



Universidade de Brasília
Instituto de Ciências Biológicas
Departamento de Biologia Celular
Pós-Graduação em Biologia

**Integração de proteômica *bottom-up* e
top-down para caracterização de
proteoformas em microrganismos com
relevância médica e biotecnológica**

REYNALDO MAGALHÃES MELO

Orientador: Prof. Dr. Carlos André Ornelas Ricart

Co-orientador: Prof. Dr. Luis Henrique Ferreira do Vale

Brasília, outubro de 2023



Universidade de Brasília
Instituto de Ciências Biológicas
Departamento de Biologia Celular
Pós-Graduação em Biologia

Integração de proteômica *bottom-up* e *top-down* para caracterização de proteoformas em microrganismos com relevância médica e biotecnológica

REYNALDO MAGALHÃES MELO

Tese de doutorado apresentada ao Programa de Pós-Graduação em Biologia Molecular, associado ao Instituto de Ciências Biológicas da Universidade de Brasília, como requisito para a obtenção do título de Doutor em Biologia Molecular, na área de concentração Bioquímica. Orientador: Prof. Dr. Carlos André Ornelas Ricart.

Brasília, outubro de 2023

Banca examinadora

Presidente: Prof. Dr. Carlos André Ornelas Ricart

Universidade de Brasília, Departamento de Biologia Celular

Prof. Dr. Sébastien Olivier Charneau

Universidade de Brasília, Departamento de Biologia Celular

Dra. Lyris Martins Franco de Godoy

Instituto Carlos Chagas, Fundação Oswaldo Cruz, Curitiba-PR

Prof. Dr. Magno Rodrigues Junqueira

Universidade Federal do Rio de Janeiro, Instituto de Química

**Dr. Agenor de Castro Moreira dos Santos Júnior
(suplente)**

Hospital da Criança de Brasília José de Alencar, Laboratório de
Pesquisa Translacional

Local: Híbrida, Microsoft Teams/Instituto de Ciências Biológicas, auditório 3.

Data: 21/03/2023

Horário: 09:00

Dedico esse trabalho a todos os cientistas brasileiros, àqueles que conseguem seguir esse caminho árduo e também aos que infelizmente não continuam ou começam por falta de oportunidades. Em especial, dedico o capítulo III dessa tese ao Agnelo Rodrigues de Souza, quem iniciou e participou dos trabalhos descritos nesse capítulo, mas que infelizmente faleceu antes de sua conclusão.

Agradecimentos

Agradeço imensamente ao meu orientador, Prof. Carlos André Ornelas Ricart, por ter compreensão nos momentos difíceis e incentivar o desenvolvimento da pesquisa nos momentos necessários. Também agradeço ao meu co-orientador, Prof. Luis Henrique Ferreira do Vale, por instigar o meu desenvolvimento relacionado à proteômica *top-down*.

Agradeço também a todos professores do laboratório de bioquímica e química de proteínas (LBQP), em especial ao Prof. Marcelo Valle de Sousa, coordenador do projeto PRONEX da FAP-DF responsável por financiar grande parte dessa pesquisa, ao Prof. Wagner Fontes, por estimular a discussão e instigar meu conhecimento sobre espectrometria de massas, e ao Prof. Sébastien Olivier Charneau, por ter me convidado para participar de publicações junto ao seu grupo. Agradeço também à um ex-membro do LBQP, Samuel Coelho Mandacaru, quem confiou e me estimulou a aprender sobre análise de dados.

Agradeço os técnicos do LBQP, Jaques Miranda Ferreira de Souza e Nuno Manoel Rodrigues, pela ajuda durante as análises de espectrometria de massas e análise de aminoácidos.

Agradeço a todos os meus colegas do LBQP, por sempre manterem um ótimo ambiente de trabalho, além de possibilitarem a discussão de diversos temas tratados nessa tese. Em especial, gostaria de agradecer à minha colega e amiga Anna Fernanda Vasconcellos, por ter me convidado a participar de dois artigos, juntamente com o Prof. Sébastien Olivier Charneau.

Outro membro do LBQP a quem eu agradeço imensamente é o Agnelo Rodrigues de Souza, com quem eu pude aprender muito sobre questões humanas, sociais e laboratoriais. O Agnelo teve influência direta na estrutura e desenvolvimento dessa tese, sendo assim, parte dessa é dedicada a ele, assim como foi apresentado na sessão de dedicatória.

Ainda como membro do LBQP, agradeço a minha parceira Isabelle Souza Luz, quem me apoiou nos momentos de conclusão da tese, auxiliando muitas vezes durante a escrita e discussão dos resultados.

Agradeço também aos membros do laboratório de interação patógeno-hospedeiro por terem me ajudado durante os experimentos realizados com *Trypanosoma cruzi*. Em especial à colega Tarcila Zuleica Zapparolli, e à Prof. Izabela Marques Dourado Bastos Charneau, que me auxiliaram no cultivo dos parasitos.

Agradeço às agências de fomento FAP-DF pelo projeto que financiou grande parte dessa tese e à CAPES pelo fornecimento da bolsa de estudos que possibilitou a minha dedicação.

Agradeço a minha família, em especial à minha mãe, Rozângela Rodrigues Magalhães, por sempre apoiar minha vida na carreira acadêmica, incentivando e auxiliando em todos os momentos, sendo meu porto seguro nos momentos mais difíceis dessa jornada.

Resumo

A proteômica é uma área de pesquisa que utiliza metodologias capazes de identificar, quantificar e caracterizar proteínas e suas modificações pós-traducionais (PTMs). Existem duas abordagens proteômicas principais baseadas em espectrometria de massas (MS), as quais são denominadas *bottom-up* e *top-down*. A primeira analisa peptídeos provenientes de digestão proteolítica, e a segunda analisa diretamente proteínas intactas do proteoma. Recentemente, a integração das duas abordagens tem possibilitado o aprimoramento na identificação e caracterização de proteoformas, definidas como diferentes formas moleculares produzidas por um único gene. O presente trabalho aplicou as abordagens proteômicas *bottom-up* e *top-down* para o estudo de *Corynebacterium glutamicum*, importante plataforma industrial para produção de aminoácidos, e de *Trypanosoma cruzi*, agente causador da doença de Chagas. Em *C. glutamicum*, a análise proteômica *top-down*, revelou novas proteoformas de Odhl, importante regulador da produção de glutamato, e mepB, peptidase relacionada ao metabolismo de envelope celular. Ademais, a análise proteômica *bottom-up* quantitativa de *C. glutamicum*, comparando condições controle e sob indução para produção de glutamato, revelou proteínas pouco caracterizadas envolvidas no início desse processo. Comparando as mesmas condições, foi iniciado um estudo de integração das abordagens *bottom-up* e *top-down* para *C. glutamicum*, o qual sugeriu influência de proteoformas relacionadas à assimilação de nitrogênio e metabolismo de aminoácidos de cadeia ramificada no processo de produção de glutamato. Além disso, o uso de abordagem *bottom-up* para identificação de PTMs e subsequente utilização dessa informação para caracterização de proteoformas resultou em aumento considerável no número de proteoformas identificadas. Em relação ao *T. cruzi*, foi aplicada a abordagem *bottom-up* para análise proteômica quantitativa de formas axênicas e intracelulares de amastigotas revelando grande número de proteínas reguladas entre as condições. Entre as proteínas reguladas estão presentes trans-sialidases, peptidases *Leishmanolysin-like*, e proteínas associadas ao kDNA, sugerindo importância dessas proteínas no processo de replicação desses parasitos. Finalmente, foram gerados os primeiros dados de proteômica *top-down* para epimastigotas de *T. cruzi*. Esses dados sugerem a identificação de proteoformas causadas por substituição de resíduos em proteínas relacionadas a respostas a estresses ambientais. A análise inicial também apontou metodologias que podem ser otimizadas para melhor caracterização das proteoformas de *T. cruzi*, como a identificação de PTMs por abordagem *bottom-up* e utilização dos dados para identificação de proteoforma por *top-down*.

Palavras chaves: Proteômica *top-down*, Proteômica *bottom-up*, Proteoformas, produção de aminoácidos, doença de Chagas, *Corynebacterium glutamicum*, *Trypanosoma cruzi*.

Abstract

Proteomics is a field of study that employs methodologies capable of identifying and quantifying proteins and their post-translational modifications (PTMs). There are two main proteomic approaches based on mass spectrometry (MS), known as bottom-up and top-down. The former analyzes peptides derived from proteolytic digestion, and the latter directly examines intact proteins. Recently, the integration of these two approaches has enabled improvements in the identification and characterization of proteoforms, defined as different molecular forms produced by a single gene. This study applied bottom-up and top-down proteomic approaches to investigate *Corynebacterium glutamicum*, an important industrial workhorse for amino acid production, and *Trypanosoma cruzi*, the causative agent of Chagas disease.

In *C. glutamicum*, the top-down proteomic analysis revealed new proteoforms of Odhl, an important regulator of glutamate production, and mepB, a peptidase related to cell envelope metabolism. Additionally, quantitative bottom-up proteomic analysis of *C. glutamicum*, comparing control and glutamate production-inducing conditions, unveiled uncharacterized proteins involved in the early stages of this process. Further, a study was initiated to integrate bottom-up and top-down approaches for *C. glutamicum* under the same conditions, suggesting the influence of proteoforms related to nitrogen assimilation and branched-chain amino acid metabolism in the glutamate production process. Moreover, using the bottom-up approach for PTM identification and subsequently employing this information to top-down proteomics resulted in a significant increase in the number of identified proteoforms. In the case of *T. cruzi*, the bottom-up approach was applied for quantitative proteomic analysis of axenic and intracellular amastigote forms, revealing several regulated proteins between conditions. Among the regulated proteins were trans-sialidases, Leishmanolysin-like peptidases, and proteins associated with kDNA, suggesting the importance of these proteins in the replication process of *T. cruzi*. Lastly, the top-down proteomic analysis was performed for *T. cruzi* epimastigotes, allowing the identification of proteoforms formed by amino acid residue substitutions in proteins related with responses to environmental stress. The preliminary analysis also indicated methodologies that can be optimized for a better characterization of *T. cruzi* proteoforms, such as the identification of PTMs using the bottom-up approach and the use of these data for proteoform identification via top-down.

Keywords: Top-down Proteomics, Bottom-up Proteomics, Proteoforms, amino acid production, Chagas disease, Corynebacterium glutamicum, Trypanosoma cruzi.

Sumário

LISTA DE ABREVIATURAS.....	XV
LISTA DE FIGURAS	XVII
Considerações Iniciais	26
Capítulo I. Introdução e referencial teórico.....	27
1.1 Proteômica	28
1.1.1 Proteômica baseada em espectrometria de massas	29
1.1.2 Espectrômetro de massas	29
1.1.3 Abordagens proteômicas <i>top-down</i> e <i>bottom-up</i>	32
1.1.4 Integração de proteômica <i>top-down</i> e <i>bottom-up</i>	37
1.2 <i>Corynebacterium glutamicum</i>	39
1.3 <i>Trypanosoma cruzi</i>	44
Capítulo II. Justificativa, Objetivo e Metas.....	49
2.1 Justificativa.....	50
2.2 Objetivo e metas.....	51
Objetivo geral.....	51
Metas.....	51
Capítulo III. Metodologia Comum	53
3.1 Abordagem <i>bottom-up</i>	54
3.1.1 Digestão para análises quantitativas	54
3.1.2 Digestão multi-enzimática.....	55
3.1.3 Dessalinização.....	59
3.1.4 LC-MS/MS	59

3.1.5	Bioinformática.....	60
3.2	Abordagem <i>top-down</i>	62
3.2.1	Preparo de amostra.....	62
3.2.2	LC-MS/MS.....	63
3.2.3	Análise de dados.....	64
Capítulo IV. Proteômica de <i>Corynebacterium glutamicum</i>		67
4	Resumo.....	68
4.1	Análise <i>top-down</i> descritiva das proteoformas de <i>C. glutamicum</i>	69
4.2	Análise <i>bottom-up</i> quantitativa comparando condição controle vs condição produtora de glutamato.....	69
4.2.1	Objetivos.....	69
4.2.2	Métodos.....	70
Cultivo e cepas.....		70
Análise de aminoácidos.....		72
Extração e quantificação de proteínas.....		73
Análise do perfil de proteínas por SDS-PAGE.....		74
Análise proteômica.....		74
4.2.3	Resultados e discussão.....	74
Cultivo e crescimento bacteriano.....		74
Análise de aminoácidos.....		75
Análise de perfil proteico por SDS-PAGE.....		78
Análise proteômica.....		82
4.2.4	Conclusões.....	96
4.3	Integração de proteômica <i>bottom-up</i> multi-enzimática e <i>top-down</i> para caracterização de proteoformas da bactéria <i>C. glutamicum</i>	97
4.3.1	Objetivos.....	97

4.3.2	Métodos	98
	Cultivo e extração de proteínas.....	98
	Análises proteômicas <i>bottom-up</i> e <i>top-down</i> , e integração dos dados	98
4.3.3	Resultados	99
	Análise global da proteômica <i>top-down</i>	99
	Análise de proteínas com variação de massa (Δm) – Identificações por TopPic	110
	Análise comparativa entre controle e Pen-G.....	114
	Análise proteômica <i>top-down</i> pelo software MetaMorpheus	125
	Abordagem <i>bottom-up</i> multi-enzimática.....	127
	Integração de proteômica <i>bottom-up</i> e <i>top-down</i> por meio da expansão das proteoformas do banco de dados.....	131
4.3.4	Conclusões	132
	Capítulo V. Proteômica de <i>Trypanosoma cruzi</i>	133
	Dedicatória	134
5	Resumo.....	135
5.1	Manuscript: Proteomics of axenic and intracellular amastigote lifeforms of <i>Trypanosoma cruzi</i> in replicative and non-replicative conditions.....	136
5.1.1	Abstract: Proteomics of axenic and intracellular amastigote lifeforms of <i>Trypanosoma cruzi</i> in replicative and non-replicative conditions	136
5.1.2	Introduction	137
5.1.3	Methods	139
	Growth and purification of intracellular amastigotes	139
	Fluorescence microscopy.....	140
	Axenic amastigotes cultures.....	140
	Protein extraction and sample preparation for mass spectrometry analysis	141

LC-MS/MS.....	141
Protein identification and quantification	142
Statistical analysis and biological annotations.....	143
5.1.4 Results.....	144
Analysis of intracellular amastigote purification by DEAE-cellulose chromatography	144
Quantitative proteomic analysis of <i>T. cruzi</i> amastigote forms in replicative and non-replicative conditions.....	147
Trans-sialidases in amastigote replication.....	155
Leishmanolysin-like peptidases and kDNA associated proteins may be involved in amastigote replication.....	157
Proteins with unknown functions	160
5.1.5 Concluding remarks	160
5.1.6 References.....	161
5.2 Integração de proteômica <i>bottom-up</i> e <i>top-down</i> para caracterização de proteoformas da forma de vida epimastigota de <i>Trypanosoma cruzi</i>	168
5.2.1 Objetivos	168
5.2.2 Métodos	169
Células e extração de proteínas.....	169
Preparo de amostra e análise <i>top-down</i> e <i>bottom-up</i> por LC-MS/MS.....	170
Análise de dados <i>bottom-up</i> , <i>top-down</i> e integração das abordagens.....	170
5.2.3 Resultados	171
Análise global das proteoformas detectadas.....	171
Proteoformas identificadas por TopPic Suite	175
Proteoformas identificadas por MetaMorpheus.....	185
Integração de abordagem <i>bottom-up</i> multi-enzimática com <i>top-down</i>	187

5.2.4	Conclusões	188
	Capítulo VI. Considerações finais e perspectivas	189
	Referências	192
	Anexos	210
5.2.5	Anexo 1: Artigo “Revealing <i>Corynebacterium glutamicum</i> proteoforms through top-down proteomics”	210
5.2.6	Anexo 2: Produção bibliográfica e participações em eventos científicos ao longo do doutorado	222
	Primeiro autor.....	222
	Co-autor	222
	Eventos científicos	223

LISTA DE ABREVIATURAS

CZE-*Capillary Zone Electrophoresis*

DDA-*Data-dependent Acquisition*

FDR-*False-discovery rate*

FT-*Fourier transform*

GELFrEE-*Gel Eluted Liquid Fractionation Entrapment Electrophoresis*

HSP-*Heat-shock protein*

ICR-*Ion-Cyclotron Resonance*

IT-*Ion Trap*

LC- *Liquid chromatography*

MS-*Mass spectrometry*

PAGE-*Polyacrylamide Gel Electrophoresis*

PEPPI-MS-*Passively Eluting Proteins from Polyacrylamide gels as Intact species for MS*

PPIase-*peptidil-prolil cis-trans isomerase*

PTCR-*Proton-Transfer Charge Reduction*

PTM(s)-*Post-Translational Modification*

Q-*Quadrupolo*

RT-*Retention Time*

S/N-*Signal/Noise*

SDS-*Sodium dodecyl sulfata*

SDS-PAGE-*Sodium dodecyl sulfata- Polyacrylamide Gel Electrophoresis*

sSEC-*serial Size Exclusion Cromatography*

TOF-*Time-of-Flight*

TxP-*triparedoxina peroxidase*

Δm -variação de massa

LISTA DE FIGURAS

Figura 1. Possíveis origens para diferentes proteoformas. Em nível de DNA é apresentada uma mutação de citosina (C) para timina (T) no exon 1. Entre o pre-RNA e o mRNA são mostradas diferentes possibilidades de splicing alternativos, gerando diferentes mRNA. Após a formação do mRNA, é demonstrada a tradução desses em proteínas, e suas modificações pós-traducionais, acrescentando outras possíveis proteoformas, como acetilação (Ac), fosforilação (P) e clivagem (representada pela diminuição da sequência apresentada em verde). O grupo de todas proteoformas originadas a partir desse mesmo gene apresentado é chamado de família de proteoformas. Adaptado de Schaffer et al., 2019. 33

Figura 2. Representação de abordagem proteômica bottom-up e sua possível influência na identificação de PTMs. São representadas cinco proteoformas proveniente de uma mesma sequência primária de aminoácidos, cada uma delas possuindo diferentes níveis de PTMs biológicas comuns, como clivagem (representada na proteoforma 1), ubiquitinação (Ub, em roxo), fosforilação (P, em amarelo) e metilação (Me, em pastel). A sequência dos eventos da abordagem proteômica bottom-up é então apresentada para cada proteoforma e demonstra como a informação dessas PTMs podem ser perdidas após a digestão em peptídeos trípticos, sendo identificados somente peptídeos não modificados, a inferência de proteoforma identificada por essa abordagem seria a proteoforma sem PTMs. Fonte: Adaptado de Carbonara, Andonovski, & Coorsen, 2021. 34

