological activities. A new energetics-based method for target identification is presented: PePTID - Pulse Proteolysis and Precipitation for Target Identification. In this method, proteins incubated with or without a ligand and submitted to a brief proteolytic pulse are directly analyzed and compared using a label-free semi-quantitative mass spectrometry strategy, dispensing the SDS-PAGE readout, and greatly improving the throughput. As a proof-of-concept, we apply the PePTID

method to identify ATP-binding proteins in *Mycobacterium smegmatis*, a model system for *Mycobacterium tuberculosis*, the etiological agent of tuberculosis (TB).

INTRODUCTION

Contrary to early expectations,¹ the advances in the post-genomic era have not boosted drug discovery efforts.² The shift towards a target-centric approach for drug discovery could have contributed to the decline in the productivity of pharmaceutical industry.^{3,4} As an example, even with most companies adopting a target-based drug discovery (TDD) approach, the majority of small-molecule first-in-class drugs approved by the US Food and Drug Administration (FDA) between 1999 and 2008 were discovered using phenotypic assays.⁵ Therefore, a “neoclassic” phenotypic approach, similar to early physiology-based strategies but enhanced by recent technical advances, has gained ground.^{2,6} An advantage of phenotypic drug discovery (PDD) approaches is the absence of a preconceived idea about neither the molecular mechanism of action nor the target identity of drug candidates.⁶ This is particularly important for antibacterial discovery campaigns, since, aside few exceptions, there have been only three targets or pathways explored: the ribosome, cell wall synthesis, and DNA topoisomerase.⁷

However, both the elucidation of the mechanism of action of a bioactive compound and the identification of its cellular targets are critical for development of new drugs.⁸ Once the molecular targets of bioactive compounds are identified, structure-guided SAR studies can be performed, providing essential information for the rational optimization of compound’s activity.⁹ Also of importance is the identification of unknown off-targets, even for compounds with known mechanisms of action, as these interactions could raise undesired effects. Therefore, the

identification of these off-targets might give subsidies for selection of more promising compounds for pre-clinical development, reducing attrition rates and increasing R&D productivity.^{5, 10, 11}

New approaches for drug target identification were developed in the last years driven by the need for methods which are high-throughput and suitable for analysis of complex biological mixtures.¹² In this context, energetics-based approaches are promising alternatives for target identification, as the binding reaction takes place in solution phase, eliminating the need for affinity tags or any other chemical modification on the bioactive compound of interest. These approaches rely on the effect of ligand binding on the conformational stability of the target protein: drug binding frequently increases its thermodynamic stability and consequently slows down its unfolding.^{13, 14}

Pulse proteolysis is an energetics-based method for identification of molecular targets, which determines the fraction of unfolded protein after a brief incubation with a protease.¹³ Some advantages of pulse proteolysis are: no need to modify test molecules, eliminating possible adverse effects on binding; the experiment stringency can be modulated by changing ligand concentration; and protein-ligand interactions of any nature can be identified, such as interactions with substrates, regulatory and signaling molecules, inhibitors and drugs.¹³ It can also be used to detect variations in unfolding thermodynamics under different denaturing conditions.¹³ Moreover, pulse proteolysis is a low-cost method, with potential applications in the fields of protein engineering and drug discovery.¹⁵

A simplified approach for energetics-based target identification using pulsed proteolysis was previously reported.¹⁴ In order to reduce the complexity of cell lysates, samples are firstly fractionated by ion exchange chromatography. Each isolated fraction is then treated with different

concentrations of urea to induce protein unfolding with varying stringencies in the presence or in the absence of the ligand of interest, followed by pulse proteolysis using thermolysin. Then, proteins displaying different proteolytic susceptibilities upon unfolding in the presence of the ligand, when compared with control samples, are identified using SDS-PAGE and mass spectrometry. This strategy was applied to identify previously known but also novel ATP-binding proteins in *Escherichia coli*.¹⁴

Here, we report a new energetics-based approach using pulse proteolysis and precipitation for target identification: PePTID. Instead of looking for differences in proteolytic susceptibilities by visual inspection of protein samples after SDS-PAGE^{14, 16} or 2D gel electrophoresis (2D-GE),¹³ we focus our analysis on the small peptides generated after the pulse-proteolysis step. We apply a label-free semi-quantitative approach to directly detect and identify the differences in the peptide content of control and treated samples in an automated way, without relying on visual inspection of samples. Data from technical and biological replicates can be easily integrated into a single analysis, adding confidence on the results obtained. Moreover, the PePTID approach greatly reduces sample complexity, abrogating the need for sample fractionation before pulse proteolysis and ligand binding, simplifying the protocol and minimizing both sample and ligand requirements. As a proof-of-concept, we apply this approach to identify ATP-binding proteins in *Mycobacterium smegmatis*, a model system for *Mycobacterium tuberculosis*, the etiological agent of tuberculosis (TB).

MATERIALS AND METHODS

***Mycobacterium smegmatis* lysate extraction**

Mycobacterium smegmatis MC2 155 was grown in Lysogeny Broth media (LB) with 0.05% Tween 80 (Sigma-Aldrich). The culture was incubated with shaking (180 RPM) (New Brunswick Scientific Innova) at 37 °C for three days, collected by centrifugation (11,800 x g, 30 min, 4 °C) (Hitachi himac CR21G), washed with bi-distilled water and resuspended in buffer Tris-HCl 20 mM, pH 8.0 containing 1 mM Ethylenediamine tetraacetic acid (EDTA - Sigma-Aldrich) and 1 mM dithiothreitol (DTT - Sigma-Aldrich). The cell lysate protocol was adapted from Chang et al.¹⁴ Briefly, the cells were lysed by sonication (Sonic VCX-400 Ultrasonic Cell Disruptor) in ice (15 pulses of 20 s with amplitude of 60%) and centrifuged (38,900 x g, 30 min, 4 °C) (Hitachi himac CR21G). The supernatant was collected and incubated with 50 U/mL of benzonase (Sigma-Aldrich) and 5 mM of MgCl₂ (Merck) at 25 °C for 1.5 h to hydrolyze nucleic acids. EDTA and NaCl (Synth) were added to the final concentration of 10 mM e 25 mM, respectively. To remove metabolites and nucleic acid debris a gel filtration chromatography was performed with a 5 mL HiTrap desalting column.

ATP γ S incubation and pulse proteolysis

The lysate (200 μ g of proteins) was incubated with (treatment) and without (control) 1 mM Adenosine 5'-[γ -thio]triphosphate (ATP γ S) (Sigma-Aldrich) in 20 mM Tris-HCl (pH 8.0) buffer containing 5 mM MgCl₂ and 1 mM DTT for 10 min and then urea (USB) was added to the final concentrations of 0, 1, 2, 3 and 4 M for 1.5 hours at 25 °C, in a final volume of 90 μ L. Proteolysis was started by adding 0.20 mg/mL thermolysin (Sigma-Aldrich). After 1 min, the reactions were quenched by adding phosphoramidon (Sigma-Aldrich) and EDTA to the final

concentrations of 60 μ M and 20 mM, respectively. Trichloroacetic acid (Sigma-Aldrich) was added to both reactions to a final concentration of 20%, incubated for 10 min at 4 °C and centrifuged (17,000 x g, 5 min, 4 °C) to precipitate and remove intact and partially degraded proteins. The peptide content of both reactions was purified and enriched by chromatography using POROS 50 R2 resin in a homemade pipette tip column, with 0.1 % formic acid/90% acetonitrile solution as elution buffer. The samples were dried by speedvac (vacuum centrifuge). Experiments were performed in biological triplicates for 1 M urea incubations. For the other urea concentrations just one incubation were performed.

Protein Digestion

Peptide pellets were resuspended in 500 mM Tris-HCl, pH 7.5, and digested with a protocol adapted from Klammer and MacCoss.¹⁷ Briefly, disulphide bonds were reduced in 5 mM dithiothreitol (DTT) for 20 min at 37 °C and then cysteines were alkylated in 25 mM iodoacetamide for 20 min at room temperature in the dark. Peptides were digested with trypsin (0.5 μ g) overnight at 37 °C. Formic acid was added to quench the reaction (5% v/v, final concentration).

LC/MS-MS Analysis

Samples were analyzed in technical triplicates via LC-MS/MS. Chromatographic separations of peptides from the complex mixtures were performed on an Eksigent nanoLC Ultra 1D plus with AS-2 autosampler (Sciex, USA) connected to a LTQ-XL Orbitrap Discovery hybrid instrument (Thermo Fischer Scientific, USA). Fused-silica capillary pre-columns and analytical columns (150 μ m ID/360 μ m OD and 100 μ m ID/360 μ m OD with 5 μ m ID tip, respectively; Polymicro

Technologies, Phoenix, USA) were prepared in house by slurry packing 2 cm and 20 cm of C18 reversed phase resin (5 μm ODS-AQ C18 Yamamura Chemical Lab, USA), respectively.

Samples (10% of the samples) were loaded into the system through the autosampler at 1 $\mu\text{L}/\text{min}$ for 15 min. Mobile phases were 5% acetonitrile/0.1% formic acid as buffer A, and 90% acetonitrile/0.1% formic acid as buffer B. Gradient elution was 4 h long at 300 nL/min, as follows: 0-5% B in 5 min; 5-25% B in 120 min; 25-50% B in 65 min; 50-80% B in 40 min; 80% B for 5 min; 80-0% B in 1 min; 5% B for 14 min.

Data were collected in a top 8 mode, meaning that each FT full-scan (400-1800 m/z range; 30,000 resolution) was followed by data-dependent CID MS/MS spectra acquisition of the 8 most intense ions in the ion trap, with dynamic exclusion applied. Electrospray voltage was set to 2.3 kV, capillary temperature to 275 $^{\circ}\text{C}$ and voltage to 35 V, and tube lens voltage was set to 110 V. The mass spectrometer and HPLC were controlled by the Xcalibur data system (Thermo Fischer Scientific, USA).

Data analysis

Tandem mass spectra were searched directly from Thermo RAW files with the software Comet¹⁸ in the platform PatternLab for Proteomics (Version 3.2.0.3).¹⁹ The database contained a non-redundant *M. smegmatis* reference proteome (Proteome ID UP000000757, www.uniprot.org) and the reverse sequences of all entries. Search space included all fully and semitryptic peptide candidates. Carbamidomethylation of cysteine was considered as static modification. The validity of the peptide spectra matches (PSMs) was assessed using Patternlab for Proteomics module Search Engine Processor (SEPro), with a false discovery rate of 1%. Identified proteins were grouped by maximum parsimony. Approximately area-proportional Venn diagram analysis was

performed considering a minimum 2 replicates per class to access proteins totally stabilized and totally destabilized by the ligand. Proteins identified in both treatment and control samples were compared by TFold analysis,²⁰ using spectral counting for fold change calculation, minimum 2 replicates per class, 0.02 BH q-value (2% FDR), 0.4 L-Stringency and optimized F-Stringency.

RESULTS AND DISCUSSION

PePTID Strategy

The PePTID approach is depicted in Figure 1. The whole cell lysate, without previous fractionation, is separated in a treatment and a control sample. The ligand is added to the treatment sample and allowed to interact with the targets we want to identify. As in other energetics-based methods, protein-ligand interactions are proposed to stabilize specific conformations, thereby rendering the target proteins less susceptible to proteolysis. However, differently from previous methods, the partially digested and undigested proteins that remain in samples after the pulse proteolysis step are not analyzed. They are removed from samples by a trichloroacetic acid (TCA) precipitation step. Only the small peptides that are retained in solution after the precipitation step are analyzed. Treatment and control samples are analyzed by shotgun proteomics and compared using a label-free semi-quantitative approach.

Data Analysis Pipeline

Our data analysis pipeline is outlined in Figure 2A. Proteins can be identified in treatment, in control, or in both samples. Conceivably, the increase in conformational stability after ligand binding reduces the unfolded population for a given target protein. Therefore, we expect either a reduction of PSMs for a target in treatment samples or no matches at all in this condition if the unfolded population is below the detection limit of the experiment. Proteins with an increase of

PSMs in treatment samples or the ones identified exclusively in treatment samples are either false positives or cases in which ligand binding increases the proteolytic susceptibility of the target protein.¹⁶ In order to cope with the noise intrinsic in shotgun proteomics experiments and make PePTID a robust method for target identification, we performed three technical replicates for each of three independent biological replicates, totalizing 9 runs per sample condition. In each biological replicate, proteins should be detected in at least two technical replicates to be considered for further analysis, and only proteins detected as differential digested in all three biological replicates were considered a hit. We applied the TFold test in order to detect differences in the level of proteins identified in both samples.²⁰ If a protein satisfies cutoffs for both p value and Fold Change in the TFold analysis on all three biological replicates, it is considered as differentially digested (Figure 2B) (see Methods for details). The same criteria of reproducibility apply for proteins detected exclusively in one of the samples (see Figure 2). With these very stringent criteria, we expect to reduce the level of false positives, even at the expense of some targets not being detected. Depending on the particular target identification project and the number of positive IDs obtained, the selection criteria could be relaxed accordingly.

Identification of ATP-Binding Proteins in *Mycobacterium smegmatis* Using PePTID

To validate PePTID for target discovery, we looked for ATP-binding proteins in *Mycobacterium smegmatis* proteome using a non-hydrolysable ATP (ATP γ S) as a ligand. *M. smegmatis* lysates were incubated in 0, 1, 2, 3 or 4 M urea with and without ATP γ S, and each incubation was submitted to a pulse proteolysis reaction with thermolysin followed by a TCA protein precipitation step (see Methods for details).

One of the major limitations of energetics-based methods for target identification lies in the fact that we do not detect differences in stabilities after ligand binding for proteolysis-resistant proteins. Therefore, only the fraction of the proteome susceptible to proteolysis in a given set of conditions is effectively screened by these methods, which defines the proteome coverage. Different proteases were tested with the DARTS approach in order to improve proteome coverage, with limited success.¹⁶ The introduction of a denaturing agent such as urea to induce unfolding, followed by a pulse proteolysis, allowed the proteome coverage to be improved.¹³ In our hands, we found a larger number of proteins identified as differentially digested upon ATP γ S incubation using 1 M urea than in others urea concentrations (Table S1). Therefore, further experiments were performed in this urea concentration.

We identified 29 proteins in *M. smegmatis* that are stabilized and only 3 that are destabilized by the presence of ATP γ S (Table 1). We found in the literature direct biochemical evidence of ATP binding for only 2 out of the 29 stabilized proteins: Acyl-CoA synthase (MSMEG_5114 – A0R2H4)²¹ and a universal stress protein family member (MSMEG_3811 – A0QYW6).²² However, 25 out of the remaining 27 stabilized proteins display features that suggest ATP or at least nucleotide binding. Indeed, based on a Gene Ontology (GO) analysis using the UniProtKB database, 12 of the 29 proteins protected by ATP γ S binding are ATP-binding proteins (GO:0005524) (see Table 1). The gene products of MSMEG_3942, MSMEG_3743 and MSMEG_0759 have the P-loop motif, known to interact with ATP/GTP nucleotides.^{23, 24} Moreover, three Universal Stress Proteins (USPs) were identified, MSMEG_3811 (whose ATP-binding activity was experimentally confirmed²²), MSMEG_3945 and MSMEG_3950.