Figura 3. Processos de fermentação utilizando *C. glutamicum* apresentando seus substratos, possíveis fontes de origem, produtos da fermentação e suas aplicações na sociedade. As fontes de carbono e produtos produzidos sem necessidade de engenharia genética por *C. glutamicum* são representados por caixas com cores mais escuras, enquanto os que já são possíveis produzir, mas essa bactéria precisou passar por modificações genéticas são apresentados pelas caixas com colorações mais claras. Adaptado de Becker & Wittmann, 2012. 41

Figura 4. Ciclo de vida do parasito T. cruzi. Descrição das quatro formas de vida do T. cruzi e suas diferenciações, replicações e infecções. Adaptado de Moretti et al., 2020.....	46
Figura 5. Esquema ilustrado de microcoluna para fracionamento em pH básico. As cores são meramente ilustrativas.	58
Figura 6. Fluxograma de preparo de amostra e análise de dados de proteômica top-down.	63
Figura 7. Fluxograma para integração de dados de proteômica bottom-up e top-down.	66
Figura 8. Fluxograma do cultivo de C. glutamicum até a aquisição de amostras para análises proteômica e de aminoácidos. Ctrl = Controle, Pen-G = Indução por penicilina-G (8 µg/mL), T-40 = Indução por tween-40 (4 mg/mL).....	72
Figura 9. Curva de crescimento de C. glutamicum ATCC 13032 em condição controle, e com indução para produção de glutamato por Penicilina-G (Pen-G) e tween-40. A curva foi feita a partir de dados de OD ₆₀₀ e plotada de acordo com a média e desvio padrão das quatro réplicas biológicas de cada condição.	75
Figura 10. Análise de aminoácidos quantificados e detectados no cultivo de C. glutamicum. A) Concentração de treonina 6,75 h e 9 h após o inóculo. B) Concentração de arginina 6,75 h e 9 h após o inóculo. C) Concentração de glutamato nas condições Pen-G e T-40 9 h, 24 h e 36 h após o inoculo. D) Cromatograma das condições T-40, Pen-G, e controle após 36 h de cultivo sobrepostos a um padrão de glutamina com 25 µM (seta verde). Foi utilizado o teste-t para comparar a concentração dos aminoácidos entre as condições e tempos. As comparações são indicadas pelas barras entre as duas condições analisadas pelo teste-t, e as diferenças significativas estão representadas pelos “*”. * = p-valor ≤ 0,05; ** = p-valor ≤ 0,01.	77
Figura 11. SDS-PAGE de quatro réplicas biológicas das diferentes condições analisadas. Foi realizada eletroforese em malha de poliacrilamida, utilizando gel concentrador de 5% e separador gradiente de 8% a 18%, utilizando o marcador de peso molecular pré-corado BLUeye (Sigma-Aldrich), indicado pela letra “M” com as massas moleculares descritas em kDa. Em cada poço foi adicionado as proteínas	

intracelulares extraídas de cada réplica biológica. Bandas com intensidade de cor diferente entre as condições estão indicadas pelas setas verdes nas condições que sugerem maior abundância, e vermelha nas condições com menor abundancia. A) Análise temporal da condição Controle. B) Análise das condições, Pen-G, T-40 e controle após 24 h de cultivo. C) Análise temporal da condição Pen-G. D) Análise temporal da condição T-40..... 81

Figura 12. Análise de parâmetros globais de quantificação e identificação das proteínas. A) Boxplot da abundância das proteínas. B) Número de proteínas observadas em pelo menos x amostras. C) Análise de outliers por Robust Mahalanobis Distance (RMD). D) Distribuição das abundâncias de todas as proteínas quantificadas em todas as condições..... 83

Figura 13. Análise das diferenças entre as condições analisadas. A) Probabilistic Principal Component Analysis (PPCA) das abundâncias e número de missing values das proteínas analisadas em cada réplica das diferentes condições. B) Volcano plot das comparações analisadas nesse estudo. 84

Figura 14. Proteínas reguladas na condição PenG_24h em comparação com Controle_24h e os processos biológicos relacionados. O número de proteínas reguladas positivamente na condição PenG_24h está apresentado em verde, enquanto as negativamente estão em vermelho. Os termos de Gene Ontology super-representados para os grupos de proteínas de acordo com sua regulação são apresentados ao lado de cada grupo. Análise de super-representação realizada om software shinyGO (Ge, Jung, & Yao, 2020) utilizando cutoff de p-valor = 0,05..... 86

Figura 15. Análise da diferença temporal do cultivo de C. glutamicum, sem indução por penicilina. A) Número de proteínas reguladas. B) Volcano plot das proteínas analisadas. 88

Figura 16. Análise de cluster sensível a variância. Clusterização das abundancias das proteínas sem missing values em três clusters como determinado pela distância mínima entre os centroides (quadro iv). Cada condição foi analisada separadamente: controle (A); PenG (B). Os clusters formados foram nomeados em ordem de acordo com o número de proteínas agrupadas, e são apresentados nos diferentes quadros (i, ii, iii) de cada condição. Análise realizada pelo software

VSclust (Schwämmle & Jensen, 2018) com os dados normalizados pelo MetaMorpheus e transformados por log2.....	89
Figura 17. Análise de super-representação das proteínas de cada cluster obtido pela clusterização sensível à variância. As proteínas de cada cluster foram submetidas separadamente a análise de super-representação em ShinyGO (Ge et al., 2020) e os 5 termos mais enriquecidos (com maior fold enrichment) foram selecionados e plotados em forma de gráfico de bolhas. A) Termos super-representados nos clusters da condição controle. B) Termos super-representados nos clusters da condição PenG.....	91
Figura 18. Interações entre proteínas reguladas com $\log_2FC > 1$ ou $\log_2FC < -1$. As interações das proteínas reguladas foram obtidas por meio do Cytoscape stringapp (N.T. Doncheva, Morris, Gorodkin, & Jensen, 2019), em seguidas foram clusterizadas de acordo com suas interações e os grupos formados foram submetidos a análise de super-representação para identificar funções biológicas relacionadas aos grupos. As funções biológicas super-representadas e as proteínas pertencentes dos grupos estão destacadas em diferentes cores.....	94
Figura 19. Fluxograma do preparo de amostra para análises proteômicas bottom-up e top-down das amostras de <i>C. glutamicum</i>.	98
Figura 20. SDS-PAGE das frações de GELFrEE obtidas a partir das amostras das condições controle e Pen-G. As frações são representadas por números que indicam o tempo em minutos após a saída do azul de bromofenol durante o fracionamento por GELFrEE.	100
Figura 21. Massa dos features detectados e deconvoluídos de acordo com o tempo de eluição de cada fração dos pools analisados. Os espectros adquiridos de cada fração foram deconvoluídos pelo programa TopFD (Basharat et al., 2023) e representados em imagem com auxílio do programa VisioProt-MS (Locard-Paulet et al., 2019). A: Pool da condição Pen-G; B: Pool da condição controle.....	102
Figura 22. Histograma com a diferença de massa (Δm) calculada entre todos os features e relação de intensidade e contagem dessas diferenças da fração GF0. Histograma construído com binwidth = 0,01 pelo programa MStopDiff (Kaulich et al., 2022). A: Condição controle. B: Condição Pen-G.	104

Figura 23. Histogramas com cálculo de diferença de massas entre features (Δm). A: Aproximação na região de baixa massa (0 a 50 Da) do histograma da fração GF0, condição controle. B: Histograma da fração GF15 da condição controle.	107
Figura 24. Análise de massa de todas as proteínas PrSMs de <i>C. glutamicum</i>. A: Histograma da massa de todas as proteínas do proteoma de referência de <i>C. glutamicum</i> disponibilizado no Uniprot. B: Histograma de massa de todos os PrSMs de <i>C. glutamicum</i> identificados nesse estudo. Em ambos os histogramas foi utilizado binwidth de 5 kDa.	109
Figura 25. Análise do número de identificações e massa das proteoformas identificadas nas frações de GELFrEE. A: Número de PrSMs identificados em cada fração a partir dos pools das condições analisadas. B: Violin plot das massas das proteoformas identificadas nas frações de GELFrEE analisadas.	110
Figura 26. Interação entre proteínas identificadas com variação de massa e representação das variações mais comuns por cores. Interações entre proteínas foram mantidas após a clusterização com granularidade de 4, apresentando os termos super-representados de cada grupo.	111
Figura 27. Histogramas da distribuição de variações de massas (Δm) identificadas em <i>C. glutamicum</i>. A) Histograma com todas as variações de massas identificadas. B) Histograma com as variações de massas após exclusão daquelas entre -2,5 Da e 2,5 Da. O binwidth dos dois histogramas foi de 1 Da.	112
Figura 28. Análise das proteínas formiladas e suas demais proteoforms. A) Termos super-representados nas proteínas identificadas com o N-terminal formilado em relação ao genoma de <i>C. glutamicum</i> . B) Número total de proteoformas identificadas em cada proteína com N-terminal formilado.	114
Figura 29. Diagrama de Venn das identificações das condições controle e penicilina-G (Pen-G). A) Comparação na identificação de proteínas. B) Comparação da identificação de proteoformas. C) Termos super-representados para as proteínas exclusivas da condição controle.	115
Figura 30. Análise das proteoformas da proteína reguladora do nitrogênio PII (UniprotAC: Q79VF2). São apresentadas duas proteoformas da proteína PII, com	

as evidências de espectro e sequência de cada uma. No topo de cada figura está apresentada a sequência da proteoforma identificada, com a região modificada preenchida em azul e os fragmentos identificados são representados por traços em azul. No meio das figuras é apresentada o MS1, apresentando o precursor e a janela de isolamento de cada proteoforma. Na parte inferior da figura é apresentado o espectro de massas do MS2, demonstrando os fragmentos identificados de cada proteoforma com o quadro de erro abaixo do espectro. A) Proteoforma com $\Delta m < 1$ Da, provavelmente decorrente de determinação errada de isótopo identificada na condição Pen-G. B) Proteoforma com Δm próximo ao N-terminal com aproximadamente 28 Da, identificada na condição controle. 117

Figura 31. Análise das características bioquímicas previstas da proteína reguladora do nitrogênio PII de *C. glutamicum*. A) Predição de domínio transmembrana por deepTMHMM (Hallgren et al., 2022). B) Índice de hidrofobicidade, de acordo com Kyte & Doolittle (Kyte & Doolittle, 1982). 118

Figura 32. Análise dos espectros das proteoformas da subunidade pequena da acetolactato sintase (uniprotAC: Q79VG4). A sequência de resíduos de aminoácidos das duas proteoformas subunidade pequena da acetolactato sintase são apresentadas no topo de cada figura, com a região modificada preenchida em azul e os fragmentos identificados são representados por traços da mesma cor. No meio das figuras é apresentado o MS1, com o precursor e a janela de isolamento de cada proteoforma. Na parte inferior da figura é apresentado o espectro de MS2, indicando os íons fragmentos identificadores de cada proteoforma. A) Proteoforma canônica. B) Proteoforma com Δm de aproximadamente 14 Da..... 120

Figura 33. Análise dos espectros e sequências das proteoformas de SepF (Q8NNN6). A sequência de resíduos de aminoácidos das duas proteoformas de SepF são apresentadas no topo de cada figura, com a região modificada preenchida em vermelho e os fragmentos identificados são representados por traços em azul. No meio das figuras é apresentado o MS1, com o precursor e a janela de isolamento de cada proteoforma. Na parte inferior da figura é apresentado o espectro de massas, indicando os fragmentos identificadores de cada proteoforma. A) Proteoforma canônica. B) Proteoforma N-terminal acetilado. 123

Figura 34. Proteínas dos proteomas de referência de <i>C. glutamicum</i> (UP000000582) e <i>E. coli</i> (UP000000625) anotadas com modificações pós-traducionais. As informações foram extraídas e interpretadas da base de dados UniprotKB (Bateman et al., 2021). O campo Mod_res se refere às informações armazenadas no campo modified residue do Uniprot, que engloba todas as PTMs, menos lipidações e glicosilações.	126
Figura 35. Análise por MALDI-TOF dos peptídeos obtidos por digestão em FASP utilizando as enzimas Arg-C, Asp-N, Glu-C, Lys-C, Quimiotripsina e Tripsina.	128
Figura 36. Análise de peptídeos modificados identificados por abordagem multi-enzimática. A) Número de peptídeos modificados identificados pelas diferentes enzimas utilizadas. B) 10 modificações mais identificadas nos peptídeos modificados.	130
Figura 37. Comparação das abordagens utilizadas para identificação. As abordagens analisadas são a busca tolerante a erro (TopPic), busca restrita com banco de dados do Uniprot (MetaMorpheus) e busca restrita com banco de dados aumentado a partir das PTMs identificadas por abordagem bottom-up (MetaMorpheus_(uPTM-DB). A) Comparação em número de espectros identificados (PrSMs). B) comparação em nível de proteínas e proteoformas identificadas.	131
Figura 38. Fluxograma para preparo de amostra da forma de vida epimastigota de <i>T. cruzi</i> para análises proteômicas bottom-up e top-down.	169
Figura 39. SDS-PAGE das frações de GELFrEE do lisado das células epimastigotas de <i>T. cruzi</i>. As frações são representadas por números que indicam o tempo em minutos após a saída do azul de bromofenol durante o fracionamento por GELFrEE.....	171
Figura 40. Análise dos eventos detectados por LC-MS e suas massas moleculares de acordo com o tempo de retenção. A) Frações de GELFrEE 0, 1, 2, 3, 4 e 5. B) Frações de GELFrEE 7, 9, 11, 13, 15 e 20. Análise realizada com threshold de 20% para intensidade dos eventos detectados usando o programa VisioProt-MS (Locard-Paulet et al., 2019).	172

Figura 41. Análise de diferença de massas (Δm) de eventos de MS1 detectados. A) Δm calculadas na fração GF0. B) Δm calculadas na fração GF11. Ambos histogramas foram construídos com binwidth = 0,01. Calculos e construção de histograma realizados pelo programa MStopDiff (Kaulich et al., 2022).....	174
Figura 42. Análise dos PrSMs, proteoformas e proteínas identificadas em cada fração. A) Número de PrSMs e proteínas identificadas em cada fração. B) Violin plot representando a distribuição de massa das proteoformas presentes em cada fração.	176
Figura 43. Análise de super-representação das proteínas identificadas com variação de massa (Δm). Foram considerados termos super-representados se q-valor < 0,05 e mínimo de duas proteínas no termo. Análise realizada com ShinyGO 0.77 (Ge et al., 2020).	177
Figura 44. Análise dos PrSMs com Δm mais frequentes ou N-terminal acetilado. A) O número de PrSMs para as variações de massas foi calculado considerando o valor de Δm arredondado, sem casas decimais. B) Possíveis causas para as variações de massas decorrentes de substituições de resíduos de aminoácido. Obtido em unimod.org.....	178
Figura 45. Análise de função das proteoformas com N-terminal acetilado. Códigos das proteínas foram submetidos a análise de interação e clusterização (inflation = 10) em cytoscape stringAPP (Nadezhda T. Doncheva, Morris, Gorodkin, & Jensen, 2019).	179
Figura 46. Análise de espectros e sequencias das proteoformas identificadas com Δm de 27 Da. Em A e B são apresentadas análises dos PrSMs identificados. No topo dessas imagens está a sequência identificada, com traços em azul representando os fragmentos identificados, destaque em azul na região de identificação para a variação de massa, e colchetes em vermelho representando onde ocorreu uma possível clivagem da proteína. No meio das imagens, é mostrado o MS1 e a janela de exclusão utilizada no precursor identificado. Na parte inferior está o espectro de massas e gráfico de erros em ppm da massa dos fragmentos identificados. A) Análise de PrSM de HSP 10 kDa (UniprotAC: Q4DFB0). B) Análise de PrSM de HSP 10 kDa (UniprotAC: Q4D000). C) Alinhamento global entre as três	

proteínas HSP 10 kDa (UniprotACs: Q4DFB0; Q4D000; Q4DFA8) identificadas com Δm de 27 Da. A região onde ocorre 100 % de identidade está preenchida em azul, e onde não ocorre está destacada em vermelho. Alinhamento realizado pelo programa Clustal omega (Sievers & Higgins, 2018) e imagem gerada por Jalview (Waterhouse, Procter, Martin, Clamp, & Barton, 2009). 181

Figura 47. Análise dos PrSMs identificadores da proteína peptidil-prolil cis-trans isomerase (PPIase, Q4E4L9). A) Análise de PrSM da proteoforma identificada como mais próxima da canônica. No superior está a sequência da proteína com os fragmentos identificados em MS2 representados por traços azuis, possível perda de resíduos apresentada por um colchete em vermelho e local de modificação por acetilação do N-terminal apresentada por um sombreamento em vermelho. No meio está o precursor dos fragmentos identificados dessa proteoforma. Na parte inferior está apresentado o espectro de massas dos fragmentos identificados juntos a sua barra de erros (ppm). B) Precursores identificadores de diferentes proteoformas de PPIase. 184

Figura 48. Análise do tempo de retenção (RT) das proteoformas de TxP e PPIase de acordo com suas massas...... 185

Figura 49. Análise de PTMs anotadas em bancos de dados de T. cruzi e S. cerevisiae. Informações obtidas do UniprotKB a partir dos proteomas de referência de T. cruzi CL Brener (UP000002296) e S. cerevisiae (UP000002311) (Bateman et al., 2021). 187

Considerações Iniciais

Esta tese está organizada em três capítulos. O capítulo I apresenta uma Introdução Geral sobre o tema: “Proteômica”, que é o arcabouço principal da tese. Serão assim apresentados termos e conceitos fundamentais da espectrometria de massas e das abordagens *top-down* e *bottom-up* para proporcionar uma melhor compreensão das análises proteômicas apresentadas nos capítulos posteriores. Também é apresentada revisão bibliográfica sobre os dois organismos objetos de estudo dessa tese: *Corynebacterium glutamicum* e *Trypanosoma cruzi*.

Considerando o referencial teórico apresentado, no capítulo II são apresentadas a justificativa, o objetivo, e as metas desse trabalho. Por sua vez, no capítulo III são abordadas as metodologias proteômicas gerais utilizadas.

O capítulo IV tem como tema os seguintes estudos conduzidos com a bactéria *C. glutamicum*: a) “Análise proteômica *top-down* descritiva das proteoformas de *C. glutamicum*”; b) “Análise *bottom-up* quantitativa comparando condição controle Vs condição produtora de glutamato”; e c) “Integração de proteômica *bottom-up* multi-enzimática e *top-down* para caracterização de proteoformas de *C. glutamicum*”.

O capítulo V diz respeito aos estudos com o protozoário *Trypanosoma cruzi* relativos à análise proteômica de formas amastigotas, assim como resultados iniciais de proteômica *top-down* de formas epimastigotas. Além disso, são discutidas as possibilidades do estudo de integração proteômica *bottom-up* e *top-down* para caracterização das proteoformas de *T. cruzi*.