The USP superfamily is a conserved group of proteins found in bacteria, Archea, fungi and plants.²⁵ The USP expression is stimulated by several stress conditions and could be important for survival during growth arrest. USPs can be divided into two groups: those which bind ATP and those which do not. The putative ATP-binding USPs have the G2xG9xGS(V/T) motif²⁶ and the three USP proteins identified have this conserved motif (Figure 3).

IMPDH was totally protected by ATP γ S. Peptides were identified only in control sample and correspond to the CBS domains (figure 4), which was shown to bind ATP in the human enzyme, and also in IMPDHs from other microorganisms, with consequences on the global shape of the enzyme.^{30, 31} A second sequence related to IMPDH was also identified within the proteins stabilized by ligand incubation (MSMEG_1603 - A0QSU4).

Another identified protein (MSMEG_0131 - A0QNQ9) possesses an adenylating motif (PF00501) that is found in enzymes that act via an ATP-dependent covalent binding of AMP to the substrates.³² We also identified the subunit L of dihydroxyacetone kinase (MSMEG_2122 – A0QU92) which shows glycerone kinase activity by GO (GO:0004371). The subunit DhaL corresponds to a domain (PS51480) found in both PEP-dependent and ATP-dependent dihydroxyacetone kinases.³³ Orthologues of 6 proteins stabilized by ATP γ S treatment were found to bind related nucleotides, such as adenosine, NAD or FAD (see Table 1).

Three proteins were found to be destabilized by the presence of ATP γ S: alcohol dehydrogenase (MSMEG_2048 – A0QU20), 50S ribosomal protein L20 (MSMEG_3791 – A0QYU6), and DoxX (MSMEG_2370 – A0QUX4). One possibility is that direct ligand binding effectively promotes an increase in proteolytic susceptibility on these proteins. The increase in proteolytic susceptibility upon ligand binding was demonstrated for different protein-ligand interactions,

such as haemin binding to bovine serum albumin (BSA),³⁴ binding of the chromogenic reporter 4'-hydroxyazobenzene-2-benzoic acid (HABA) to Proteinase K,³⁵ and heparin binding to Pigment Epithelium-Derived Factor (PEDF).³⁶ However, in contrast to most proteins stabilized upon ATP γ S incubation, we did not find any structural feature suggesting ATP-binding activity for these 3 differentially-digested proteins. An interesting possibility is that their conformational stabilities are changed upon ATP γ S incubation by indirect mechanisms.

Finding the Niche of PePTID among Target Identification Strategies

Different approaches for target deconvolution were developed in recent years. The application of these approaches can provide complementary information and increase the success of target-identification efforts.³⁷ Affinity chromatography is the most used method for target discovery. The bioactive compound is immobilized on a matrix and target candidates are isolated based on affinity. However, the compound must be chemically modified or labeled in order to be attached on the matrix, which could compromise binding with targets. The target protein should also be abundant to be detected by this approach.³⁸ Functional genomics approaches, on the other hand, makes use of genetic tools to identify functionally relevant cellular components, not necessarily being the direct targets that physically interact and elicit the compound's activities.³⁷ Moreover, cellular profiling and knowledge-based approaches were also developed. In the former, a gene expression or metabolomic signature obtained from the treated sample is compared with a reference set of data using pattern-matching algorithms. In the latter, focused on the compound properties and not on the target functions, bioactivity data are collected and compared to compound-target associations available for other compounds using cheminformatics methods.

Energetics-based methods explore the well-known fact that protein folding and ligand binding are linked.³⁹ The conformational stability of a protein is affected upon ligand binding. Frequently, the formation of a ligand-binding interaction increases the thermodynamic stability and consequently slows down protein unfolding.^{13, 14} These methods were originally introduced to make thermodynamic measurements of protein folding and stability.¹² As much less protein is required to perform experiments, they served as surrogates of classical biophysical techniques such as optical spectroscopies, nuclear magnetic resonance, and calorimetry.¹² Another useful advantage of energetics-based methods is the possibility to use unpurified proteins, which prompted their use in target-identification efforts.

Energetics-based methods use chemical or enzymatic modifications to probe the equilibrium unfolding of proteins or protein-ligand complexes. They include methods such as Stability of Proteins from Rates of Oxidation (SPROX),⁴⁰ Stability of Unpurified Proteins from Rates of amide H/D Exchange (SUPREX),⁴¹ slow histidine H/D exchange,⁴² Drug Affinity Responsive Target Stability (DARTS),¹⁶ pulse proteolysis, Quantitative Cysteine Reactivity (QCR),^{43, 44} among others.¹² Except for DARTS, in all other methods the equilibrium distribution of folded and unfolded forms are experimentally controlled by chemical or thermal denaturation. The structural probes vary among methods and ultimately determine the experimental readout employed.¹² Techniques that probe structure using hydrogen/deuterium (H/D) exchange, methionine oxidation or lysine amidination make use of mass spectrometry and all tools available for robust analysis of MS-derived data. However, each of these methods have its own limitations in terms of the structural probes employed. SUPREX require high amounts of target protein to be detected in cell lysates, SPROX require the presence and detection of peptides containing

methionine from target proteins, and lysine amidination require the presence of lysine residues buried in protein structure, which is a rare event. In contrast, pulse proteolysis is a very cheap and straightforward method for getting a snapshot of conformational equilibria between folded and unfolded structures of target proteins. The main issue of DARTS, related to the large fraction of proteolytically-resistant proteins found in proteomes, was successfully avoided by the use of chemical or thermal-induced denaturation in pulse proteolysis protocols. Therefore, a larger fraction of the proteome can be structurally probed with these methods. However, the main limitation of current approaches based on pulse proteolysis lies on the analytical readouts: SDS-PAGE or 2D electrophoresis. In both cases, the identification of differentially-digested proteins depends on visual inspection. The poor reproducibility of 2D electrophoresis experiments were partially relieved by the use of SDS-PAGE, but that adaptation required the sample fractionation before ligand binding and pulse proteolysis,¹⁴ greatly increasing the number of gel runs and ligand required for the analysis. Summing up, currently available energetics-based methods relying on pulse proteolysis for target identification are not yet developed for high-throughput efforts. Precisely for that reason we developed PePTID. In this method, we achieved to bring together the advantages of pulse proteolysis as a structural probe for the folding/unfolding conformational distribution of target proteins or protein-ligand complexes with the robustness of a MS-based analytical readout. By focusing our analysis on the small peptides generated after the pulse-proteolysis step, and not on the intact or large protein fragments that resisted proteolysis, we reduced sample complexity. Moreover, the quantity of these small peptide fragments can be compared among different samples using label-free nanoLC-MS/MS proteomic experiments. The reduction of sample complexity and the improved analytical resolution achieved with this MS

approach allowed us to avoid the sample fractionation step, improving the throughput and reliability of the method.

ASSOCIATED CONTENT

Supporting Information. Global analysis of PePTID data for all urea concentrations (Table S1), analysis of PePTID data for biological replicates of 1M urea incubations (Table S2). This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

* E-mail: cristiano.bizarro@puers.br Tel: +55-51-33203629; Fax: +55-51-33203629.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by funds from the National Institute of Science and Technology on Tuberculosis (INCT-TB), Decit/SCTIE/MS-MCT-CNPq-FNDCT-CAPES (Brazil) to C.V. Bizarro and D. S. Santos. D. S. Santos (CNPq, 304051/1975-06) is Research Career Awardee of

the National Research Council of Brazil (CNPq). The fellowships from FAPERGS (R. V. Trindade) and CNPq (A. F. M. Pinto) are also acknowledged.

REFERENCES

1. Chanda, S. K.; Caldwell, J. S., Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discovery Today* **2003**, *8* (4), 168-174.
2. Lee, J. A.; Uhlik, M. T.; Moxham, C. M.; Tomandl, D.; Sall, D. J., Modern Phenotypic Drug Discovery Is a Viable, Neoclassic Pharma Strategy. *Journal of Medicinal Chemistry* **2012**, *55* (10), 4527-4538.
3. Sams-Dodd, F., Target-based drug discovery: is something wrong? *Drug Discovery Today* **2005**, *10* (2), 139-147.
4. Sams-Dodd, F., Is poor research the cause of the declining productivity of the pharmaceutical industry? An industry in need of a paradigm shift. *Drug Discovery Today* **2013**, *18* (5-6), 211-217.
5. Swinney, D. C.; Anthony, J., How were new medicines discovered? *Nature Reviews Drug Discovery* **2011**, *10* (7), 507-519.
6. Lee, J. A.; Berg, E. L., Neoclassic Drug Discovery: The Case for Lead Generation Using Phenotypic and Functional Approaches. *Journal of Biomolecular Screening* **2013**, *18* (10), 1143-1155.
7. Lewis, K., Platforms for antibiotic discovery. *Nature Reviews Drug Discovery* **2013**, *12* (5), 371-387.
8. Hart, C. P., Finding the target after screening the phenotype. *Drug Discovery Today* **2005**, *10* (7), 513-519.

9. Futamura, Y.; Muroi, M.; Osada, H., Target identification of small molecules based on chemical biology approaches (vol 9, pg 897, 2013). *Molecular Biosystems* **2013**, *9* (12), 3210-3210.
10. Lounkine, E.; Keiser, M. J.; Whitebread, S.; Mikhailov, D.; Hamon, J.; Jenkins, J. L.; Lavan, P.; Weber, E.; Doak, A. K.; Cote, S.; Shoichet, B. K.; Urban, L., Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **2012**, *486* (7403), 361-+.
11. Ziegler, S.; Pries, V.; Hedberg, C.; Waldmann, H., Target Identification for Small Bioactive Molecules: Finding the Needle in the Haystack. *Angewandte Chemie-International Edition* **2013**, *52* (10), 2744-2792.
12. Geer, M. A.; Fitzgerald, M. C., Energetics-Based Methods for Protein Folding and Stability Measurements. *Annual Review of Analytical Chemistry, Vol 7* **2014**, *7*, 209-228.
13. Liu, P.-F.; Kihara, D.; Park, C., Energetics-Based Discovery of Protein-Ligand Interactions on a Proteomic Scale. *Journal of Molecular Biology* **2011**, *408* (1), 147-162.
14. Chang, Y.; Schleich, J. P.; VerHeul, R. A.; Park, C., Simplified proteomics approach to discover protein-ligand interactions. *Protein Science* **2012**, *21* (9), 1280-1287.
15. Park, C. W.; Marqusee, S., Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. *Nature Methods* **2005**, *2* (3), 207-212.
16. Lomenick, B.; Hao, R.; Jonai, N.; Chin, R. M.; Aghajan, M.; Warburton, S.; Wang, J.; Wu, R. P.; Gomez, F.; Loo, J. A.; Wohlschlegel, J. A.; Vondriska, T. M.; Pelletier, J.; Herschman, H. R.; Clardy, J.; Clarke, C. F.; Huang, J., Target identification using drug

- affinity responsive target stability (DARTS). *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106* (51), 21984-21989.
17. Klammer, A. A.; MacCoss, M. J., Effects of modified digestion schemes on the identification of proteins from complex mixtures. *Journal of Proteome Research* **2006**, *5* (3), 695-700.
18. Eng, J. K.; Jahan, T. A.; Hoopmann, M. R., Comet: An open-source MS/MS sequence database search tool. *Proteomics* **2013**, *13* (1), 22-24.
19. Carvalho, P. C.; Lima, D. B.; Leprevost, F. V.; Santos, M. D. M.; Fischer, J. S. G.; Aquino, P. F.; Moresco, J. J.; Yates, J. R., III; Barbosa, V. C., Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0. *Nature Protocols* **2016**, *11* (1), 102-117.
20. Carvalho, P. C.; Yates, J. R., III; Barbosa, V. C., Improving the TFold test for differential shotgun proteomics. *Bioinformatics* **2012**, *28* (12), 1652-1654.
21. Galandrin, S.; Guillet, V.; Rane, R. S.; Leger, M.; Radha, N.; Eynard, N.; Das, K.; Balganes, T. S.; Mourey, L.; Daffe, M.; Marrakchi, H., Assay Development for Identifying Inhibitors of the Mycobacterial FadD32 Activity. *Journal of Biomolecular Screening* **2013**, *18* (5), 576-587.
22. Banerjee, A.; Adolph, R. S.; Gopalakrishnapai, J.; Kleinboelting, S.; Emmerich, C.; Steegborn, C.; Visweswariah, S. S., A Universal Stress Protein (USP) in Mycobacteria Binds cAMP. *Journal of Biological Chemistry* **2015**, *290* (20), 12731-12743.

23. Saraste, M.; Sibbald, P. R.; Wittinghofer, A., The P-Loop - A Common Motif In ATP-Binding And GTP-Binding Proteins. *Trends in Biochemical Sciences* **1990**, *15* (11), 430-434.
24. Chen, K.; Wang, D. C.; Kurgan, L., Systematic investigation of sequence and structural motifs that recognize ATP. *Computational Biology and Chemistry* **2015**, *56*, 131-141.
25. Hingley-Wilson, S. M.; Loughheed, K. E. A.; Ferguson, K.; Leiva, S.; Williams, H. D., Individual Mycobacterium tuberculosis universal stress protein homologues are dispensable in vitro. *Tuberculosis* **2010**, *90* (4), 236-244.
26. Sousa, M. C.; McKay, D. B., Structure of the universal stress protein of Haemophilus influenzae. *Structure* **2001**, *9* (12), 1135-1141.
27. Szabo, A.; Langer, T.; Schroder, H.; Flanagan, J.; Bukau, B.; Hartl, F. U., The ATP hydrolysis-dependent reaction cycle of the Escherichia-coli hsp70 system - DnaK, DnaJ, and GrpE. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91* (22), 10345-10349.
28. Breda, A.; Martinelli, L. K. B.; Bizarro, C. V.; Rosado, L. A.; Borges, C. B.; Santos, D. S.; Basso, L. A., Wild-Type Phosphoribosylpyrophosphate Synthase (PRS) from Mycobacterium tuberculosis: A Bacterial Class II PRS? *Plos One* **2012**, *7* (6).
29. Meena, L. S.; Chopra, P.; Bedwal, R. S.; Singh, Y., Nucleoside diphosphate kinase-like activity in adenylate kinase of Mycobacterium tuberculosis. *Biotechnology and Applied Biochemistry* **2003**, *38*, 169-174.