No capítulo VI são feitas as considerações finais e apresentadas as perspectivas geradas por esse trabalho.

Capítulo I. Introdução e referencial teórico

1.1 Proteômica

A proteômica é o estudo de proteínas em larga escala (Pandey & Mann, 2000). As proteínas apresentam grande diversidade funcional, desde catálise de reações bioquímicas, sinalização, transdução de sinal, estrutural, transporte de moléculas, entre outras (Sinha & Mann, 2020). O conjunto de proteínas de determinado organismo ou fração de um organismo expresso sob condições definidas é chamado de proteoma (Zimmer, Monroe, Qian, & Smith, 2006). A análise proteômica apresenta grande relevância na elucidação da função biológica dos produtos gênicos (proteínas), sendo esses os agentes ativos das funções celulares (Pandey & Mann, 2000). Além disso, a proteômica possibilita a quantificação relativa entre condições, determinação da localização espacial de proteínas, e a identificação de modificações pós-traducionais (PTMs - *Post-Translational Modifications*), mudanças essas que não podem ser previstas por informações gênicas (Smith et al., 2002). As PTMs são capazes de afetar a função, a estrutura, a velocidade de reação, entre outras características de proteínas (Grangeasse, Stülke, & Mijakovic, 2015), demonstrando assim a importância da análise proteômica na interpretação de fenótipos e descrição molecular de processos biológicos.

Em meados do século 2000, a proteômica estava frequentemente relacionada a análise de géis de eletroforese bidimensional (Pandey & Mann, 2000). A partir do início da década de 1990, abordagens por espectrometria de massas (MS - *Mass spectrometry*) passaram a ganhar muito espaço na análise de biomoléculas, entre elas as proteínas. Essa mudança foi impulsionada pelo sequenciamento de diversos genes e genomas e desenvolvimento dos métodos de ionização de proteínas (Aebersold & Mann, 2003). Mais recentemente, a proteômica está frequentemente relacionada ao estudo quantitativo do proteoma, sendo regularmente utilizada na comparação de diferentes condições (Sinha & Mann, 2020) e tem como abordagem de grande valor a espectrometria de massas (Schaffer et al., 2019).

1.1.1 Proteômica baseada em espectrometria de massas

A espectrometria de massas (MS – *Mass spectrometry*) é uma poderosa ferramenta analítica, amplamente utilizada na caracterização, identificação e quantificação de biomoléculas, como proteínas e metabólitos (Savaryn, Toby, & Kelleher, 2016). Por sua vez, a proteômica baseada em MS é uma abordagem que acopla diferentes campos de estudos, muitas vezes utilizando conhecimentos de bioquímica para o preparo de amostras, química analítica, para compreensão da instrumentação, e de biologia computacional e estatística para a análise de dados (Sinha & Mann, 2020). Essas análises são diversificadas, podendo variar de acordo com a pergunta biológica, tipo de amostra, complexidade, entre outros fatores (Savaryn et al., 2016). Apesar dessa diversidade de possibilidade, todas as análises tem em comum a utilização de um espectrômetro de massas (Shuken, 2023). Nessa sessão serão abordados de maneira introdutória os diferentes campos de conhecimentos utilizados para a compreensão da proteômica baseada em espectrometria de massas: Espectrômetro de massas; abordagens proteômicas; e métodos computacionais na análise proteômica.

1.1.2 Espectrômetro de massas

Um espectrômetro de massas é por definição constituído por três componentes principais: 1) fonte de ionização; 2) analisador de massas, 3) detector (Sinha & Mann, 2020). Esses equipamentos são comumente referidos de acordo com seus analisadores de massas, por exemplo: *Ion-Trap* (IT), *Time-Of-Flight* (TOF), *orbitrap*, *Ion-Cyclotron Resonance* (ICR), *quadrupolos* (Q) (Savaryn et al., 2016). O espectrômetro de massas tem como principal função o aferimento/separação de íons na fase gasosa de acordo com sua relação massa/carga (m/z), e funcionam a partir da aplicação de correntes elétricas em frequências variadas, que geram campos eletromagnéticos para condução dos íons em fase gasosa (Savaryn et al., 2016; Shuken, 2023). Atualmente, existe uma grande diversidade de espectrômetros de massas comercialmente disponíveis, combinando diferentes tipos de analisadores de massas. Diferentes analisadores

de massas são baseados em princípios diferentes para a determinação do m/z (Savaryn et al., 2016). Serão mais bem abordados os princípios de funcionamento de analisadores dos tipos Quadrupolo, TOF e Orbitrap.

Um analisador de massas quadrupolo funciona como um filtro de massas, permitindo a passagem de todos os íons, ou somente a de um conjunto com determinado m/z . Esse analisador é constituído por quatro bastões que tem aplicação de correntes com frequências de amplitudes iguais, mas sinais opostos para gerar o campo eletromagnético. A partir da amplitude da frequência utilizada são filtrados os íons de relação m/z menores, de forma que ao aumentar a amplitude íons com menor m/z são mais suscetíveis a se chocarem com os bastões ou não conseguem passar pelo quadrupolo. Enquanto a partir da diferença de carga entre os bastões são filtrados os íons com maior m/z , íons carregados positivamente são atraídos para o polo negativo, no entanto o campo eletromagnético gerado pela aplicação de corrente os puxa de volta para o centro do quadrupolo, apesar disso, íons com m/z maiores tem muita inercia, o que dificulta sua volta a trajetória e ocasiona no choque com os bastões ou saída do quadrupolo (Savaryn et al., 2016). Normalmente os íons que passam pelo quadrupolo são direcionados a um detector, outro analisador de massas, ou a uma câmara de fragmentação.

Analisadores de massas do tipo TOF se baseiam na velocidade das moléculas em uma região sem campo elétrico, normalmente chamada de *drift tube*, após aceleração homogênea por um campo elétrico de aproximadamente 20 kV para determinar a sua relação m/z (Sinha & Mann, 2020). Esse analisador normalmente está acoplado a um detector de íons do tipo MCP (*microchannel plate*), os quais são estimulados pela chegada de íons que resulta na ejeção de elétrons da superfície do detector, resultando em uma propagação de sinal, o que resulta em alta sensibilidade (Sinha & Mann, 2020).

Por sua vez, analisadores de massas do tipo Orbitrap determinam a relação m/z por meio da aplicação de um campo eletromagnético que resulta na movimentação dos íons em uma trajetória oscilatória. O Orbitrap é dividido em dois polos, e a partir da movimentação dos íons, ao redor desses polos, eles são

detectados e é determinado em domínio de tempo a oscilação de cada íon. A informação em domínio de tempo é então transformada para frequência usando transformada de Fourier (FT – *Fourier transform*) e a frequência dessas moléculas pode ser utilizada para determinar a relação m/z (Savaryn et al., 2016).

Espectrômetros detectam moléculas carregadas em fase gasosa, pois essas precisam ser suscetíveis a campos eletromagnéticos para serem analisadas. A ionização das moléculas é feita pelo componente comumente referido como fonte de ionização (Sinha & Mann, 2020). Dois tipos de ionização muito utilizados na análise proteômica são ESI (*electrospray ionization*) e MALDI (*Matrix-assisted laser desorption ionization*) (Gao & Yates, 2019), a relevância do desenvolvimento e descoberta desses métodos foi reconhecida pelo prêmio nobel em 2002 (Aebbersold & Mann, 2003). Os peptídeos ou proteínas muitas vezes precisam sair da fase aquosa para a fase gasosa e serem ionizados nesse processo. A ionização por *electrospray* possibilita isso essa transformação por meio da formação do bombeamento do líquido contendo os peptídeos junto a uma fonte de carga, normalmente um ácido, até a saída desse líquido por um orifício de dimensão micrométrica, o qual está sob diferença de potencial de alta voltagem. Esse processo resulta na formação rápida de micro gotículas cheias de carga, as quais evaporam rapidamente, gerando as moléculas carregadas que são direcionadas para os analisadores de massas (Sinha & Mann, 2020).

A determinação da m/z da proteína ou peptídeos analisados fornece informação valiosa sobre a molécula, no entanto, normalmente não é suficiente para identifica-la. Para adquirir outro nível de informação sobre essas moléculas, elas normalmente são fragmentadas, e seus fragmentos analisados novamente em analisadores de massas. A fragmentação mais comum em análises proteômicas é a baseada em colisão com gás inerte, podendo ser do tipo CID (*Collision-induced dissociation*) ou HCD (*Higher-energy collisional dissociation*). No entanto, métodos alternativos como ETD (*Electron-transfer dissociation*) podem fornecer informações complementares (Gao & Yates, 2019).

Análises proteômicas normalmente utilizam a interface da fonte de ionização por ESI ao espectrômetro de massas acoplada a uma cromatografia líquida (LC – *Liquid chromatography*). O acoplamento de cromatografia líquida de alta performance à espectrômetro de massas normalmente é referido como LC-MS ou LC-MS/MS (Aebersold & Mann, 2016). Dessa forma o sistema de cromatografia líquida acoplado a fonte de ionização realiza as funções de separação parcial de proteínas/peptídeos em picos cromatográficos e os concentra para ionização por ESI (Shuken, 2023).

1.1.3 Abordagens proteômicas *top-down* e *bottom-up*

A proteômica baseada em espectrometria de massas faz uso de duas abordagens principais denominadas *bottom-up* e *top-down*. A principal diferença entre essas abordagens é que na *bottom-up* as proteínas a serem analisadas são clivadas por uma protease (normalmente tripsina) e os peptídeos resultantes são depois submetidos a análise por MS, enquanto na abordagem *top-down* as proteínas não passam por processo de clivagem enzimática e são submetidas à MS em sua forma intacta (KELLIE et al., 2012). A abordagem *top-down* é vantajosa em relação à *bottom-up* quando se trata da análise de PTMs (Catherman, Skinner, & Kelleher, 2014).

Recentemente foi alcunhado o termo proteoforma para uniformizar e descrever a complexidade de formas que um produto gênico pode assumir. Proteoformas são todas as formas proteicas geradas a partir de um mesmo gene. Essas podem ser resultado de modificações pós-transcricionais, como *splicing* alternativo, e PTMs (L. M. Smith & Kelleher, 2013). Mais atualmente, esse termo foi expandido para “famílias de proteoformas”, descritas como grupo de proteoformas provenientes de um mesmo gene (Cesnik et al., 2018; Schaffer et al., 2019) (figura 1).

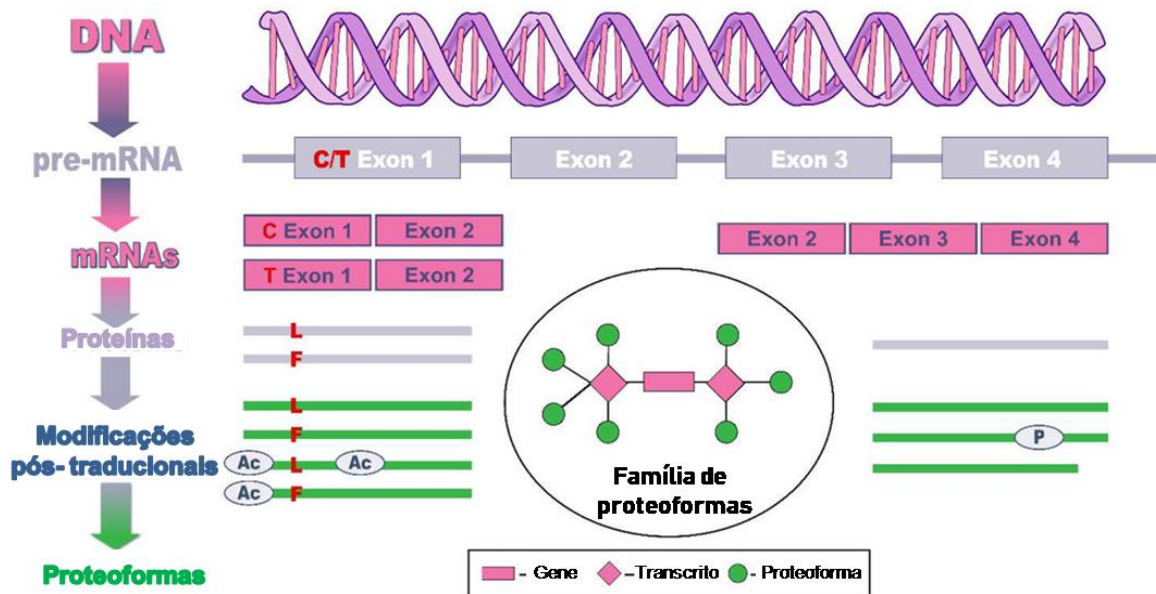


Figura 1. Possíveis origens para diferentes proteoformas. Em nível de DNA é apresentada uma mutação de citosina (C) para timina (T) no exon 1. Entre o pre-RNA e o mRNA são mostradas diferentes possibilidades de *splicing* alternativos, gerando diferentes mRNA. Após a formação do mRNA, é demonstrada a tradução desses em proteínas, e suas modificações pós-traducionais, acrescentando outras possíveis proteoformas, como acetilação (Ac), fosforilação (P) e clivagem (representada pela diminuição da sequência apresentada em verde). O grupo de todas proteoformas originadas a partir desse mesmo gene apresentado é chamado de família de proteoformas. Adaptado de Schaffer et al., 2019.

Na abordagem proteômica *bottom-up*, a perda de informação de PTMs pode ocorrer por diversos processos, como o a perda por clivagem, a não ionização de peptídeo modificado e a não fragmentação dos peptídeos portadores de PTMs. Sendo assim, os peptídeos responsáveis pela identificação de proteínas podem não conter PTMs, no entanto essa informação pode estar presente em outras sequências não identificadas (figura 2). Já na abordagem *top-down*, ao analisar a proteína intacta, a detecção anterior a fragmentação sempre vai informar a massa da proteína com todas as suas modificações presentes no momento (Carbonara, Andonovski, & Coorsen, 2021; Catherman et al., 2014).

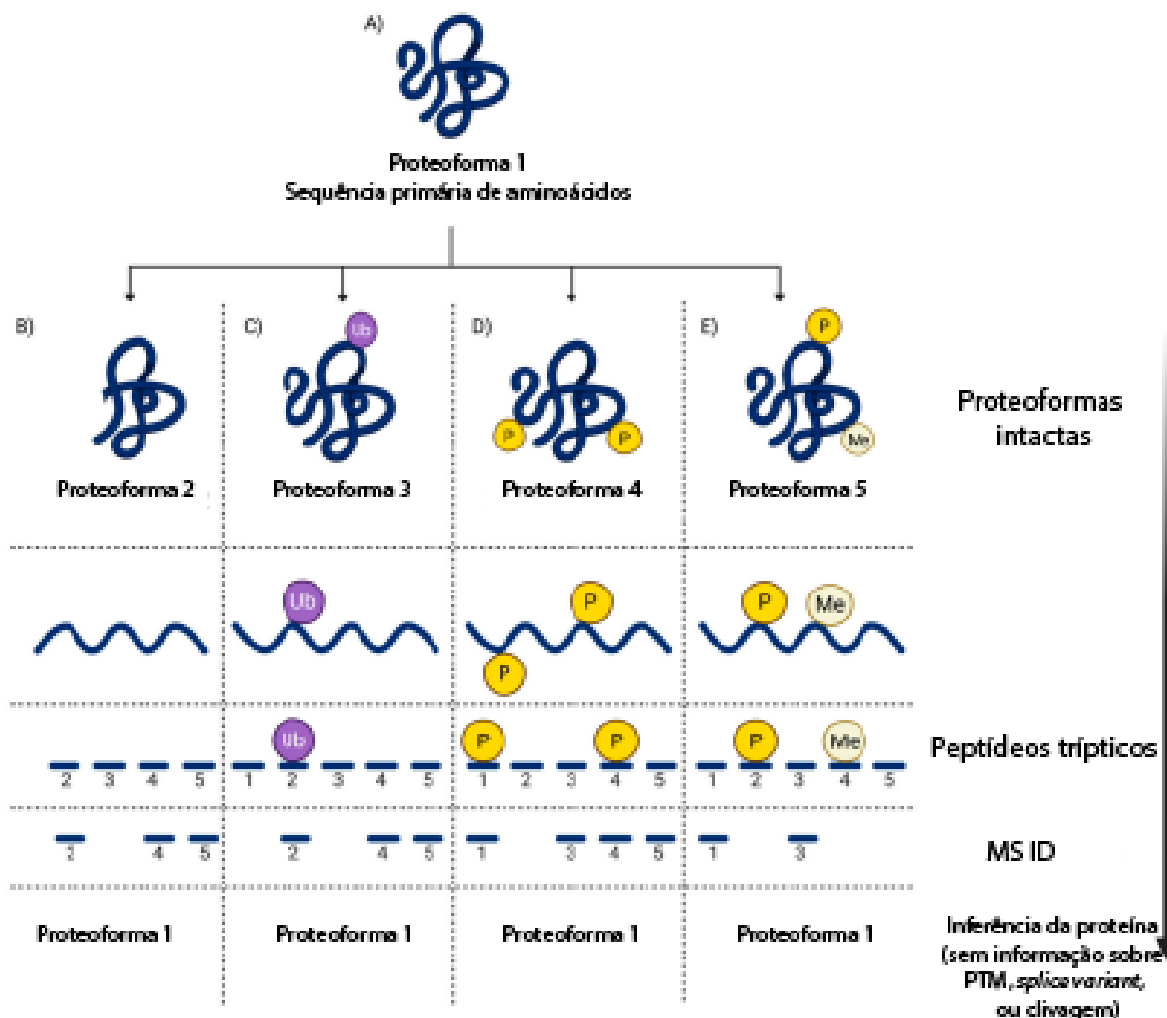


Figura 2. Representação de abordagem proteômica *bottom-up* e sua possível influência na identificação de PTMs. São representadas cinco proteoformas proveniente de uma mesma sequência primária de aminoácidos, cada uma delas possuindo diferentes níveis de PTMs biológicas comuns, como clivagem (representada na proteoforma 1), ubiquitinação (Ub, em roxo), fosforilação (P, em amarelo) e metilação (Me, em pastel). A sequência dos eventos da abordagem proteômica *bottom-up* é então apresentada para cada proteoforma e demonstra como a informação dessas PTMs podem ser perdidas após a digestão em peptídeos tripticos, sendo identificados somente peptídeos não modificados, a inferência de proteoforma identificada por essa abordagem seria a proteoforma sem PTMs. Fonte: Adaptado de Carbonara, Andonovski, & Coorsen, 2021.