30. Labesse, G.; Alexandre, T.; Vaupre, L.; Salard-Arnaud, I.; Him, J. L. K.; Raynal, B.; Bron, P.; Munier-Lehmann, H., MgATP Regulates Allostery and Fiber Formation in IMPDHs. *Structure* **2013**, *21* (6), 975-985.
31. Alexandre, T.; Rayna, B.; Munier-Lehmann, H., Two Classes of Bacterial IMPDHs according to Their Quaternary Structures and Catalytic Properties. *Plos One* **2015**, *10* (2).
32. Conti, E.; Franks, N. P.; Brick, P., Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure* **1996**, *4* (3), 287-298.
33. Erni, B.; Siebold, C.; Christen, S.; Srinivas, A.; Oberholzer, A.; Baumann, U., Small substrate, big surprise: fold, function and phylogeny of dihydroxyacetone kinases. *Cellular and Molecular Life Sciences* **2006**, *63* (7-8), 890-900.
34. Shin, W. S.; Yamashita, H.; Hirose, M., Multiple effects of haemin binding on protease susceptibility of bovine serum-albumin and a novel isolation procedure for its large fragment. *Biochemical Journal* **1994**, *304*, 81-86.
35. Ellison, D.; Hinton, J.; Hubbard, S. J.; Beynon, R. J., Limited proteolysis of native proteins - the interaction between avidin and proteinase K. *Protein Science* **1995**, *4* (7), 1337-1345.
36. Valnickova, Z.; Petersen, S. V.; Nielsen, S. B.; Otzen, D. E.; Enghild, J. J., Heparin binding induces a conformational change in pigment epithelium-derived factor. *Journal of Biological Chemistry* **2007**, *282* (9), 6661-6667.

37. Schirle, M.; Jenkins, J. L., Identifying compound efficacy targets in phenotypic drug discovery. *Drug Discovery Today* **2016**, *21* (1), 82-89.
38. Titov, D. V.; Liu, J. O., Identification and validation of protein targets of bioactive small molecules. *Bioorganic & Medicinal Chemistry* **2012**, *20* (6), 1902-1909.
39. Wyman, J.; Gill, S. J. *Binding and Linkage: Functional Chemistry of Biological Macromolecules*; University Science Books: Mill Valley, CA, 1990.
40. West, G. M.; Tang, L.; Fitzgerald, M. C., Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass-spectrometry based strategy. *Analytical Chemistry* **2008**, *80* (11), 4175-4185.
41. Ghaemmaghami, S.; Fitzgerald, M. C.; Oas, T. G., A quantitative, high-throughput screen for protein stability. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97* (15), 8296-8301.
42. Tran, D. T.; Banerjee, S.; Alayash, A. I.; Crumbliss, A. L.; Fitzgerald, M. C., Slow Histidine H/D Exchange Protocol for Thermodynamic Analysis of Protein Folding and Stability Using Mass Spectrometry. *Analytical Chemistry* **2012**, *84* (3), 1653-1660.
43. Isom, D. G.; Marguet, P. R.; Oas, T. G.; Hellinga, H. W., A miniaturized technique for assessing protein thermodynamics and function using fast determination of quantitative cysteine reactivity. *Proteins-Structure Function and Bioinformatics* **2011**, *79* (4), 1034-1047.

44. Isom, D. G.; Vardy, E.; Oas, T. G.; Hellinga, H. W., Picomole-scale characterization of protein stability and function by quantitative cysteine reactivity. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107* (11), 4908-4913.

List of Figures

Figure 1. PePTID approach for target identification. Cell lysate containing target proteins (blue) and many non-binders (gray) is separated in two samples, the (+) Ligand (treatment) and (-) Ligand (control). Both samples are maintained in the same buffer conditions. After incubation of (+) Ligand sample with the bioactive compound (red diamonds), both samples are subjected to a pulse proteolysis step. Target proteins are less susceptible to proteolysis in the (+) Ligand sample than in the (-) Ligand sample (see main text for details), resulting in more peptide fragments from target proteins in the latter (blue fragments). After a trichloroacetic acid (TCA) precipitation step, peptides remaining on solution are analyzed by shotgun proteomics and compared using a label-free semi-quantitative approach.

Figure 2. (A) PePTID data analysis pipeline. (B) TFold volcano plot: blue dots represent proteins which satisfied both fold change and p -value cutoffs; Orange dots represent proteins filtered out by L-stringency criterium; Green dots represent proteins which satisfied fold change criterium but the statistical analysis indicates this happened by chance; Red dots represent proteins which did not meet fold and p -value criteria.

Figure 3. Alignment of identified USPs sequence segments containing the ATP binding motif (G2xG9xGS/T).

Figure 4. IMPDH sequence. The sequences shaded in red and green correspond to CBS domains 1 and 2, respectively. The peptides identified in control sample (Blue) belong to the conserved CBS domains.

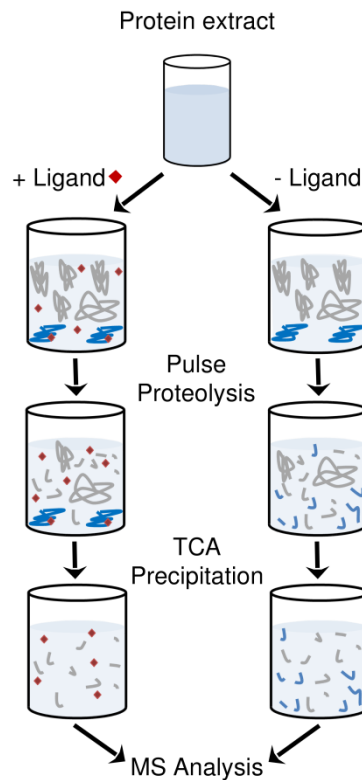


Figure 1. PePTID approach for target identification. Cell lysate containing target proteins (blue) and many non-binders (gray) is separated in two samples, the (+) Ligand (treatment) and (-) Ligand (control). Both samples are maintained in the same buffer conditions. After incubation of (+) Ligand sample with the bioactive compound (red diamonds), both samples are subjected to a pulse proteolysis step. Target proteins are less susceptible to proteolysis in the (+) Ligand sample than in the (-) Ligand sample (see main text for details), resulting in more peptide fragments from target proteins in the latter (blue fragments). After a trichloroacetic acid (TCA) precipitation step, peptides remaining on solution are analyzed by shotgun proteomics and compared using a label-free semi-quantitative approach.