Atualmente, a abordagem *bottom-up* está mais bem estabelecida e possui mais recursos instrumentais, analíticos e bioinformáticos quando comparada à *top-down*. Isso se deve, em parte, devido às limitações técnicas de espectrômetros para analisarem moléculas grandes, como proteínas intactas, no início do desenvolvimento da proteômica por espectrometria de massas (CHAIT, 2006). Essa vantagem da metodologia *bottom-up* se sobressai quando é necessário fazer

comparações quantitativas entre condições, fato evidenciado pelas várias estratégias de quantificação relativa de proteínas já desenvolvidas (XIE et al., 2011). A abordagem *bottom-up* também é mais apropriada para caracterizar proteínas de massa molecular acima de 30 kDa, devido a limitações analíticas da MS, entre elas a diluição de sinal das proteoformas de alta massa (Fornelli & Toby, 2022). No entanto, a proteômica *bottom-up* apresentam algumas desvantagens, por exemplo, a adição da etapa de digestão no preparo da amostra, já que qualquer etapa adicionada anterior à espectrometria de massa pode afetar o resultado final e a reprodutibilidade do experimento (Schubert, Röst, Collins, Rosenberger, & Aebersold, 2017). Além dessa diferença, vantagem excepcional da abordagem *top-down* é não necessitar de uma etapa de inferência, que ocorre em *bottom-up*, considerando que nessa última abordagem não se tem informação experimental da massa molecular da proteína precursora dos peptídeos analisados (Yansheng Liu, Beyer, & Aebersold, 2016).

Apesar das vantagens da metodologia *bottom-up*, a análise de proteínas intactas vem se tornando cada vez mais robusta com os avanços tecnológicos, aumento de acurácia e resolução dos espectrômetros de massa e melhorias em técnicas que antecedem a etapa de espectrometria de massa, como o GELFrEE (*Gel-eluted liquid fraction entrapment electrophoresis*) (Tran & Doucette, 2009), o sSEC (*serial Size Exclusion Chromatography*) (Cai et al., 2017), CZE (*Capillary Zone Electrophoresis*) (McCool et al., 2018), *capillary* HILIC-MS (Gargano et al., 2018), e PEPPPI-MS (*Passively Eluting Proteins from Polyacrylamide gels as Intact species for MS*) (Takemori et al., 2020). Esses métodos são tipos de fracionamento, que causam a divisão de amostras iniciais em frações menos complexas. Esses procedimentos são importantes para a análise de proteínas por MS, por resultarem na redução da complexidade das amostras, o que melhora a proporção de sinal/ruído e a cobertura de identificação e quantificação, principalmente de peptídeos e proteínas em pequenas quantidades na amostra (MOSTOVENKO et al., 2013). Além do fracionamento de proteínas, a dificuldade de solubilização dessas moléculas em soluções amenas para análise por LC-MS/MS também tem avançado nos últimos anos com a implementação de solubilização de proteínas

utilizando ácido fórmico concentrado gelado (Doucette, Vieira, Orton, & Wall, 2014) e desenvolvimento de detergentes compatíveis com espectrometria de massas, como o Azo (K. A. Brown et al., 2019). Em relação a otimizações de espectrômetros de massas e seus métodos para análise de proteínas intactas, mais especificamente proteínas de alta massa, as seguintes técnicas tem apresentado resultados promissores: *Proton-Transfer Charge Reduction* (PTCR) (Huguet et al., 2019); MS nativa; abordagem com aquisição *medium-high* (Fornelli et al., 2017), entre outros (Fornelli & Toby, 2022). Alternativamente métodos de aquisição de íons individuais (*individual ion mass spectrometry* – I²MS) estão sendo desenvolvidos com resultados que já possibilitaram a detecção com resolução isotópica de proteínas grandes como anticorpos (Fornelli & Toby, 2022).

Outro desafio enfrentado pela abordagem *top-down* é a análise de dados, caracterizada por três etapas: preparação dos espectros, deconvolução e caracterização das proteoformas (Zhong et al., 2021). A identificação de proteoformas é um desses desafios e as estratégias atuais para sua solução podem ser divididas em dois tipos: 1) expansão do banco de dados de proteoformas; e 2) busca cega de modificações das sequências de proteínas nos bancos de dados. A primeira abordagem basicamente faz uso de informações de modificações anotadas em bancos de dados, normalmente depositadas no Uniprot, enquanto a segunda estratégia é baseada na comparação de espectros de fragmentos experimentais com teóricos, sem levar em consideração o precursor, ou minimizando a sua influência, reduzindo o espaço/tempo de busca por meios computacionais como algoritmos de alinhamento de espectros ou *graph-based models* (Zhong et al., 2021).

Considerando as vantagens e limitações das abordagens proteômicas *top-down* e *bottom-up*, muitas vezes elas são consideradas complementares e estudos são realizados a partir da integração das duas estratégias (Ntai et al., 2016; Petras, Heiss, Süssmuth, & Calvete, 2015).

1.1.4 Integração de proteômica *top-down* e *bottom-up*

A integração de abordagens *top-down* e *bottom-up* foi utilizada, por exemplo, para a confirmação da identificação de proteoformas do veneno da cobra rei, *Ophiophagus hannah*, além de possibilitar a identificação de proteoformas com especificidade de sítios de PTMs (Petras et al., 2015). Outro estudo de integração de *bottom-up* com *top-down* foi feito analisando xenógrafos de câncer de mama basal e câncer de mama do tipo luminal B. A partir de seus resultados foi possível detectar proteoformas com SNPs (*single nucleotide polymorphism*) específicos de cada tipo (Ntai et al., 2016). A relevância de PTMs também foi demonstrada na bactéria patogênica *Neisseria meningitidis*, que permanece na garganta em sua forma assintomática ou muda para formas sintomáticas, causando sepse ou meningite. Células de *N. meningitidis* podem ser liberadas dos agregados formados na garganta, dando acesso aos outros sistemas e podendo ocasionar em suas formas sintomáticas. A liberação de células a partir de agregados de colônias de *N. meningitidis* depende de uma fosfogliceroilação (adição de grupamento fosfoglicerol) em organelas filamentosas, chamadas de pili IV (Chamot-Rooke et al., 2011). A proteína Pile é o maior constituinte do pili tipo IV. Além da adição de grupamento fosfoglicerol, também já foram identificados diferentes tipos de glicosilações e fosforilações nessa proteína. Em uma abordagem integrando proteômica *bottom-up* e *top-down* foi demonstrado a necessidade do uso das duas abordagens para a localização e caracterização de todas as proteoformas de Pile. Foi identificado que essas proteoformas eram diferentes pela massa de um glicano, DATDH (2,4-diacetamido 2,4,6-trideoxy α -D-hexose). A estratégia *bottom-up* foi capaz de identificar os sítios de modificação e identificar as proteoformas de menor e maior massa, no entanto duas proteoformas de massas intermediárias só foram possíveis de serem determinadas por meio da abordagem *top-down* (Gault et al., 2014).

A integração das abordagens proteômicas *bottom-up* e *top-down* foi realizada recentemente de forma mais automatizada utilizando o programa *Proteoform Suite*, de forma que ele utiliza as identificações de *bottom-up* para auxiliar na identificação das proteoformas e a localização de PTMs usando dados

top-down (Schaffer, Millikin, Shortreed, Scalf, & Smith, 2020). O programa Proteoform Suite também já fez uso de outra abordagem chamada de *Enhanced Global Post-Translational Modification Discovery* (G-PTM-D), que possibilita a identificação de PTMs não anotadas nos bancos de dados de maneira precisa, a partir de dados *bottom-up* (Solntsev, Shortreed, Frey, & Smith, 2018). A união do G-PTM-D a uma estratégia multi-enzimática para caracterização dos peptídeos possibilitou o aumento de cobertura, com consequente redução de ambiguidade de identificação de proteínas, e aumento no número de PTMs identificadas (Miller et al., 2019). A combinação dessas abordagens para descoberta e anotação de PTMs com alta confiança e construção de bancos de dados para identificação de proteínas intactas e usando abordagem *top-down* foi explorada em diferentes organismos, como camundongos, humanos e *Escherichia coli*, demonstrando melhora na identificação e caracterização de proteoformas (Yunxiang Dai et al., 2019, 2017; Schaffer et al., 2018). Outra ferramenta que busca essa integração é o *ProteoCombiner*, capaz de integrar e visualizar dados provenientes de diferentes ferramentas de identificação de proteínas por abordagem *bottom-up* e *top-down* (Lima et al., 2020). Esse programa auxilia na desambiguação de proteoformas usando dados de *bottom-up*, além de gerar um *score* (*combined score*) para as proteoformas identificadas baseado nas identificações de *bottom-up* e *top-down* conjuntamente. Como foi demonstrado, o desenvolvimento de integração de dados proteômicos de *bottom-up* e *top-down* está crescente e atualmente é essencial para a melhoria das identificações de proteoformas tanto em quesito de número, quanto em qualidade e caracterização das mesmas.

1.2 *Corynebacterium glutamicum*

A ordem *Corynebacteriales*, pertencente ao filo *Actinobacteria*, é composta por bactérias que possuem ácido micólico na parede celular (Houssin et al., 2020). As principais bactérias patogênicas dessa subordem são a *Corynebacterium diphtheria*, agente causador da difteria em humanos, *Corynebacterium ulcerans*, causador de doenças em animais domésticos e selvagens, *Mycobacterium tuberculosis*, além dos patógenos oportunistas *M. abscessus* e *C. jeikeium* (Georgi, Rittmann, & Wendisch, 2005; Houssin et al., 2020). O gênero *Corynebacterium*, o qual possui 90 espécies, compreende organismos relevantes à saúde humana, animal e também de interesse industrial (Tauch & Burkovski, 2015). *Corynebacterium glutamicum* é o representante com maior importância biotecnológica desse grupo taxonômico. (Kalinowski et al., 2003). Além disso, essa bactéria tem sido utilizada como modelo para o estudo de bactérias patogênicas filogeneticamente próximas (Tauch & Burkovski, 2015).

C. glutamicum possui a forma de cajado, não é esporulante, geralmente é encontrada no solo e quando cultivada em meio sólido, forma colônias amareladas (Tryfona & Bustard, 2005). Foi isolada pela primeira vez em 1956 em um zoológico do Japão quando, na época, a companhia Kyowa Hakko buscava microrganismos selvagens capazes de produzir aminoácidos. A descoberta de microrganismos produtores de aminoácidos foi marcante pois até então acreditava-se que microrganismos não seriam capazes de produzir metabólitos primários em grande quantidade (Hirasawa & Shimizu, 2016).

Ao longo dos anos, algumas dessas cepas passaram a ser utilizadas para produzir aminoácidos em escala industrial e foi conferido a elas o *status* de “geralmente reconhecida como segura” ou *status* GRAS (*Generally Recognized as Safe*), o que aumenta seu potencial de aplicação biotecnológica (Lee, Na, Kim, Lee, & Kim, 2016).

C. glutamicum possui um sistema metabólico com grande capacidade para produção de biomoléculas de interesse industrial (Lee et al., 2016). Além da produção de aminoácidos, já foram descritos para *C. glutamicum* processos de

produção de isobutanol (Yamamoto, Suda, Niimi, Inui, & Yukawa, 2013), lactato (Okino, Inui, & Yukawa, 2005), pantotenato (Hüser et al., 2005), succinato (Litsanov, Brocker, & Bott, 2012) , alguns polímeros (Q. Liu, Ouyang, Kim, & Chen, 2007; Matsumoto, Kitagawa, Jo, Song, & Taguchi, 2011), dentre outras biomoléculas (Becker & Wittmann, 2012) (figura 3). Recentemente, também vem sendo investigada a capacidade dessa bactéria como agente de bioremediação (Ray et al., 2022). Apesar do potencial de produção de diferentes biomoléculas, a principal utilização industrial da bactéria permanece sendo a de aminoácidos, cujo principal método para a produção industrial é a fermentação, sendo os principais organismos utilizados pela indústria as bactérias *C. glutamicum* e *Escherichia coli* (Leuchtenberger, Huthmacher, & Drauz, 2005).

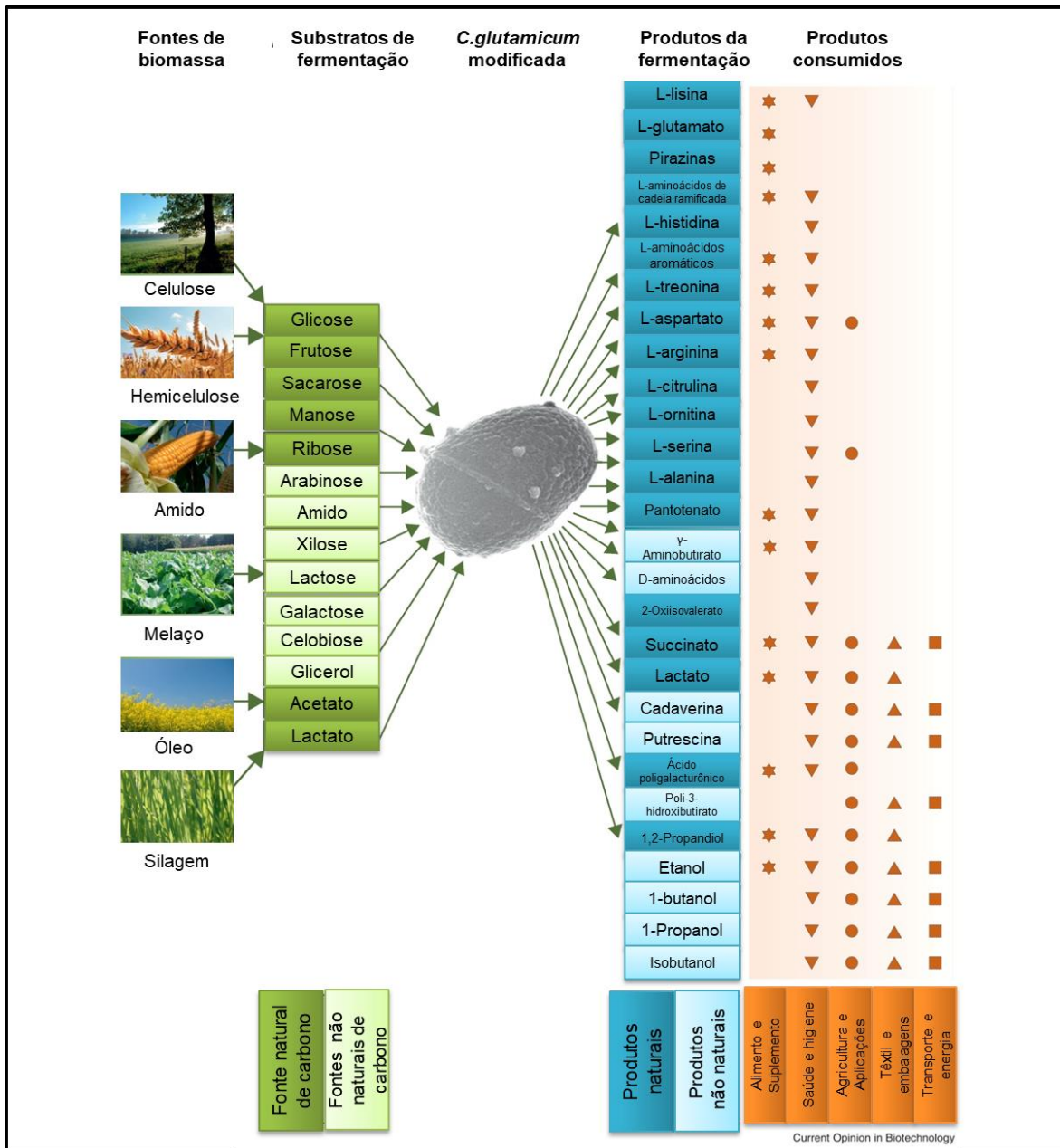


Figura 3. Processos de fermentação utilizando *C. glutamicum* apresentando seus substratos, possíveis fontes de origem, produtos da fermentação e suas aplicações na sociedade. As fontes de carbono e produtos produzidos sem necessidade de engenharia genética por *C. glutamicum* são representados por caixas com cores mais escuras, enquanto os que já são possíveis produzir, mas essa bactéria precisou passar por modificações genéticas são apresentados pelas caixas com colorações mais claras. Adaptado de Becker & Wittmann, 2012.

Aminoácidos são o terceiro grupo de moléculas mais produzidas pelo método de fermentação industrial, atrás apenas da produção de etanol e antibióticos,

respectivamente primeiro e segundo colocados (Leuchtenberger et al., 2005). Em 2016, a produção global de aminoácidos alcançou 2,5 milhões de toneladas/ano (Hirasawa & Shimizu, 2016), e foi estimada para alcançar 10 milhões de toneladas em 2022. Os valores de mercado dessa produção chegaram a aproximadamente 13 bilhões de dólares em 2016, com crescimento anual de 7 % (Ikeda & Takeno, 2020).

Os principais aminoácidos produzidos pelo processo fermentativo industrial são o ácido glutâmico e a lisina. A produção de ácido glutâmico é direcionada principalmente para a indústria alimentícia, sendo um dos constituintes do tempero glutamato monossódico, comercializado pela empresa multinacional de origem japonesa Ajinomoto. A lisina está no grupo dos aminoácidos que tem como finalidade comercial a suplementação humana e animal, servindo principalmente como suplemento alimentar de aves domésticas (Leuchtenberger et al., 2005). Outros aminoácidos que integram o grupo de suplementação são a metionina, treonina e o triptofano, enquanto aspartato e fenilalanina são utilizados na indústria alimentícia para a produção do adoçante Aspartame (Ikeda & Takeno, 2020).

A superprodução de ácido glutâmico por *C. glutamicum* pode ser estimulada pela adição de detergentes, e antibióticos durante o crescimento da bactéria ou a utilização de concentrações sub-ótimas de biotina ao longo do crescimento celular (Kanzaki, Isobe, Okazaki, Motizuki, & Fukuda, 1967; Kimura, 2002; Lubitz & Wendisch, 2016; Nara, Samejima, & Kinoshita, 1964).

Ao longo dos anos, estudos evidenciaram a importância do complexo proteico oxoglutarato desidrogenase (ODHC, EC: 1.2.4.2) sobre o mecanismo de superprodução de glutamato por *C. glutamicum* (Boulay et al., 2010; J. Kim et al., 2009; Shimizu et al., 2003). Localizada em um importante ponto da via metabólica para a produção de aminoácidos, na integração do ciclo do ácido cítrico com a produção de glutamato, foi provado que a atividade específica de ODHC está suprimida quando é estimulada a produção de ácido glutâmico (Shimizu et al., 2003). A ODHC é um complexo proteico constituído por três subunidades diferentes: E1 (codificado por *odhA*), E2 (codificado por *aceF*) e E3 (Codificado por *lpd*). (J. Kim

et al., 2009). Em *C. glutamicum* esse complexo é encontrado na forma de um supercomplexo, interagindo com o complexo da piruvato desidrogenase (PDH), compartilhando as subunidades E2 e E3 (Kinugawa et al., 2020).

Além das proteínas constituintes deste complexo, a proteína OdhI (Inibidor de oxoglutarato desidrogenase) executa papel fundamental na regulação do ODHC, a partir do seu estado de fosforilação, ela inibe ou estimula a atividade de ODHC por meio de interação com a subunidade OdhA, também conhecida como E1 (Niebisch, Kabus, Schultz, Weil, & Bott, 2006). A fosforilação da OdhI é feita predominantemente pela enzima *serine/threonine protein kinase* PknG (Niebisch et al., 2006). A fosforilação da Thr14 da OdhI ocasiona uma mudança de conformação, prevenindo a interação do seu domínio *forkhead associated* (FHA) com a ODHC e, conseqüentemente inibe a inibição da enzima (Barthe et al., 2009). Em um estudo recente, foi proposto um modelo para a via de transdução de sinal, resultando na fosforilação de OdhI, a partir da interação de GlnH com glutamato ou aspartato. Foi demonstrado que GlnH é capaz de se ligar a glutamato e aspartato, e foi hipotetizado que esse possa interagir com domínio periplasmático de GlnX na presença desses aminoácidos. Por sua vez, a GlnX foi suposta como responsável pela transferência de sinal periplasmático detectado pela GlnH para a quinase intracelular PknG, por meio de seu domínio intracelular, possivelmente resultando na sua ativação, o que desencadearia a fosforilação da OdhI e diminuição da inibição de ODHC (Sundermeyer et al., 2022). Essa cadeia de sinais foi sugerida como um controle para a adaptação do fluxo metabólico para a assimilação de nitrogênio ou geração de energia, dependendo da disponibilidade de doadores de grupamento amina no ambiente (glutamato ou aspartato) (Sundermeyer et al., 2022).

Outra fonte de regulação da OdhI foi recentemente identificadas pelas suas proteoformas succiniladas e acetiladas, demonstrado que proteoformas succiniladas da OdhI são capazes de afetar a maneira como essa proteína interage com a ODHC (Komine-Abe et al., 2017; Mizuno et al., 2016). Mais recentemente, um estudo conduzido por nosso grupo mostrou a identificação de duas proteoformas

putativas, uma com adição de um grupamento de 70 Da e outra com clivagem do dipeptídeo SD do N-terminal da proteoforma madura da Odhl (Melo et al., 2023).

1.3 *Trypanosoma cruzi*

O *Trypanosoma cruzi* é o agente etiológico da doença de Chagas ou Tripanossomíase americana. Essa doença foi descoberta em 1909 por Carlos Chagas (Rassi, Rassi, & Marcondes de Rezende, 2012). A doença de Chagas afeta atualmente entre 6 e 7 milhões de pessoas no mundo, com ocorrência principalmente na América Latina (WHO, 2022). Apesar disso, essa doença é considerada como endêmica no sul dos EUA e movimentos de migração tem aumentado a preocupação de contaminação com essa doença nos EUA, Europa e Ásia (Moretti, Mortara, & Schenkman, 2020). Essa doença pode ser dividida em fase aguda e fase crônica, a qual normalmente começa dois a três meses após a infecção inicial (Rassi et al., 2012). Na fase crônica, normalmente 15 a 20 anos após a infecção, cerca de 30% dos infectados desenvolvem a forma cardíaca da doença de Chagas, caracterizada por bradiarritmias, taquicardias, falhas cardíacas, entre outros, podendo chegar a morte repentina. Além da forma cardíaca, também é possível desenvolver a forma digestiva da doença de Chagas crônica, descrita por alterações no motor secretor e funções de absorção dos esôfago e sistema gastrointestinal (Rassi et al., 2012).

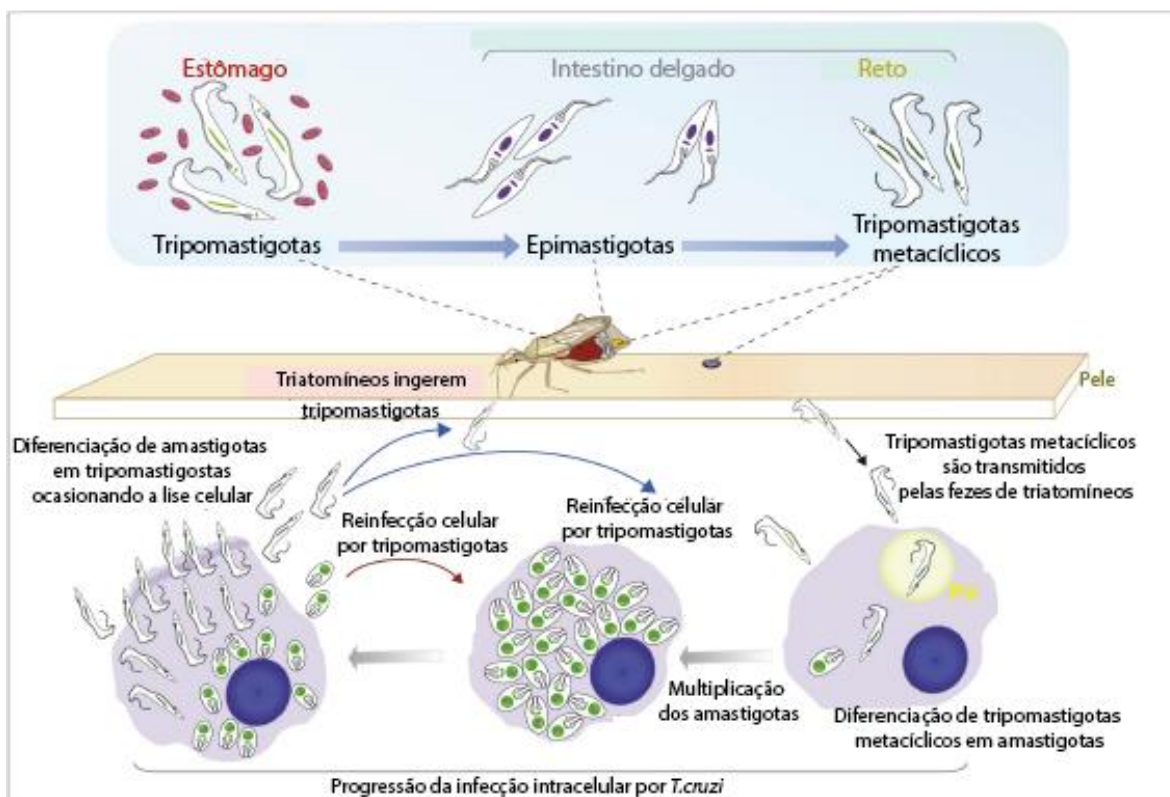
O *T. cruzi* é um protozoário da ordem Kinetoplastida, família Trypanosomatidae. Ele é caracterizado pela presença de um kinetoplasto, região específica da mitocôndria onde está concentrado o DNA mitocondrial (kDNA). O kinetoplasto está localizado próximo ao corpo basal do flagelo desse protozoário. Além de ser transmitido para humanos, o *T. cruzi* também é capaz de infectar diversos outros mamíferos. Sua transmissão comumente ocorre por meio de vetores artrópodes sugadores de sangue, vetores comuns são: *Triatoma infestans*, *Triatoma brasiliensis*, *Rhodnius prolixus*, e *Triatoma dimidiata*. A transmissão de *T. cruzi* também pode ser ocasionada pelo consumo de alimentos contaminados com

esse parasito, além de transfusões e transmissão congênita (Rassi et al., 2012). Tendo em vista o espectro de infecção desse protozoário, é esperado que esse apresente um ciclo de vida complexo. Dessa forma, o *T. cruzi* possui quatro estágios de vida, dois no inseto vetor (epimastigota e tripomastigota metacíclico) e dois no hospedeiro mamífero (amastigota e tripomastigota sanguíneo). As formas amastigotas e epimastigotas são replicativas, enquanto as tripomastigotas são infectivas. Apesar disso, já foram descritas formas amastigotas capazes de infecção (Herrerros-Cabello, Callejas-Hernández, Gironès, & Fresno, 2020).

Durante seu ciclo de vida, o protozoário é adquirido por um vetor hematófago através da sucção do sangue do hospedeiro, na forma tripomastigota sanguíneo. No intestino médio do inseto os parasitas se diferenciam para a forma epimastigota. Os epimastigotas se multiplicam no vetor por meio de fissão binária e então uma pequena porção desses parasitas se diferencia em tripomastigotas metacíclicos (metaciclogênese) no reto do inseto.

Em um novo repasto sanguíneo, o vetor libera os tripomastigotas metacíclicos nas fezes que são capazes de infectar um novo hospedeiro mamífero via mucosa ou ferimentos na pele. Após a infecção de células nucleadas do hospedeiro (como macrófagos, fibroblastos ou células epiteliais), os tripomastigotas se diferenciam em formas amastigotas (amastigogênese). No citosol eles começam a se multiplicar por divisão binária e diferenciar em tripomastigotas sanguíneos, ocasionando no rompimento da célula do hospedeiro e liberação dos parasitas no meio extracelular (Moretti et al., 2020; W. Souza, 2005) (figura 4).

Diferenciação de formas não-infectivas em formas infectivas



Trends in Parasitology

Figura 4. Ciclo de vida do parasito *T. cruzi*. Descrição das quatro formas de vida do *T. cruzi* e suas diferenciações, replicações e infecções. Adaptado de Moretti et al., 2020.

Após a amastigogênese, os amastigotas permanecem sem mudanças morfológicas por 35 h, quando começam a fissão binária. Apesar desse longo tempo sem mudanças morfológicas, foi demonstrado que ocorre a síntese de DNA. Durante a interfase, a cromatina se condensa em áreas próximas da periferia da membrana nuclear. No início da divisão celular de *T. cruzi* a cromatina e o nucléolo se dispersam ao longo do núcleo, e desaparecem. Em seguida, microtúbulos se revelam dentro do núcleo, e esse começa a assumir uma forma mais oval. Então ocorre a divisão do kinetoplasto, e o alongamento do núcleo continua. Após esse alongamento, ocorre a sua divisão por constrição, e ao final o material genético nuclear reaparece e é reorganizado para a localização inicial, ao mesmo tempo em que os microtúbulos desaparecem. O núcleo permanece intacto durante toda a divisão celular de *T. cruzi*, caracterizando uma mitose fechada (W. Souza, 2005).

O *T. cruzi* apresenta alta complexidade em seus mecanismos e estruturas moleculares. A começar pelo seu genoma, cerca de 50% dele é composto de regiões repetitivas não codificantes, e as regiões codificantes podem ser divididas em dois compartimentos: central (caracterizado principalmente pela presença de genes muito conservados com funções conhecidas) e disruptivo, que é caracterizado pela presença de genes que evoluem constantemente, como os genes de famílias multigênicas, incluindo as trans-sialidases, MASPs e as mucinas (Herrerros-Cabello et al., 2020). Ademais, os genes de trypanosomas não possuem regiões promotoras reguladoras, o que faz com que a regulação da expressão seja feita principalmente em nível pós-transcricional. Sobre a transcrição desses parasitos, ela é feita de maneira policistrônica, e os transcritos policistrônicos são processados para gerarem mRNA maduros (Herrerros-Cabello et al., 2020).

Os mecanismos moleculares do ciclo celular de *T. cruzi* foram investigados durante a divisão celular da forma epimastigota, revelando grandes variações na abundância de proteínas e fosfopeptídeos durante a fase de síntese (S). Essas mudanças estavam principalmente relacionadas com redução da abundância de proteínas ribossomais. Além disso, também foram observadas várias *heat shock proteins* (HSPs) e duas tubulinas com mesmo padrão de regulação, abundância diminuída na fase S. Em contrapartida, duas proteínas transportadoras ABC, subunidades da proteína motor dineína, proteína centrina 4, proteína de membrana de kinetoplastídeo e uma trans-sialidase apresentaram maior abundância na fase S. (Santos Júnior et al., 2021). Ao analisar as proteínas fosforiladas durante o ciclo celular, também foi possível observar fosfopeptídeos relacionados aos metabolismos de piruvato e carbono com redução de abundância na fase S, enquanto fosfopeptídeos relacionados com proteínas ligantes de nucleotídeos estavam com maior intensidade nessa mesma fase (Santos Júnior et al., 2021). Esse estudo também mostrou a baixa correlação entre a variação da abundância de proteínas e seus correspondentes transcritos comparando com um estudo transcriptômico prévio do ciclo celular de *T. cruzi* (Chávez et al., 2017; Santos Júnior et al., 2021). Os mecanismos moleculares descritos nesse estudo do ciclo celular

de *T. cruzi* demonstram a relevância do estudo proteômico para compreensão desse parasito.

Capítulo II. Justificativa, Objetivo e Metas

2.1 Justificativa

A proteômica é um campo de pesquisa bem estabelecido usado para a caracterização de processos biológicos em diversos organismos, comumente por meio de análises quantitativas entre condições e identificações de PTMs. Dentro desse contexto, a abordagem proteômica *top-down* tem avançado em diversos aspectos recentemente, em campos como o preparo de amostra, fracionamento de proteínas, espectrômetros de massa otimizados para análise de proteínas intactas, *softwares* para análise de dados obtidos por MS, e até mesmo por padronizações em todas essas etapas (Donnelly et al., 2019a; Melby et al., 2021; Nickerson et al., 2021; Schaffer et al., 2019). Conhecendo-se as vantagens e limitações das abordagens proteômicas *bottom-up* e *top-down*, a integração de ambas tem demonstrado capacidade de caracterização de proteínas de maneira individual (Gault et al., 2014; Ntai et al., 2016) e mais recentemente tem sido explorada de forma *high-throughput* e automatizada (Lima et al., 2020; Ntai et al., 2016; Schaffer et al., 2020).

A bactéria *C. glutamicum* é uma importante plataforma industrial, principalmente na produção de aminoácidos, além de possuir proximidade filogenética à outras bactérias de relevância clínica. O estudo de proteínas e suas PTMs nessa bactéria já revelou diferentes mecanismos de regulação metabólica de importância para a produção de aminoácidos (Boulahya et al., 2010; Komine-Abe et al., 2017; Niebisch et al., 2006). Por outro lado, o protozoário *T. cruzi* é o agente causador da doença de Chagas, enfermidade que afeta entre 6 e 7 milhões de pessoas no mundo (WHO, 2022). Esse parasito possui grande complexidade em seus mecanismos de regulação de abundância de proteínas, que ocorrem, predominantemente, de maneira pós-transcricional (Clayton, 2019). Além disso, a relevância da análise de proteínas e suas PTMs nesse organismo foi demonstrada em diversos estudos (Queiroz et al., 2014; Santos Júnior et al., 2021).

Considerando a relevância biotecnológica e médica de *C. glutamicum* e *T. cruzi*, respectivamente, bem como seus complexos mecanismos de regulação demonstrados em níveis proteicos, a análise proteômica por meio de integração de

abordagens *bottom-up* e *top-down* apresenta grande potencial para a descoberta de novas PTMs e proteoformas que possam afetar processos biológicos relevantes desses microrganismos. Além disso, a implementação da abordagem proteômica *top-down* no laboratório em que foi desenvolvido esse projeto possibilita estudos futuros de proteoformas com diferentes questões e organismos.

2.2 Objetivo e metas

Objetivo geral

Aumentar a compreensão sobre o processo de produção de glutamato por *C. glutamicum*, e o mecanismo de replicação celular em *T. cruzi*, por meio de análises proteômicas *bottom-up* quantitativa e multi-enzimática, bem como qualitativas por abordagem *top-down*.

Metas

- Implementar análise de proteínas intactas por abordagem proteômica *top-down* utilizando a bactéria *C. glutamicum* como modelo.
- Caracterizar proteoformas de *C. glutamicum* utilizando análise proteômica *top-down*.
- Comparar cultivos de *C. glutamicum* nas condições com ou sem estímulo para produção de glutamato por análises proteômicas: *bottom-up* quantitativa; e *top-down* qualitativa.
- Comparar diferentes cultivos da forma de vida amastigota (intracelular, e extracelulares replicativas ou não replicativas) de *T. cruzi*, usando análise proteômica quantitativa por abordagem *bottom-up*.
- Executar análise de proteínas intactas de *T. cruzi* por abordagem *top-down*.
- Identificar PTMs de *T. cruzi* e *C. glutamicum* por meio da união dos métodos *enhanced Global Post-Translational Modification Discovery* (G-PTM-D) e abordagem multi-enzimática.
- Caracterizar as proteoformas de *C. glutamicum* e *T. cruzi* utilizando a integração dos dados das abordagens proteômicas *bottom-up* e *top-down*.

Capítulo III. Metodologia Comum

Nesse capítulo são abordados descritos os materiais e métodos utilizados em comum ao longo dessa tese, principalmente aqueles diretamente relacionados à análise proteômica por abordagem *bottom-up* e *top-down*. Sendo assim, as informações metodológicas das condições de cultivo e extração de proteínas são apresentadas nos respectivos estudos de *C. glutamicum* e *T. cruzi*, focos dos capítulos II e III. Ademais, análises complementares de cada estudo também são descritas nesses capítulos, como: Análise de aminoácidos, análise de microscopia de fluorescência, entre outros.

3.1 Abordagem *bottom-up*

3.1.1 Digestão para análises quantitativas

O protocolo de digestão e remoção do SDS foi seguido de acordo com o padrão e otimizações do *filter-aided sample preparation* (FASP) (J. R. Wiśniewski, 2017; Jacek R. Wiśniewski, 2018; Jacek R. Wiśniewski, Zougman, Nagaraj, & Mann, 2009). Inicialmente, alíquotas de 100 µg das amostras foram reduzidas pela adição de DTT para concentração final de 0,1 M e incubação a 56 °C por 1 h. Em seguida, após refrigerar as amostras para temperatura ambiente, foi adicionado 200 µL de TU (8 M ureia em 0,1 M Tris–HCl, pH 8,5). Então as amostras foram aplicadas em filtros Microcon de 30 KDa (Milipore) e centrifugadas a 10.000 x g por 10 min (todas as centrifugações subsequentes usaram as mesmas condições, a não ser que seja mencionado o contrário). Após a centrifugação foi adicionado 100 µL de solução IAA (Iodocetamida 0.05M em TU) e essas foram incubadas no escuro por 20 min a 21 °C para alquilar as cisteínas reduzidas. Ao final dessa incubação, foram feitas duas lavagens com 150 µL de TU por meio de duas centrifugações e em seguida foram feitas duas lavagens com tampão DT (0,05 M NH₄HCO₃, pH 7,8). Após as lavagens com tampão DT, 1 µg de tripsina foi adicionado a 95 µL de tampão DT e esse foi adicionado ao filtro, que foi incubado a 37 °C por 16 h. Ao final da incubação, a solução foi centrifugada para passagem dos peptídeos digeridos, e foi adicionado mais 45 µL de tampão DT e a centrifugação foi repetida para extrair qualquer peptídeo remanescente. A suspensão de peptídeos resultante foi acidificada pela

adição de TFA para concentração final de 1 %. Ao final, os digestos trípticos foram submetidos à processo de dessalinização por *StageTips* de C18 de acordo com Rappsilber, et al. (Rappsilber, Mann, & Ishihama, 2007), e detalhadamente descrito no tópico “Dessalinização”.