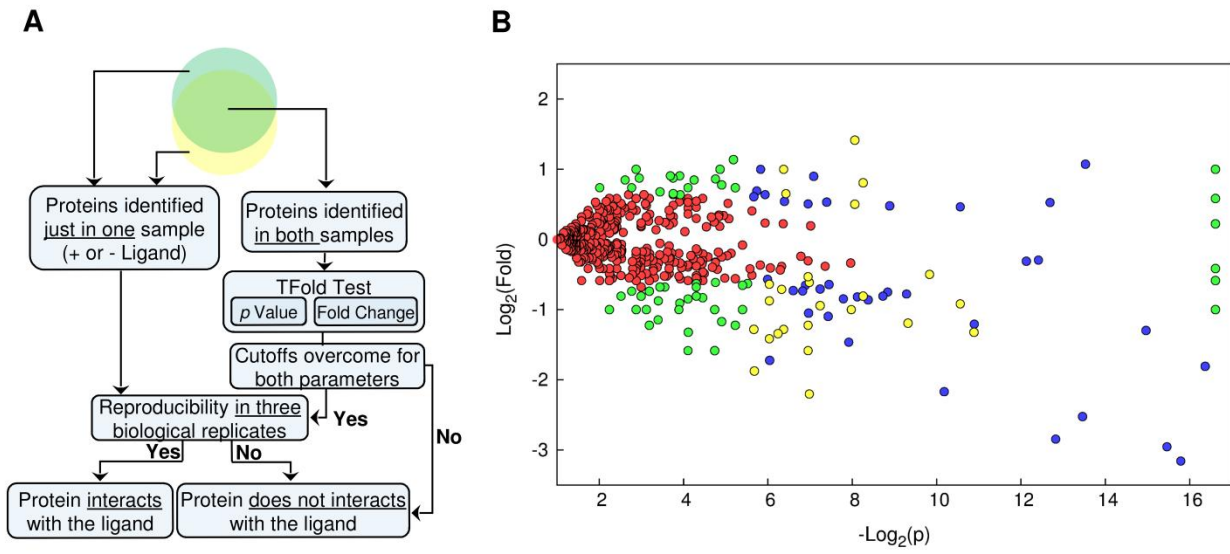


Figure 2. (A) PePTID data analysis pipeline. (B) TFold volcano plot: blue dots represent proteins which satisfied both fold change and p -value cutoffs; Orange dots represent proteins filtered out by L-stringency criterium; Green dots represent proteins which satisfied fold change criterium but the statistical analysis indicates this happened by chance; Red dots represent proteins which did not meet fold and p -value criteria.

```

sp | MSMEG_3950   1  MVQS--ATEYGILVGVDS SAESDAVRWAAREASLHDAPITLMHVIA-----PVVVS
tr | MSMEG_3811   1  ----MSAYQTVVGT DGS DSSLRAVDRAGQIAA-----A
tr | MSMEG_3945   1  -----MEWAAKEAVLHKVPLRIVHVMQW-----PPGSAT

sp | MSMEG_3950  51  WPAGPYMATVLECEENARHAEQAOKVVADCLGETHGLTVQTEIRKESVARTLIDASKS
tr | MSMEG_3811  31  SNAKLIITATAYFPQSE-----DSRAADV LKDEGVK MAGNAPIYAILRE
tr | MSMEG_3945  30  WSEIPAPTTLADQLKRNAEQLLKEARKVAEDVVG GAVEIHVDV VVAVG SVVAALTELSSE

sp | MSMEG_3950 111  AQMVVVGNRGMGALGRVLLGSTSTSLHYASGPVVV VHGDDQA-A--HDSRLPVLLGIDG
tr | MSMEG_3811  74  ANDRAKAAGAT-----DIEE
tr | MSMEG_3945  90  SRLLVVGCRGLGPIGRRL LGSVTAGLRHARCPVAVIHGESKASA--IATTAPVVVGVVDG

sp | MSMEG_3950 168  SPASEVATSHAFDEASRRG-----VDLVALHVWIDVGDIPPIGPTWEEQEETGRALLAER
tr | MSMEG_3811  89  RPVVGAPVDALVELADEV-----
tr | MSMEG_3945 148  SPASEKALEFAFDASRRG-----APLLAVHAWSDANVAGYPCVVRWDDLRLVQAEESLSER

sp | MSMEG_3950 223  LAGWQERYPDVKVHRRVERAQPAYWLL EEAKQAQLV VVGSHGRGGFT CMLLGSVSSRVAQ
tr | MSMEG_3811 107  -----KADLVVGVNGLSTHAGRL LGSVPANVAR
tr | MSMEG_3945 203  LAGWQERYPDVHVRREVVLDRPDRRL KRSEIAQLTVV VVGSHGRGGFACLLGSVSTSVAE

                                                                    G2xG9xGS/T

sp | MSMEG_3950 283  SATTPVMVVRPR
tr | MSMEG_3811 136  RSKTDVLI VHTS
tr | MSMEG_3945 263  AERFPVVVVRG-

```

Figure 3. Alignment of identified USPs sequence segments containing the ATP binding motif (G2xG9xGS/T).

```

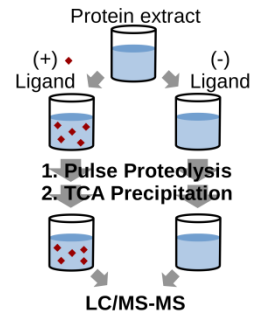
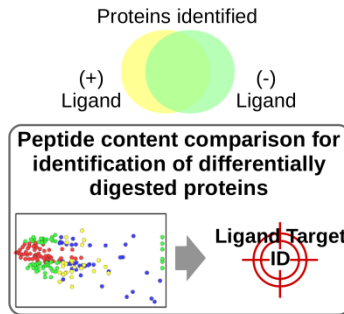
001 MSIAESSVPI AVPVPTGGDD PTKVAMLGLT FDDVLLLPAA SDVVPATADT SSQLTKRIRL 060
061 RVPLVSSAMD TVTESRMAIA MARAGGMGVL HRNLPVAEQA GQVETVKRSE AGMVTDPVTC 120
121 SPDNTLAEVD AMCARFRISG LPVVDDTGEL VGIITNRDMR FEVDQSKPVS EVMTKAPLIT 180
181 AKEGVSAEAA LGLLRHKIE KLPIVDGHGK LTGLITVKDF VKTEQFPLST KDSDGRLLVG 240
241 AAVGVGDDAW TRAMTLVDAG VDVLIVDTAH AHNRGVLDLV SRLKQAVGER VDVVGGNVAT 300
301 RAAAAALVEA GADAVKVGVG PGSICTTRVV AGVGAPQITA ILEAVAACKP YGVPVIADGG 360
361 LQYSGDIAKA LAAGASTAML GSLLAGTAES PGELIFVNGK QFKSYRGMGS LGAMQGRGAA 420
421 KSYSKDRYFQ DDVLSKDLV PEGIEGRVPP RGPLGTVIHQ LTGGLRAAMG YTGSATIEQL 480
481 QQAQFVQITA AGLKESHPhD ITMTVEAPNY YTR

```

Figure 4. IMPDH sequence. The sequences shaded in red and green correspond to CBS domains 1 and 2, respectively. The peptides identified in control sample (Blue) belong to the conserved CBS domains.

Table 1. ATP-binding proteins identified in *Mycobacterium smegmatis* using PePTID