3.1.2 Digestão multi-enzimática

Para a análise multi-enzimática de *C. glutamicum* foi feita um *pool* das condições analisadas por LC-MS/MS, enquanto para a análise de *T. cruzi*, a digestão foi realizada com apenas uma réplica. A digestão por proteases de ambas amostras foi realizada da mesma maneira. Alíquotas de 100 µg de proteínas foram submetidas separadamente a digestão pelas seguintes proteases: tripsina, quimiotripsina, Lys-C, Glu-C, Asp-N, e Arg-C, de acordo com Miller et al. (Miller et al., 2019). O protocolo de digestão e remoção do SDS foi seguido de acordo com o padrão e otimizações do *filter-aided sample preparation* (FASP) (J. R. Wiśniewski, 2017; Jacek R. Wiśniewski, 2018; Jacek R. Wiśniewski et al., 2009) até o último passo, como descrito no tópico anterior (Digestão para análises quantitativas). Ao final da lavagem com tampão DT, foram seguidos o protocolo para análise de com múltiplas proteases (Miller et al., 2019). Brevemente, foram feitas mais duas lavagens com 150 µL dos tampões correspondentes a cada enzima, e as proteases foram adicionadas em 95 µL de seus respectivos tampões, aplicadas aos seus filtros pré-lavados, e incubadas por tempo e temperatura adequados a cada enzima (tabela 1). Ao final da incubação, os filtros foram centrifugados, e o residual de peptídeos foi filtrado com mais 45 µL dos tampões de cada enzima e adicionado à solução total de peptídeos. Ao final da filtração foi adicionado TFA para concentração final de 1 % para parar a digestão enzimática e preparar para a dessalinização. Além disso, os peptídeos/proteínas retidos no filtro foram recuperados de acordo com orientações do fabricante. Rapidamente, foi adicionado 45 µL de tampão das respectivas enzimas, esses dispositivos foram sonicados em banho de ultrassom por 3 min, e passaram por spin com o filtro invertido.

Tabela 1 Proteases e condições de incubação para digestão. Adaptado de Miller et al. 2019 (Miller et al., 2019).

Protease	Proporção (µg) Proteína:Enzima	Tampão de digestão	Temperatura (°C)	Tempo (h)
		Incubação: 50 mM Tris-HCl (pH 7.6), 5 mM CaCl ₂ , 2 mM EDTA		
Arg-C	100:1	Ativação: 50 mM Tris-HCl (pH 7.6), 50 mM DTT, 2 mM EDTA	37	16
		Combinar Incubação:Ativação (9:1)		
Asp-N	100:1	50 mM Fosfato de sódio (pH 8.0)	25	16
Quimotripsina	100:1	100 mM Tris-HCl (pH 8.0), 10 mM CaCl ₂	25	12
Glu-C	100:1	25 mM bicarbonato de amônio (pH 7.8)	25	16
Lys-C	100:1	25 mM Tris-HCl (pH 8.5), 1 mM EDTA, 4 M uréia	37	16
Tripsina	50:1	50 mM bicarbonato de amônio (pH 7.8)	37	16

Os peptídeos filtrados resultantes da digestão em FASP foram então submetidos a dessalinização por *StageTips* de C18 de acordo com Rappsilber, et al. (Rappsilber et al., 2007), e detalhadamente descrito no tópico “Dessalinização”.

Análise de peptídeos de digestão multi-enzimática por MALDI-TOF

Os peptídeos resultantes de cada digestão foram quantificados por Qubit, e uma alíquota de 5 µL de cada amostra foi separada e dessalinizada por ZipTip® C18 de acordo com as instruções do fabricante. As alíquotas dessalinizadas foram utilizadas para avaliar a digestão e o perfil de peptídeos gerado por cada enzima, usando espectrômetro de massas MALDI-TOF/TOF, Autoflex II (Bruker Daltonics, Billerica, Massachusetts, EUA) em modo refletivo. Brevemente, os peptídeos dessalinizados foram plaqueados em poços de placa do tipo AnchorChip, e em seguida foi adicionado 1 µL de matriz HCCA (α-Cyano-4-hydroxycinnamic acid 10 mg/mL).

Fracionamento de peptídeos por fase reversa em pH básico

Os peptídeos dessalinizados por *StageTips* foram solubilizados em solução de fracionamento A (formiato de amônio 100 mM, pH 10, ACN 2%) e submetidos a fracionamento em microcolunas com disco de SDB e resinas POROS R2 e OLIGO R3 (figura 5). Esse protocolo foi inspirado e adaptado a partir de vários estudos que utilizam microcolunas para fracionamento de fase reversa em pH básico (Han, Jin, Woo, Min, & Kim, 2014; Han et al., 2013; H. Kim et al., 2019; Ruprecht, Zecha, Zolg, & Kuster, 2017). As principais diferenças no protocolo desenvolvido e utilizado nesse estudo foram: 1) Utilização da resina OLIGO R3 que apresenta características mais hidrofóbicas do que a POROS R2 e assim potencial para melhor retenção de peptídeos hidrofílicos em condição básica; 2) Uso de disco SDB para retenção de resinas, o que impede interações indesejadas com resinas tradicionais de C18 devido ao pH; e 3) padronização do volume de resina utilizado. O fracionamento em pH básico (HpH – *High pH*) foi feito em microcolunas com *slurry* de POROS R2 e OLIGO R3 retidos por um disco de SDB (figura 5). Essas microcolunas foram feitas com um disco para SPE de SDB (Empore™), o qual foi excisado com seringa de gauge 16 a partir de membranas e inserido em ponteiras de 200 µL *low-binding* (Axygen). Após o acoplamento do disco, a ponteira foi centrifugada a 5.000 x g por 3 min para compactar e uniformizar a posição do disco. Nas ponteiras com discos foram adicionados 10 µL de *slurry* de OLIGO R3, seguido de centrifugação a 5.000 x g por 3 min. Em seguida foram adicionados 20 µL do *slurry* de POROS R2, e as microcolunas foram novamente centrifugadas a 5.000 x g por 3 min. Ambos *slurrys* foram feitos com a ressuspendendo as resinas em metanol.

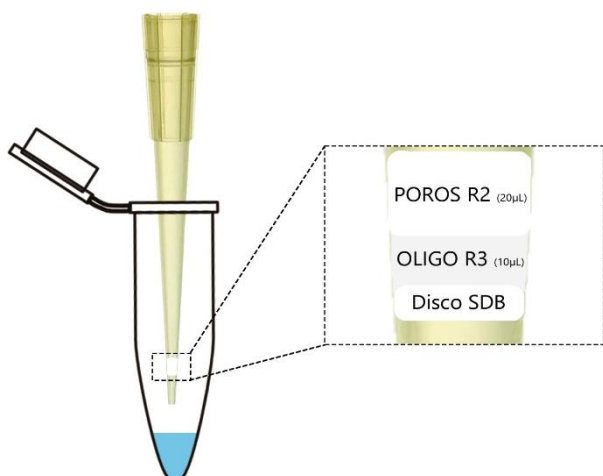


Figura 5. Esquema ilustrado de microcoluna para fracionamento em pH básico. As cores são meramente ilustrativas.

As resinas das microcolunas preparadas foram condicionadas pela passagem de 100 µL de metanol, seguida pela passagem de 100 µL de ACN. Todas as passagens de soluções pela coluna foram realizadas por centrifugação a 1500 x g por 2 min. Após condicionamento, as microcolunas foram equilibradas com 100 µL de solução de fracionamento A. Em seguida, as amostras de peptídeos dessalinizados por *StageTips* ressuspensas em solução de fracionamento A foram carregadas nas microcolunas e fracionadas em 20 frações por um gradiente de ACN em formiato de amônio 10 mM (pH = 10) (tabela 2). Ao final do fracionamento as 20 frações foram unidas em 7 frações finais (tabela 2) e secadas em *SpeedVac*.

Tabela 2. Soluções de fracionamento em pH básico.

	Acetonitrila (%)	Formiato de amônio (mM)	Fração	Fração final
SA	2	10	FT	F0
E1	5	10	1	F1
E2	7.5	10	2	F2
E3	10	10	3	F3
E4	12.5	10	4	F4
E5	15	10	5	F5

E6	17.5	10	6	F6
E7	20	10	7	F1
E8	22.5	10	8	F2
E9	25	10	9	F3
E10	27.5	10	10	F4
E11	30	10	11	F5
E12	32.5	10	12	F6
E13	35	10	13	F1
E14	40	10	14	F2
E15	50	10	15	F3
E16	60	10	16	F4
E17	70	10	17	F5
E18	80	10	18	F6
SW	100	10	W	F0

3.1.3 Dessalinização

Os peptídeos filtrados resultantes da digestão em FASP foram então submetidos a dessalinização em *StageTips* de C18 de acordo com Rappsilber, et al. (Rappsilber et al., 2007). Brevemente, foram construídas microcolunas em ponteiras de 200 µL *low-binding* (Axygen) usando seis discos de C18 extraídos com agulhas de gauge 16 de membranas para SPE de C18 (Empore™). Essas microcolunas foram condicionadas por duas lavagens com 100 µL de metanol, e duas com 100 µL solução de dessalinização D (Acetonitrila (ACN) 80%, ácido acético 0,5 %). Após condicionamento as amostras foram carregadas nas microcolunas, e lavadas duas vezes usando 100 µL de ácido acético 0,5 %. Após a lavagem as amostras foram eluídas em 100 µL usando *steps* de acetonitrila (25%-50%-80%-100%) em ácido acético 0,5%.

3.1.4 LC-MS/MS

As análises por LC-MS/MS foram realizadas em Nano-UHPLC (Ultra high performance liquid chromatography) Dionex Ultimate 3000 (Thermo Fisher Scientific™) acoplado a um espectrômetro de massas LTQ-Orbitrap Elite (espectrômetro de massa iontrap-orbitrap híbrido, ThermoFisher Scientific™).

Foram utilizadas pré-colunas (100 μm \times 3 cm) de partículas esféricas de sílica de 5 μm (Reprosil-Pur 120 Å pore C18) (Dr. Maich, Ammerbuch, Germany) e coluna analítica (75 μm \times 25 cm) com resinas C18 de 3 μm Reprosil-Pur 120 Å pore size (Dr. Maich, Ammerbuch, Germany). As fases móveis utilizadas foram: solvente A (0,1 % ácido fórmico) and B (100 % acetonitrila, 0,1 % ácido fórmico) em fluxo constante de 0,250 $\mu\text{L}/\text{min}$. O espectrômetro de massas foi configurado para aquisição DDA (*data-dependent acquisition*), de forma que os 15 íons mais intensos foram submetidos a fragmentação por HCD (Higher collisional energy dissociation), com janela de isolamento de 2,0 m/z e energia de colisão normalizada de 35. A resolução utilizada foi de 120.000 FWHM e 15.000 FWHM para o MS1 e MS2, respectivamente. Especificamente para a análise de peptídeos da abordagem multi-enzimática, foi utilizada uma coluna analítica com as dimensões 75 μm \times 25 cm, e a janela de isolamento utilizada para aquisição de MS2 foi de 3,0 m/z .

3.1.5 Bioinformática

A análise dos dados raw gerados foi realizada no programa MetaMorpheus (Solntsev et al., 2018), que, para identificação de peptídeos utiliza uma versão modificada do Morpheus (Wenger & Coon, 2013), e para quantificação *label-free* usa o FlashLFQ (Millikin, Solntsev, Shortreed, & Smith, 2018). Foram utilizados os proteomas de referência do UniProt de *C. glutamicum* ATCC 13032 (UP000000582) ou *T. cruzi* CL Brener (UP000002296) em formato xml, se acordo com o organismo sendo estudado. Os espectros em formato *raw* foram carregados no programa e foi feita uma calibração com parâmetros default. Em seguida, foram feitas buscas de PTMs não anotadas nos proteomas de referências usando a abordagem G-PTM-D, que buscou as modificações biológicas comuns, artefatos comuns e adutos metálicos nos espectros. Então foi feita a identificação permitindo duas perdas de clivagens, duas modificações por peptídeo, e tamanho mínimo de sete. Além disso, carbamidometilação de cisteína e selenocisteína foram classificadas como modificações fixas, e oxidação de metionina como variável. Parcimônia de proteínas foi aplicada na identificação, requerendo pelo menos dois peptídeos por *protein*

group. A quantificação foi feita baseado em MS1, usando MBR (Match Between Runs), com normalização baseado em mediana. *Peptide Spectrum Matches* (PSMs) foram considerados válidos se o q-valor < 0.01 e a pontuação do MetaMorpheus > 5 . Alternativamente, para a análise multi-enzimática não foram feitas quantificações e a descoberta de PTMs por *enhanced* G-PTM-D considerou a presença de modificações menos comuns (*Less common*), e glicosilações.

Para as análises estatísticas e controle de qualidade de abordagens quantitativas foi utilizada a linguagem R, por meio do RStudio (R Core Team, 2022; RStudio Team, 2021), com auxílio dos pacotes *pmartR* (Stratton et al., 2019), *VSclust* (Schwämmle & Jensen, 2018) e *GAGE* (Luo, Friedman, Shedden, Hankenson, & Woolf, 2009). Brevemente, os dados normalizados pelo MetaMorpheus foram carregados no R, passaram por transformação por Log_2 e tiveram o perfil de distribuição avaliados quanto à eficiência da normalização adotada. Após essa etapa, foi realizada a contagem de *missing values* (NA) e filtragem de proteínas que não foram quantificadas em pelo menos 14 amostras. Em seguida, os dados foram avaliados quanto a presença de *outliers*, utilizando o cálculo de *robust mahalanobis distance* (RMD), que levou em consideração as características número de NA, *skewness*, MAD (*median absolute deviation*), e correlação. Depois da análise de *outliers*, foram analisadas a reprodutibilidade e diferença global das condições por meio de *PPCA* (*Probabilistic Principal Component Analysis*) e *heatmap* de correlação. Ao final foram realizados testes ANOVA com correção de múltiplos testes (FDR), considerando como significativamente diferentes aqueles com um valor de p-valor $< 0,05$. A abundancia das proteínas sem *missing values* foram então submetidas à clusterização por *VSclust* e enriquecimento de vias por *GAGE*.

3.2 Abordagem *top-down*

3.2.1 Preparo de amostra

As proteínas obtidas nos diferentes estudos realizados passaram pelos menos processamentos. Resumidamente, foi realizada quantificação de proteínas a partir da aferição de fluorescência usando Qubit™ (Invitrogen), de acordo com as instruções do fabricante. A partir da quantificação das proteínas extraídas foi realizado um *pool* representativo das condições a serem analisadas em cada estudo. Esses *pools* foram reduzidos, adicionando DL-dithiothreitol (DTT, concentração final = 60 mM), e incubando a 56°C por 1 h, e alquilados, adicionando iodoacetoamida (IAA, concentração final = 182 mM). Para suspender a alquilação, foi adicionado mais DTT (concentração final = 120 mM). Aproximadamente 400 µg de proteínas reduzidas e alquiladas de cada *pool* foram submetidas a fracionamento por GELFrEE (*Gel Eluted Liquid Fractionation Entrapment Electrophoresis*) (Tran & Doucette, 2008). Esse fracionamento foi feito utilizando tampão tris-glicina (0,192 M glicina, 0,025 M Tris e 0,1% SDS), malhas de poliacrilamida (28,5% acrilamida, 1,5 bis-acrilamida) concentradora (5%) de 2 cm, e separadora (12%) de 1,5 cm, sob diferença de potencial constante de 240 V. Após a saída da frente de corrida, aproximadamente 20 min de corrida, a coleta das frações foi iniciada, seguindo 5 coletas a cada 1 min, em seguida 5 coletas a cada 2 min, depois 3 coletas de 5 min, seguidas por 3 coletas espaçadas em 15 min. Cerca de 10 µL das frações foram submetidos à SDS-PAGE e corados por nitrato de prata para visualização da massa molecular (MM) das proteínas de cada fração e seleção das frações submetidas à espectrometria de massas acoplada à cromatografia líquida (LC-MS/MS) (figura 6). As frações selecionadas para LC-MS/MS foram precipitadas por método metanol/clorofórmio/água (4:1:3) (Wessel & Flügge, 1984), por meio de duas lavagens com metanol e secagem em capela laminar por 20 min. Os precipitados de cada fração foram ressuspensos em 30 µL de solução de injeção (0,2% ácido fórmico, 5% acetonitrila), com auxílio de banho de ultrassom por 5 min, usando gelo. Após ressuspender, foram submetidos a centrifugação 20.000 g por 10 min e 15°C. Em seguida, 4 µL de cada fração foi submetido a LC-MS/MS.

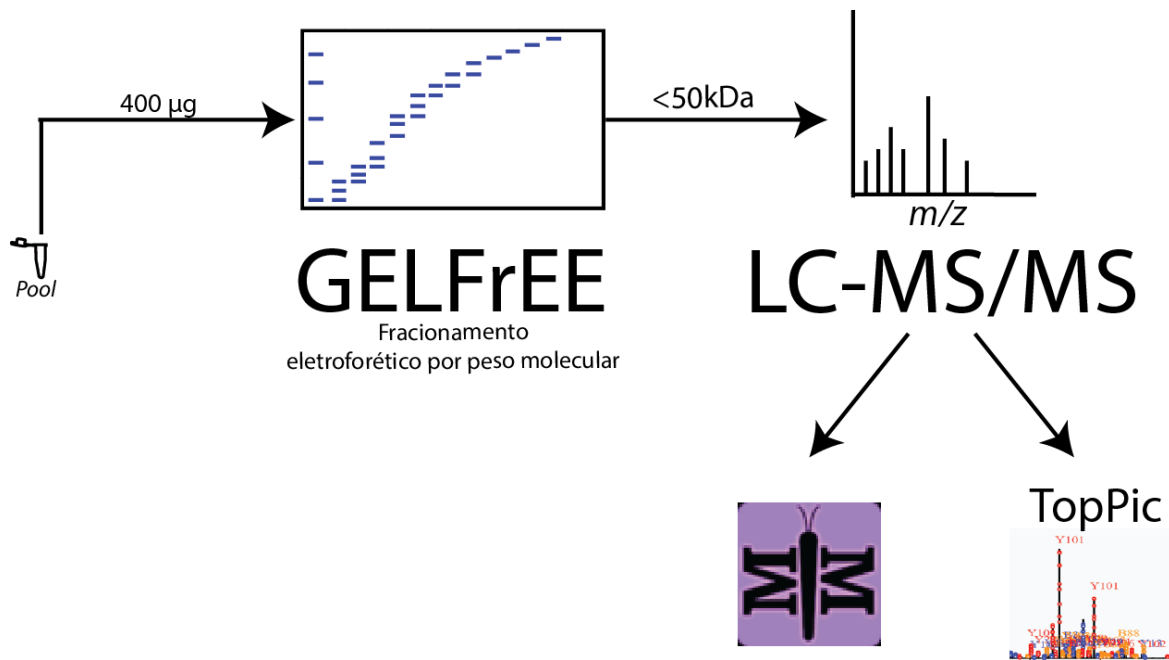


Figura 6. Fluxograma de preparo de amostra e análise de dados de proteômica *top-down*.

3.2.2 LC-MS/MS

Para abordagem *top-down* foram utilizados o mesmo Nano-UHPLC e espectrômetro de massas que na abordagem *bottom-up*. As diferenças foram nos métodos cromatográficos, colunas e método do espectrômetro.

A cromatografia líquida foi realizada utilizando pré-coluna de 2 cm (aproximadamente 200 μ de diâmetro interno), e analítica de 15 cm para separação (aproximadamente 75 μ de diâmetro interno). As fases móveis utilizadas foram A (5 % Acetonitrila, Ácido fórmico 0,2%) e B (0,2% ácido fórmico em acetonitrila 100%) e o fluxo foi de 3 μ L/min durante 10 min para carregamento na pré-coluna e 0,3 μ L/min para eluição durante 90 min. O gradiente para eluição foi 5% B por 10 min, 20% B por 45 min, 55% B por 5 min, 85% B por 20 min e por último 5% B por 10 min. A aquisição em espectrômetro foi feita ao longo dos 90 min de eluição. Foi utilizado *Source-induced dissociation* (SID) = 15 V em ESI (*Electrospray ionisation*). A resolução utilizada foi de 240.000 FWHM e 120.000 FWHM para o MS1 e MS2,

respectivamente. A aquisição do MS2 foi feita em modo *data-dependent* (DDA), fragmentando os dois íons mais intensos por *higher energy collisional dissociation* (HCD), com janela de isolamento de 15 *m/z*. Além disso, foi utilizado uma lista de exclusão dinâmica com duração de 30 s e 3 *m/z* de janela de isolamento.

3.2.3 Análise de dados

Os dados *.raw* de abordagens *top-down* gerados pelo experimento LC-MS/MS foram analisados utilizando dois softwares independentes, MetaMorpheus 1.0.2 (Solntsev et al., 2018) e TopPic Suite 1.6.3 (Kou, Xun, & Liu, 2016) (**Erro! Fonte de referência não encontrada.**). O primeiro é uma ferramenta de identificação de proteínas com pequena tolerância de variação de massa (Δm) para o precursor, entre 10 e 20 ppm, enquanto o segundo utiliza uma abordagem erro tolerante, identificando proteínas com maiores variações de massa, ± 500 Da. Os parâmetros utilizados para identificação no MetaMorpheus foram os seguintes: proteoma de referência *C. glutamicum* ATCC 13032, UP000000582. Tolerância do precursor = 5 ppm, tolerância do produto = 20 ppm, metionina inicial = variável, número máximo de modificações = 2, carbamidometilação como modificação fixa de cisteínas, *score* mínimo para identificação = 5 e *q-value* < 0.01. Para utilizar o programa TopPic, os arquivos *.raw* primeiro foram convertidos em MzML fazendo *centroid* dos espectros, utilizando o algoritmo do fabricante, pelo programa MSconvert (Chambers et al., 2012). Em seguida, os espectros foram deconvoluídos utilizando a ferramenta topFD (Basharat, Zang, Sun, & Liu, 2023) com os seguintes parâmetros: carga máxima de 30; massa máxima de 70.000 Da; MS1 S/N = 3; MS2 S/N = 1; janela de isolamento do precursor = 15 *m/z*, erro *m/z* = 0,02. Os espectros deconvoluídos foram então submetidos a identificação seguindo os seguintes parâmetros: Carga máxima = 30, Massa máxima = 70.000 Da, S/N MS1 = 3.0, S/N MS2 = 1.0, janela de exclusão = 15 *m/z*, erro de *m/z* = 0.02, carbamidometilação como modificação fixa de cisteínas, número máximo de Δm = 1, -300 Da < Δm < 300 Da, formas de N-terminal NONE, NME, NME_Acetylation, M_ACETYLATION, FDR < 0.01 em nível de PrSMs (*Proteoform spectrum matches*).

Para a integração dos dados proteômicos *bottom-up* com dados *top-down*, foram utilizadas as identificações de peptídeos provenientes da abordagem *bottom-up* multi-enzimática e todos os dados *top-down* (figura 17). Essa integração será realizada em duas etapas no programa MetaMorpheus e Proteoform suite (Cesnik et al., 2018). Inicialmente, foram construídos dois bancos de dados a partir dos peptídeos e PTMs identificados pela abordagem *bottom-up* multi-enzimática: 1) uPTM-DB (*unpruned*, Banco com todas as proteínas do proteoma de referência, incluindo as PTMs descobertas pela abordagem *bottom-up* multi-enzimática); e 2) pPTM-DB (*pruned*, Banco apenas com as proteínas que foram identificadas pela abordagem *bottom-up*, incluindo as PTMs descobertas). O banco uPTM-DB foi utilizado na identificação utilizando o programa MetaMorpheus, que faz a identificação das proteoformas a partir de dados de MS/MS. Foram utilizados os mesmos parâmetros descritos para identificação de proteoformas por *top-down*. O banco pPTM-DB foi usado no programa Proteoform Suite para a construção de um banco teórico com a massa das proteoformas possíveis. Os peptídeos, identificados por *bottom-up*, os PrSMs, identificados por *top-down*, e as massas deconvoluídas de todos espectros de *top-down* foram carregadas, juntamente ao banco de dados criado na plataforma Proteoform Suite. Ao final, essa plataforma realiza a identificação baseado na massa exata da proteoforma, informação presente nos espectros deconvoluídos, em comparação com o banco criado pelo próprio programa, gerando um FDR global para a busca.

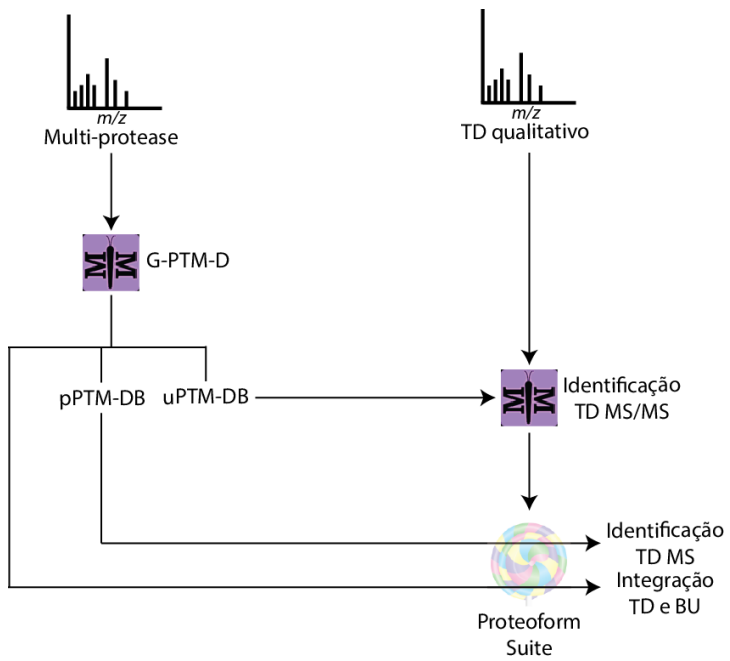


Figura 7. Fluxograma para integração de dados de proteômica *bottom-up* e *top-down*.

Capítulo IV. Proteômica de
Corynebacterium glutamicum

4 Resumo

Corynebacterium glutamicum é uma bactéria de grande relevância industrial, capaz de produzir diversas biomoléculas de interesse comercial. Essa bactéria é considerada uma plataforma industrial para a produção de aminoácidos, segundo grupo de biomoléculas mais produzido por fermentação no mundo, mercado o qual apresentou estimativas para alcançar produção de 10 milhões ton/ano em 2022. O glutamato é o aminoácido mais produzido industrialmente por fermentação. Levando em consideração a relevância industrial de *C. glutamicum* e o potencial da análise proteômica para descrever processos biológicos, foram realizados três estudos com essa bactéria buscando aumentar a compreensão sobre o seu processo de produção de glutamato: 1) Análise *top-down* descritiva das proteoformas de *C. glutamicum*; 2) Análise *bottom-up* quantitativa comparando condição controle vs condição produtora de glutamato; 3) Integração de proteômica *bottom-up* multi-enzimática e *top-down* para caracterização de proteoformas da bactéria *C. glutamicum*. Esses estudos resultaram na descoberta de modificações pós-traducionais (PTMs) e proteoformas com possíveis impactos biotecnológicos, como as proteoformas clivada e com possível butirilação de Odhl, assim como a proteoforma clivada da metalloendopeptidase mepB, sugerindo mecanismo de ativação por clivagem dessa enzima. Além disso, foram identificadas duas proteínas sem função caracterizada relacionadas ao processo inicial de produção de glutamato induzida por penicilina-G. Também foram identificadas proteoformas exclusivas na condição produtora de glutamato, como a proteoforma metilada da subunidade pequena da acetolactato sintase, sugerindo possível mecanismo de inibição dessa via por essa PTM durante produção de glutamato. Por fim, a integração de proteômica *bottom-up* e *top-down* resultou no aumento do número de identificações por ferramenta de busca com tolerância restrita, no entanto o número de proteoformas identificadas por abordagem tolerante a erro continuou maior.

4.1 Análise *top-down* descritiva das proteoformas de *C. glutamicum*

A análise proteômica qualitativa descritiva de *C. glutamicum* possibilitou a descoberta de proteoformas ainda então não reportadas em proteínas relacionadas à processos biológicos de relevância biotecnológica. Essas descobertas foram publicadas recentemente e estão descritas no artigo “*Revealing Corynebacterium glutamicum proteoforms through top-down proteomics*” (Melo et al., 2023), fornecido na sessão anexos (Anexo 1: Artigo “*Revealing Corynebacterium glutamicum proteoforms through top-down proteomics*”). Além disso, os resultados gerados tiveram impacto na divulgação científica com reportagens divulgadas pela Secretária de Comunicação da Universidade de Brasília (<https://www.unbciencia.unb.br/biologicas/717-metodo-inovador-na-analise-de-proteinas-em-bacteria-mantem-unb-na-vanguarda-cientifica>) e jornal Metrôpoles (<https://www.metropoles.com/saude/estudo-inedito-da-unb-pode-ter-impacto-na-industria-alimenticia-entenda>).

4.2 Análise *bottom-up* quantitativa comparando condição controle vs condição produtora de glutamato

4.2.1 Objetivos

Estudo anteriores já demonstraram a indução da produção de glutamato por *C. glutamicum* através da adição de antibióticos ou detergentes ao meio de cultura (J. Kim, Hirasawa, Saito, Furusawa, & Shimizu, 2011; Kimura et al., 1999; Lubitz & Wendisch, 2016). Dentre eles destacam-se antibióticos como a penicilina-G e detergentes como o Tween-40. Sendo assim, o presente estudo visou a análise de parâmetros biológicos de *C. glutamicum*, como crescimento, produção de aminoácidos decorrentes de estímulo por Penicilina-G ou tween-40 e seus efeitos no proteoma dessa bactéria.

4.2.2 Métodos

Cultivo e cepas

A bactéria utilizada nesse estudo foi a cepa *C. glutamicum* ATCC 13032 adquirida da fundação André Tosello (fat.org.br/colec_culturas/, Campinas, SP).

O cultivo de *C. glutamicum* foi realizado em meio Triptona soja ágar (TSA - 17 g/L digesto pancreático de caseína, 3 g/L digesto pancreático de soja, 5 g/L NaCl, 2,5 g/L K₂HPO₄, 2,5 g/L glicose monohidratada, 1,5 g/L ágar), quando em meio sólido, e meio de cultura CGXII líquido ((NH₄)₂SO₄ 20 g/L; ureia 5 g/L; KH₂PO₄ 1 g/L; K₂HPO₄ 1 g/L; MgSO₄.7H₂O 10 mg/L; Ácido 3-mofolinopropano sulfônico 42 g/L; CaCl₂ 10 mg/L; FeSO₄.7H₂O 10 mg/L; MnSO₄.H₂O 10 mg/L; ZnSO₄.7H₂O 1 mg/L; CuSO₄ 0,2 mg/L; NiCl₂.6H₂O 0,02 mg/L; biotina 0,2 mg/L; glicose 40 g/L; ácido protocatecuico 0,03 mg/L) (Keilhauer, Eggeling, & Sahm, 1993). Todos os cultivos em meios líquidos foram realizados em volumes de 50 mL de meio de cultura sob agitação em erlenmeyers de 500 mL com defletores (*baffled* erlenmeyer).

Para iniciar o cultivo, estoques de *C. glutamicum* armazenados em -80 °C foram inoculados em meio sólido e cultivados por aproximadamente 12 h a 30 °C. A partir desse cultivo foi coletada uma colônia isolada que foi inoculada em meio CGXII e cultivada a 30 °C, sob 120 rpm, por 17 h (pré-inóculo). Depois disso, foi aferida a densidade óptica do cultivo em 600 nm (OD₆₀₀), utilizando espectrofotômetro leitor de microplacas CLARIOstar (BMG LABTECH, Ortenberg, Alemanha). A partir da OD₆₀₀ foi calculado o volume necessário para inóculo com OD₆₀₀ de 1,0 no cultivo principal. A partir do inóculo, o meio de cultura foi incubado a 30 °C, sob 120 rpm por 36 h.

O cultivo foi realizado em três condições adaptadas a partir de estudos anteriores (J. Kim et al., 2011): 1) Controle (sem estímulo para produção de glutamato); 2) T-40 (com estímulo para produção de glutamato a partir da adição de tween-40 em concentração final de 4 mg/mL); 3) Pen-G (com estímulo para produção de glutamato a partir da adição de penicilina-G em concentração final de

8 µg/mL). Os estímulos para a produção de glutamato foram adicionados 6,75 h após o inóculo. Para acompanhar o crescimento bacteriano foi realizada uma curva de crescimento a partir de coletas 6,75 h, 9 h, 12 h, 18 h, 24 h e 36 h após o inóculo. Para análises proteômicas e quantificação de aminoácidos livres foram coletados 1 mL de cultura 6,75 h, 9 h, 24 h e 36 h após o inóculo. As amostras coletadas foram submetidas a centrifugação a 12.000 x g por 10 min. O sobrenadante resultante da centrifugação e o *pellet* foram submetidos, separadamente, a congelamento em nitrogênio líquido e armazenados em -80 °C. O sobrenadante foi então direcionado para análise de aminoácidos, enquanto o pellet foi reservado para análises proteômicas (figura 8).

Os experimentos de curva de crescimento e análises de aminoácidos foram realizados em quadruplicata biológica, enquanto para a proteômica foi realizada a análise por triplicata biológica.

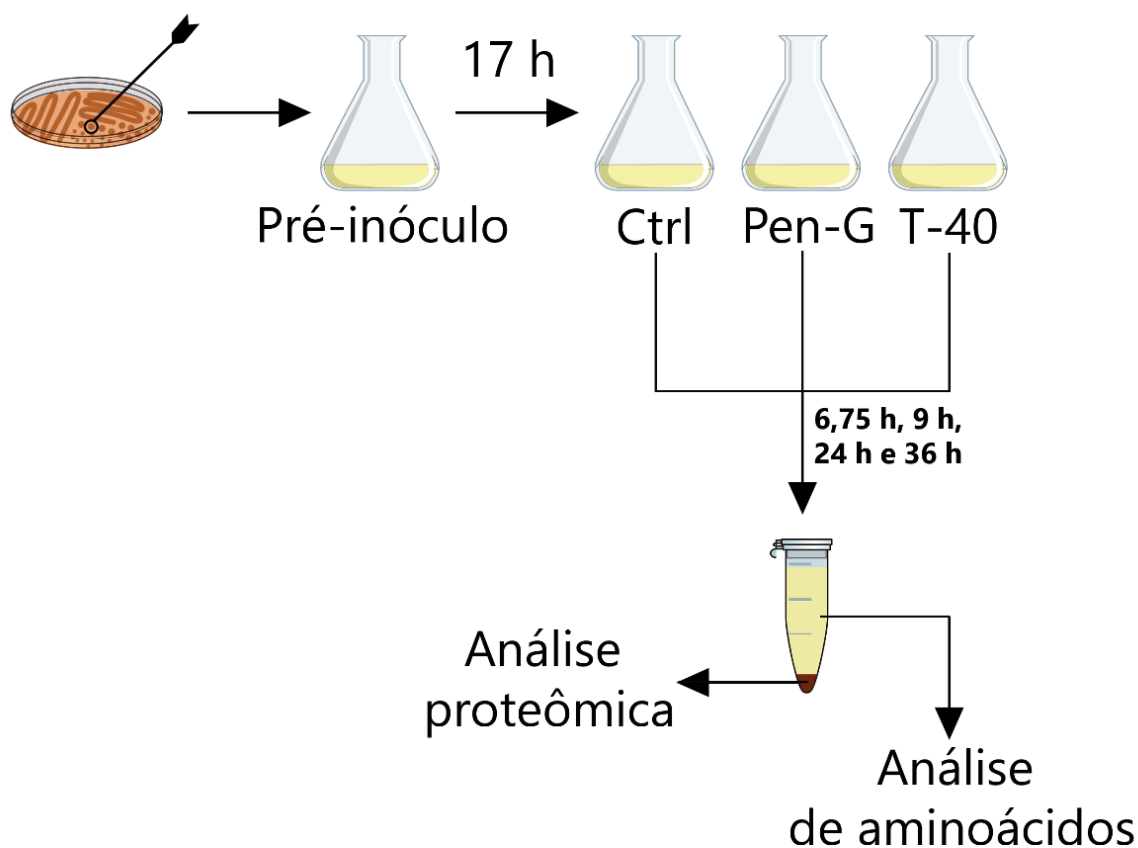


Figura 8. Fluxograma do cultivo de *C. glutamicum* até a aquisição de amostras para análises proteômica e de aminoácidos. Ctrl = Controle, Pen-G = Indução por penicilina-G (8 µg/mL), T-40 = Indução por tween-40 (4 mg/mL).

Análise de aminoácidos

Os sobrenadantes do cultivo das condições controle, T-40 e Pen-G (figura 8) foram submetidos à precipitação com metanol/clorofórmio/água (proporção 4:1:3) (Wessel & Flügge, 1984). Brevemente, foram adicionados 4 volumes de metanol à amostra, seguido de adição de 1 volume de clorofórmio e 3 volumes de água Milli-Q com agitação em *vortex* de 30 s após cada adição. A solução resultante foi submetida a centrifugação à 21.000 x *g* por 15 min, e a fase polar resultante foi coletada, secada e ressuspensa em 1 mL de água Milli-Q. Essa solução foi então submetida à remoção de contaminantes por SPE (*solid-phase extraction*).