UniprotKB	Gene name	Protein	Ligands	Evidence	Description	Ligand effect
A0QQC8	MSMEG_0709	Chaperone protein DnaK (dnaK)	ATP	Literature and GO	Orthologous evidence ²⁷ GO:0005524	Stabilization
A0R3C8	MSMEG_5427	Ribose-phosphate pyrophosphokinase (prs)	ATP and AMP	Literature and GO	Orthologous evidence ²⁸ GO:0005524	Stabilization
A0QYW6	MSMEG_3811	Universal stress protein	ATP and cAMP	Literature	Direct biochemical evidence ²²	Stabilization
A0R2H4	MSMEG_5114	Acyl-CoA synthase (fadD36)	ATP, ADP and AMP	Literature	Direct biochemical evidence ²¹	Stabilization
A0QSH8	MSMEG_1484	Adenylate kinase (adk)	ATP, ADP and AMP	Literature and GO	Orthologous evidence ²⁹ GO:0005524	Stabilization
A0QSU3	MSMEG_1602	Inosine-5'-monophosphate dehydrogenase (guaB)	IMP and NAD	Literature	Orthologous evidence ^{30, 31}	Stabilization
A0QZA1	MSMEG_3950	Universal stress protein MSMEG_3950/MSMEI_3859	-	Literature	Orthologous evidence ²⁵	Stabilization
A0QZ96	MSMEG_3945	Universal stress protein family protein	-	Literature	Orthologous evidence ²⁵	Stabilization
A0QYQ4	MSMEG_3743	SpoOJ regulator protein (soj)	-	InterPro	P-loop (InterPro: A0QYQ4)	Stabilization
A0QZ93	MSMEG_3942	Uncharacterized protein	-	InterPro	P-loop (InterPro: A0QZ93)	Stabilization
A0QQH7	MSMEG_0759	Adenylosuccinate synthetase (purA)	GTP, IMP and Adenylosuccinate	InterPro	P-Loop (InterPro: A0QQH7)	Stabilization
P0CG99	MSMEG_1019	Ribonucleoside-diphosphate reductase subunit alpha 1 (nrde1)	ATP	GO	GO:0005524	Stabilization
A0QNY8	MSMEG_0211	ABC transporter, ATP-binding protein	ATP and ADP	GO	GO:0005524	Stabilization
A0QX17	MSMEG_3140	ABC transporter ATP-binding protein, possibly in EF-3 subfamily protein	ATP and ADP	GO	GO:0005524	Stabilization
A0R1C3	MSMEG_4700	ABC-transporter protein, ATP binding component	ATP and ADP	GO	GO:0005524	Stabilization
A0R3M4	MSMEG_5525	Succinyl-CoA ligase [ADP-forming] subunit beta (sucC)	ATP and ADP	GO	GO:0005524	Stabilization
A0R3M3	MSMEG_5524	Succinyl-CoA ligase [ADP-forming] subunit alpha (sucD)	ATP and ADP	GO	GO:0005524	Stabilization
A0QU92	MSMEG_2122	Dihydroxyacetone kinase, L subunit	ATP and ADP	GO	GO:0004371	Stabilization
A0R638	MSMEG_6413	Serine--tRNA ligase (serS)	ATP and AMP	GO	GO:0005524	Stabilization
A0QUY7	MSMEG_2383	Glutamate--tRNA ligase (gltX)	ATP and AMP	GO	GO:0005524	Stabilization
A0R577	MSMEG_6094	Lysine--tRNA ligase (lysS)	ATP and AMP	GO	GO:0005524	Stabilization
A0QT14	MSMEG_1676	Adenosine deaminase (add)	Adenosine and inosine	-	-	Stabilization
A0QNQ9	MSMEG_0131	AMP-binding enzyme, putative	AMP	-	-	Stabilization
A0R6K8	MSMEG_6584	Acyl-CoA dehydrogenase	FAD	-	-	Stabilization
A0R2Q5	MSMEG_5197	Acyl-CoA dehydrogenase	FAD	-	-	Stabilization
A0R6K9	MSMEG_6585	Acyl-CoA dehydrogenase	FAD	-	-	Stabilization
A0QZR1	MSMEG_4117	2-deoxy-D-gluconate 3-dehydrogenase	NAD	-	-	Stabilization
A0QSU4	MSMEG_1603	Uncharacterized oxidoreductase MSMEG_1603/MSMEI_1564	NAD	-	-	Stabilization
A0QZ86	MSMEG_3935	Uncharacterized protein	-	-	-	Stabilization
A0QU20	MSMEG_2048	Alcohol dehydrogenase	NAD	-	-	Destabilization
A0QYU6	MSMEG_3791	50S ribosomal protein L20 (rplI)	-	-	-	Destabilization
A0QUX4	MSMEG_2370	DoxX	-	-	-	Destabilization

I. Experimental Procedure**II. Data Analysis**

For Table of Content only

6 CONSIDERAÇÕES FINAIS

Os resultados obtidos com a aplicação do método PePTID descrito nesse trabalho atestam a fiabilidade do procedimento para a identificação de alvos moleculares de compostos. A estratégia adotada resultou em um protocolo de fácil execução e custo efetivo, com potencial para identificação tanto de proteínas nas quais a interação com o ligante desempenha a atividade desejada do composto (alvo) como também de proteínas nas quais a interação exerce efeitos indesejados (interações inespecíficas com proteínas). A estratégia não somente pode ser usada para identificar os alvos de compostos inibidores, como também pode ser uma alternativa para a identificação de interações enzima-substrato e de interações em sítios regulatórios não catalíticos.

Embora o método de análise não seja trivial, ele se mostrou eficiente e mais favorável que as análises por SDS-PAGE. A análise por SDS-PAGE, mesmo sendo um procedimento rotineiro, é bastante desfavorável na plataforma experimental proposta por Chang e colaboradores quando comparada a estratégia descrita nesse trabalho. A análise por SDS-PAGE exige o fracionamento do extrato proteico e cada fração precisa ser submetida ao protocolo de proteólise pulsada em diversas concentrações do agente desnaturante para então serem analisadas no gel, de modo que são necessárias muitas análises para a obtenção de dados satisfatórios. Além disso, as comparações das incubações tratadas com o ligante e controle são visuais e as bandas que apresentam prováveis diferenças necessitam ser tratadas e analisadas individualmente por espectrometria de massa, tornando o procedimento ainda mais laborioso. A análise dos peptídeos resultantes da degradação proteolítica possui maior poder de resolução, aumentando significativamente a cobertura do proteoma, e exige apenas a execução do protocolo de proteólise pulsada e a análise por espectrometria de massa.

O método desenvolvido possui grande potencial para aplicação em plataformas de desenvolvimento de fármacos. Como perspectivas futuras para o trabalho está a aplicação do método para a identificação dos alvos moleculares de compostos-líder com atividade antimicrobacteriana provenientes de plataformas de triagem fenotípica de células.

REFERÊNCIAS

- ANDRIES, K. *et al.* A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. **Science**, v. 307, n. 5707, p. 223-227, Jan 14 2005.
- BALLELL, L. *et al.* Fueling Open-Source Drug Discovery: 177 Small-Molecule Leads against Tuberculosis. **Chemmedchem**, v. 8, n. 2, p. 313-321, Feb 2013.
- BUTERA, J. A. Phenotypic Screening as a Strategic Component of Drug Discovery Programs Targeting Novel Antiparasitic and Antimycobacterial Agents: An Editorial Miniperspectives Series on Phenotypic Screening for Antiinfective Targets. **Journal of Medicinal Chemistry**, v. 56, n. 20, p. 7715-7718, Oct 24 2013.
- CALLIGARO, G. L. *et al.* The medical and surgical treatment of drug-resistant tuberculosis. **Journal of Thoracic Disease**, v. 6, n. 3, p. 186-195, Mar 2014.
- CHANG, Y. *et al.* Simplified proteomics approach to discover protein-ligand interactions. **Protein Science**, v. 21, n. 9, p. 1280-1287, Sep 2012.
- COOPER, C. B. Development of *Mycobacterium tuberculosis* Whole Cell Screening Hits as Potential Antituberculosis Agents. **Journal of Medicinal Chemistry**, v. 56, n. 20, p. 7755-7760, Oct 24 2013.
- CORBETT, E. L. *et al.* The growing burden of tuberculosis - Global trends and interactions with the HIV epidemic. **Archives of Internal Medicine**, v. 163, n. 9, p. 1009-1021, May 12 2003.
- DARTOIS, V. The path of anti-tuberculosis drugs: from blood to lesions to mycobacterial cells. **Nature Reviews Microbiology**, v. 12, n. 3, p. 159-167, Mar 2014.
- DUCATI, R. G. *et al.* The resumption of consumption - A review on tuberculosis. **Memorias Do Instituto Oswaldo Cruz**, v. 101, n. 7, p. 697-714, Nov 15 2006.

FREUNDLICH, J. S. *et al.* Triclosan Derivatives: Towards Potent Inhibitors of Drug-Sensitive and Drug-Resistant Mycobacterium tuberculosis. **Chemmedchem**, v. 4, n. 2, p. 241-248, Feb 2009.

FUTAMURA, Y.; MUROI, M.; OSADA, H. Target identification of small molecules based on chemical biology approaches (vol 9, pg 897, 2013). **Molecular Biosystems**, v. 9, n. 12, p. 3210-3210, 2013.

GILBERT, I. H. Drug Discovery for Neglected Diseases: Molecular Target-Based and Phenotypic Approaches. **Journal of Medicinal Chemistry**, v. 56, n. 20, p. 7719-7726, Oct 24 2013.

HUGHES, J. P. *et al.* Principles of early drug discovery. **British Journal of Pharmacology**, v. 162, n. 6, p. 1239-1249, Mar 2011.

KONDREDDI, R. R. *et al.* Design, Synthesis, and Biological Evaluation of Indole-2-carboxamides: A Promising Class of Antituberculosis Agents. **Journal of Medicinal Chemistry**, v. 56, n. 21, p. 8849-8859, Nov 2013.

KOUL, A. *et al.* The challenge of new drug discovery for tuberculosis. **Nature**, v. 469, n. 7331, p. 483-490, Jan 2011.

LAWN, S. D.; ZUMLA, A. I. Tuberculosis. **The Lancet**, v. 378, p. 57-72, Jul 2011.

LEE, J. A. *et al.* Modern Phenotypic Drug Discovery Is a Viable, Neoclassic Pharma Strategy. **Journal of Medicinal Chemistry**, v. 55, n. 10, p. 4527-4538, May 24 2012.

LIU, P.-F.; KIHARA, D.; PARK, C. Energetics-Based Discovery of Protein-Ligand Interactions on a Proteomic Scale. **Journal of Molecular Biology**, v. 408, n. 1, p. 147-162, Apr 22 2011.

LOMENICK, B. *et al.* Target identification using drug affinity responsive target stability (DARTS). **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 51, p. 21984-21989, Dec 22 2009.

LOMENICK, B.; OLSEN, R. W.; HUANG, J. Identification of Direct Protein Targets of Small Molecules. **Acs Chemical Biology**, v. 6, n. 1, p. 34-46, Jan 2011.

MARRAKCHI, H.; LANEELLE, M.-A.; DAFPE, M. Mycolic Acids: Structures, Biosynthesis, and Beyond. **Chemistry & Biology**, v. 21, n. 1, p. 67-85, Jan 16 2014.

MARTINEZ-JIMENEZ, F. *et al.* Target Prediction for an Open Access Set of Compounds Active against Mycobacterium tuberculosis. **Plos Computational Biology**, v. 9, n. 10, Oct 2013.

PARK, C. W.; MARQUSEE, S. Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. **Nature Methods**, v. 2, n. 3, p. 207-212, Mar 2005.

PAYNE, D. J. *et al.* Drugs for bad bugs: confronting the challenges of antibacterial discovery. **Nature Reviews Drug Discovery**, v. 6, n. 1, p. 29-40, Jan 2007.

REMUINAN, M. J. *et al.* Tetrahydropyrazolo 1,5-a Pyrimidine-3-Carboxamide and N-Benzyl-6',7'-Dihydrospiro Piperidine-4,4'-Thieno 3,2-c Pyran Analogues with Bactericidal Efficacy against Mycobacterium tuberculosis Targeting MmpL3. **Plos One**, v. 8, n. 4, Apr 2013.

REZWAN, M. *et al.* Breaking down the wall: Fractionation of mycobacteria. **Journal of Microbiological Methods**, v. 68, n. 1, p. 32-39, Jan 2007.

RUSSELL, D. G.; BARRY, C. E., III; FLYNN, J. L. Tuberculosis: What We Don't Know Can, and Does, Hurt Us. **Science**, v. 328, n. 5980, p. 852-856, May 14 2010.

SHEVCHENKO, A. *et al.* In-gel digestion for mass spectrometric characterization of proteins and proteomes. **Nature Protocols**, v. 1, n. 6, p. 2856-2860, 2006

TAKAYAMA, K.; WANG, C.; BESRA, G. S. Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis. **Clinical Microbiology Reviews**, v. 18, n. 1, p. 81-+, Jan 2005.

UDWADIA, Z.; VENDOTI, D. Totally drug-resistant tuberculosis (TDR-TB) in India: every dark cloud has a silver lining. **Journal of Epidemiology and Community Health**, v. 67, n. 6, p. 471-472, Jun 2013.

WHO – World Health Organization, **Global tuberculosis report 2015**, 2015.

YOKOKAWA, F. *et al.* Discovery of Tetrahydropyrazolopyrimidine Carboxamide Derivatives As Potent and Orally Active Antitubercular Agents. **Acs Medicinal Chemistry Letters**, v. 4, n. 5, p. 451-455, May 2013.

ZUMLA, A. I. *et al.* New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. **Lancet Infectious Diseases**, v. 14, n. 4, p. 327-340, Apr 2014.

ZUMLA, A.; NAHID, P.; COLE, S. T. Advances in the development of new tuberculosis drugs and treatment regimens. **Nature Reviews Drug Discovery**, v. 12, n. 5, p. 388-404, May 2013.

CANTALOUBE, S. *et al.* The Mycobacterium Tuberculosis FAS-II Dehydratases and Methyltransferases Define the Specificity of the Mycolic Acid Elongation Complexes. **Plos One**, v. 6, n. 12, Dec 22 2011.

BASSO, L. A. *et al.* Mechanisms of isoniazid resistance in Mycobacterium tuberculosis: Enzymatic characterization of enoyl reductase mutants identified in isoniazid-resistant clinical isolates. **Journal of Infectious Diseases**, v. 178, n. 3, p. 769-775, Sep 1998.

QUEMARD, A. *et al.* Enzymatic characterization of the target for isoniazid in mycobacterium-tuberculosis. **Biochemistry**, v. 34, n. 26, p. 8235-8241, Jul 4 1995.

ENG, J. K.; JAHAN, T. A.; HOOPMANN, M. R. Comet: An open-source MS/MS sequence database search tool. **Proteomics**, v. 13, n. 1, p. 22-24, Jan 2013.

CARVALHO, P. C. *et al.* Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0. **Nature Protocols**, v. 11, n. 1, p. 102-117, Jan 2016.

ADACHI, J. *et al.* Proteome-Wide Discovery of Unknown ATP-Binding Proteins and Kinase Inhibitor Target Proteins Using an ATP Probe. **Journal of Proteome Research**, v. 13, n. 12, p. 5461-5470, Dec 2014.

BANSAL-MUTALIK, R.; NIKAIDO, H. Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 13, p. 4958-4963, Apr 1 2014.

PISSINATE, K. *et al.* 2-(Quinolin-4-yloxy)acetamides Are Active against Drug-Susceptible and Drug-Resistant Mycobacterium tuberculosis Strains. **Acs Medicinal Chemistry Letters**, ASAP, Jan 11 2016. DOI:10.1021/acsmchemlett.5b00324

TITOV, D. V.; LIU, J. O. Identification and validation of protein targets of bioactive small molecules. **Bioorganic & Medicinal Chemistry**, v. 20, n. 6, p. 1902-1909, Mar 15 2012.

SCHIRLE, M.; JENKINS, J. L. Identifying compound efficacy targets in phenotypic drug discovery. **Drug Discovery Today**, v. 21, n. 6, p. 82-89, Jan 2016.