Resumidamente, o pH dessa solução foi ajustado para aproximadamente 3,0 e ela foi submetida a troca catiônica utilizando *bed* de 2 mL de resina *strongly acidic dowex 50w x8-200*. Inicialmente a amostra foi ligada a essa resina, disposta em seringa de 5 mL, a qual foi posteriormente lavada com 4 volumes de água Milli-Q, e compostos eletricamente carregados foram eluídos com 3,5 volumes de hidróxido de amônio 6 M. O eluído foi então secado, ressuspendido em 1 mL de água Milli-Q e submetido à quantificação e determinação dos aminoácidos livres em cromatógrafo de troca iônica de alta eficiência acoplado a detectores amperométrico pulsátil (*High Performance Anion-Exchange with Pulsed Amperometric Detection – HPAE-PAD*) ICS-5000 (Dionex) com coluna cromatográfica AminoPAC PA10 de 2 mm X 250 mm. A análise dos dados gerados pelo ICS-5000 foi feita utilizando o programa Chromeleon (Thermo Scientific™).

Extração e quantificação de proteínas

As amostras para proteômica foram preparadas a partir das células de *C. glutamicum* obtidas de cultivo em meio CGXII, de acordo com o citado na sessão “Cultivo e cepas”. Os *pellets* obtidos por centrifugação foram ressuspendidos em 750 µL de tampão desnaturante (Tris-HCl 0,1 M; SDS 4 %, DTT 0,1 M; pH = 7,6) (Jacek R. Wiśniewski et al., 2009), suplementado com inibidor de proteases Roche *cOmplete ULTRA EDTA-free* (ROCHE) e inibidor de fosfatases PhosSTOP (ROCHE). Em seguida foram adicionados 250 µL de *beads* de sílica de 0,1 mm, e as células foram submetidas a *bead beating* por 10 min utilizando Mini-Beadbeater-16 (BioSpec Products, Bartlesville, EUA). Após esse processo, a mistura foi centrifugada a 21.000 x *g* por 15 min para separar as proteínas solúveis da fração insolúvel e das *beads* e armazenadas para quantificação de proteínas. A quantificação de proteínas foi feita a partir da aferição de fluorescência usando Qubit™ (Invitrogen), de acordo com as instruções do fabricante.

Após a quantificação de proteínas, essas amostras foram separadas para análise do perfil de proteínas por SDS-PAGE, e Análise *bottom-up* quantitativa comparativa entre as condições Pen-G e Controle.

Análise do perfil de proteínas por SDS-PAGE

O preparo da amostra para análise por SDS-PAGE consistiu na diluição de 25 µg de proteína de cada réplica biológica em 3 volumes de tampão de amostra Laemmli 4x (Tris-HCl 0,25 M, SDS 8 %, Glicerol 20 %, DTT 0,1 M, azul de bromofenol 0,01%) (Laemmli, 1970) e aquecimento a 95 °C por 10 min. Após isso, a amostra passou por eletroforese desnaturante em gel de poliacrilamida com malha concentradora de 5 % e separadora gradiente de 8 % a 18 %. Foram feitas análises comparando as três condições 24 h após o inóculo e análises temporais de cada condição 6,75, 9, e 24 h após o inóculo. A análise em SDS-PAGE comparando as condições foi realizada com as amostras coletadas após 24 h do inóculo para submeter as células a um tempo considerável de incubação sobre influência dos estímulos para a produção de glutamato, aumentando as chances de visualizar diferenças entre as condições na abundância de bandas de proteínas no gel.

Análise proteômica

Toda a metodologia de análise proteômica utilizada nesse estudo foi como o descrito no capítulo III dessa tese (Metodologia Comum).

4.2.3 Resultados e discussão

Cultivo e crescimento bacteriano

Os cultivos de *C. glutamicum* em meio CGXII foram acompanhados ao longo de 36 h após o inóculo, com OD₆₀₀ inicial = 1,0. Como descrito nos métodos, foram analisadas três condições: Controle, Pen-G (induzida por Penicilina-G 8 µM), e T-40 (induzida por tween-40 4 mg/mL). O crescimento dessas condições apresentou diferenças após a adição de Penicilina-G e Tween-40, de forma que os cultivos em condição controle alcançaram OD₆₀₀ maior que 40, enquanto a condição com tween-40 chegou a 30 de OD₆₀₀ após 12 h, mas reduziu estabilizando próximo de 25. Já a condição Pen-G foi a que apresentou menor crescimento total, estabilizando a OD₆₀₀ próximo de 20 após 12 h (figura 9). Como já foi demonstrado anteriormente, condições que aumentam a produção de glutamato em *C. glutamicum* ocasionam

em uma redução do crescimento bacteriano durante a fase exponencial (J. Kim et al., 2011; Kimura et al., 1999).

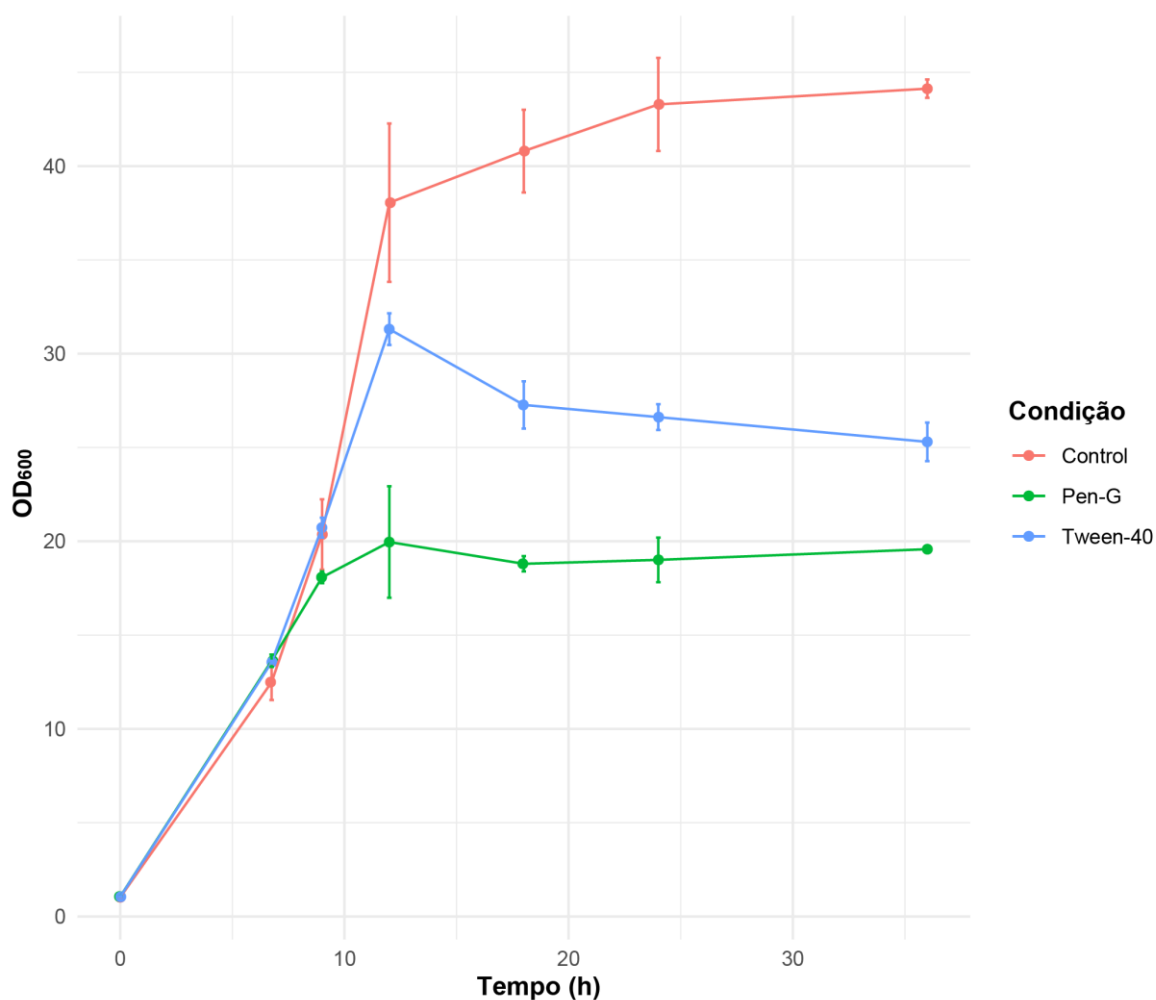


Figura 9. Curva de crescimento de *C. glutamicum* ATCC 13032 em condição controle, e com indução para produção de glutamato por Penicilina-G (Pen-G) e tween-40. A curva foi feita a partir de dados de OD₆₀₀ e plotada de acordo com a média e desvio padrão das quatro réplicas biológicas de cada condição.

Análise de aminoácidos

Como esperado, foi detectado glutamato em alta concentração nas condições T-40 e Pen-G, enquanto esse aminoácido sequer foi detectado no cultivo da condição controle (figura 10 C). Após 2,75 h de estímulo (9h de cultivo), a concentração de glutamato mostrou diferença significativa entre as condições Pen-

G ($0,63 \pm 0,05$ g/L) e T-40 ($0,03 \pm 0,01$ g/L) (figura 10 C). Hirasawa, et al. descreveram que a produção de glutamato é iniciada 1 h após a indução por penicilina e continua por pelo menos 10 h (Hirasawa, Saito, Yoshikawa, Furusawa, & Shmizu, 2018). Após 24 h de cultivo, a concentração de glutamato em ambas as condições aumentou significativamente, representando concentração de $5,03 \pm 0,94$ g/L na condição Pen-G e $1,23 \pm 0,33$ g/L na condição T-40 (figura 10 C). Além disso, nesse horário, a concentração de glutamato é significativamente maior na condição Pen-G do que em T-40 (figura 10 C). Em seguida, após 36 h de cultivo, a concentração de glutamato apresentou tendência de redução na condição Pen-G, no entanto, essa diferença não foi significativa. Enquanto isso, na condição T-40, a concentração desse aminoácido pareceu estável (figura 10 C).

A treonina e a arginina foram quantificadas nas três condições (Controle, Pen-G e T-40) nos tempos 6,75 h e 9 h. A concentração desses dois aminoácidos aumentou significativamente após 9 h de cultivo, em todas as condições (figura 10 A, B). Ademais, o tratamento com Penicilina G afetou a concentração de arginina após 9 h de cultivo, reduzindo significativamente a concentração desse aminoácido quando comparado com as condições T-40 e controle (figura 10 B).

A arginina utiliza o glutamato como um de seus precursores, sendo que já foi demonstrado que o aumento na produção de glutamato, pela diminuição na expressão de *odhA* ocasionou no diminuição da produção de arginina (Sheng et al., 2021). Além disso, a maior disponibilidade intracelular de glutamato, causada pela deleção do gene *Ncgl1221*, responsável pela expressão da proteína de excreção do glutamato, também ocasionou o aumento da produção de arginina em cepa modificada (Park et al., 2014). Por fim, análises transcriptômicas revelaram redução da expressão de diversos genes da biossíntese da arginina após o estímulo para produção de glutamato por penicilina G (Hirasawa et al., 2018). Esses estudos apontam a importância do glutamato para a produção de arginina. Isso vai ao encontro da redução da secreção de arginina na condição Pen-G e aumento da presença de glutamato no meio de cultura. Uma possível explicação para esse

resultado seria que o glutamato que seria utilizado como precursor para a produção de arginina passou a ser secretado, diminuindo produção e secreção de arginina.

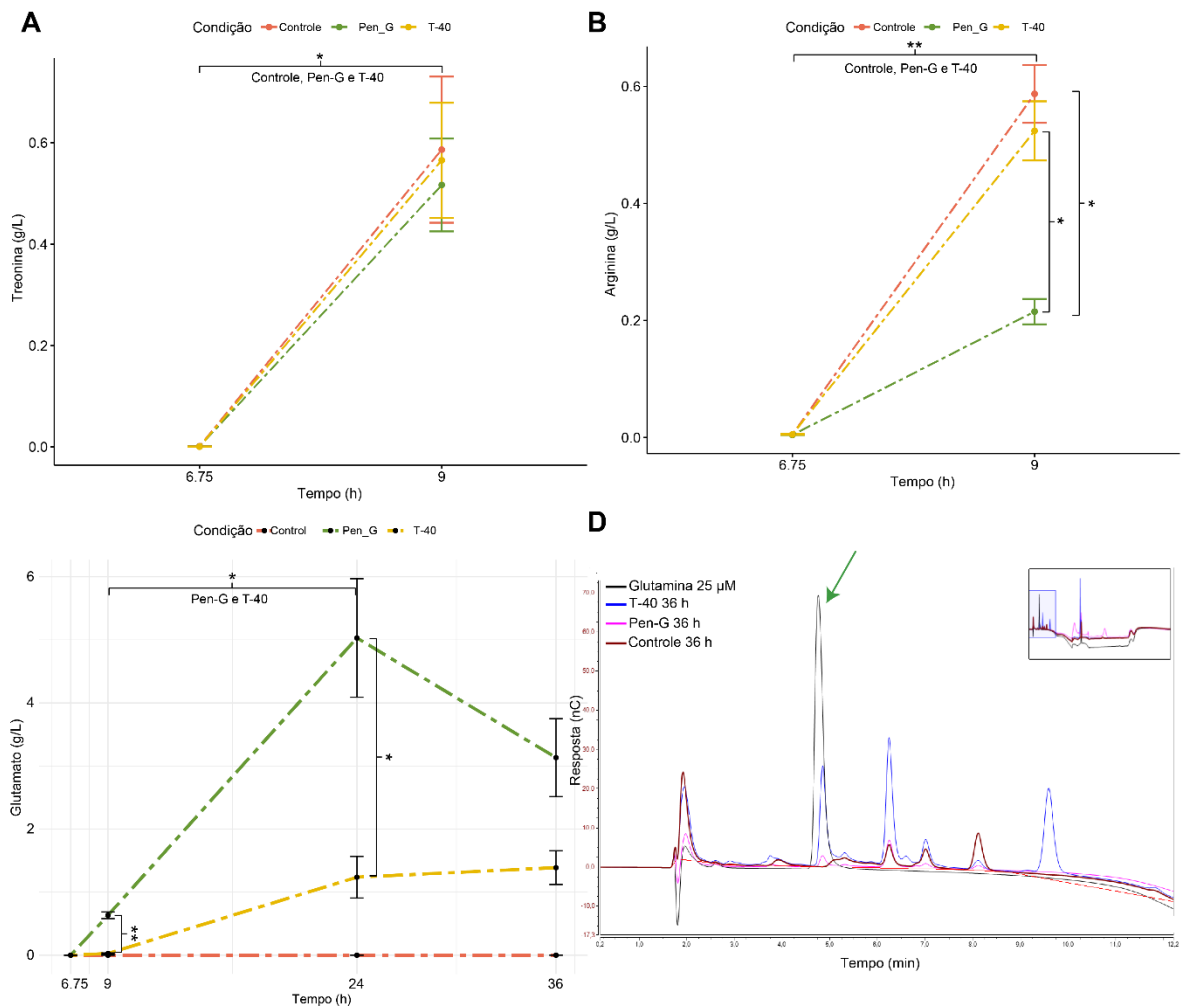


Figura 10. Análise de aminoácidos quantificados e detectados no cultivo de *C. glutamicum*. A) Concentração de treonina 6,75 h e 9 h após o inóculo. B) Concentração de arginina 6,75 h e 9 h após o inóculo. C) Concentração de glutamato nas condições Pen-G e T-40 9 h, 24 h e 36 h após o inoculo. D) Cromatograma das condições T-40, Pen-G, e controle após 36 h de cultivo sobrepostos a um padrão de glutamina com 25 μM (seta verde). Foi utilizado o teste-t para comparar a concentração dos aminoácidos entre as condições e tempos. As comparações são indicadas pelas barras entre as duas condições analisadas pelo teste-t, e as diferenças significativas estão representadas pelos “*”. * = p-valor ≤ 0,05; ** = p-valor ≤ 0,01.

Além dos aminoácidos quantificados foi possível detectar a presença de um pico cromatográfico que, de acordo com o tempo de retenção, indicava a presença de glutamina. Para verificar essa presença, foram realizadas corridas com um

padrão de glutamina, confirmando essa suspeita. Esse aminoácido foi detectado nas condições com estímulo para a produção de glutamato (Pen-G e T-40) 24 h e 36 h após o inoculo (figura 10 D). No entanto, são necessárias novas corridas para calcular a curva de calibração desse aminoácido e realizar a sua quantificação nessas amostras.

Os aminoácidos aspartato, alanina e prolina foram quantificáveis apenas em algumas condições, dificultando suas comparações. Mais especificamente, o aspartato foi detectado nas condições T-40 após 36 h de cultivo ($0,025 \pm 0,007$ g/L) e Pen-G após 9 h de cultivo ($0,007 \pm 0,002$ g/L). A alanina foi detectada nas condições T-40 ($0,19 \pm 0,01$ g/L) e controle ($0,05 \pm 0,03$ g/L) após 36 h e 24 h de cultivo, respectivamente. A prolina só foi detectada na condição T-40 após 36 h de cultivo ($0,11 \pm 0,02$ g/L).

Considerando a maior produção de glutamato pela condição Pen-G (figura 10 C), as condições com esse estímulo (6 h, 9 h, e 24 h) foram as selecionadas para serem analisadas e comparadas com a condição controle pelas abordagens proteômicas por LC-MS/MS.

Análise de perfil proteico por SDS-PAGE

A eficiência do rompimento celular pelo método de *bead beating* com SDS foi inicialmente avaliada pela quantificação de proteínas totais, utilizando método de quantificação por Qubit (tabela 3). Como pode ser observado na tabela, a quantidade total aproximada de proteínas extraídas foi sempre maior que 4 mg, variando entre 4,03 mg e 9,00 mg. Essa diferença possivelmente foi ocasionada pela discrepância na quantidade de células, como pode ser observado na figura 9. Além disso, foram observados desvios padrões pequenos relativos às suas condições, indicando boa reprodutibilidade dessa metodologia.

Tabela 3. Quantidade média de proteínas extraída em cada coleta (valores em miligramas). Foram calculadas as médias e desvio padrão da quantidade de proteínas total extraída em cada coleta de acordo com suas condições e tempo de cultivo.

Condição	Tempo de cultivo			
	6,75 h	9 h	24 h	36 h
Penicilina G	4,36 ± 0,51	5,99 ± 0,28	6,71 ± 0,42	6,71 ± 0,20
Tween 40	4,43 ± 0,05	5,70 ± 0,24	7,79 ± 0,28	9,00 ± 0,09
Controle	4,03 ± 0,41	5,38 ± 0,48	6,68 ± 0,42	6,32 ± 0,45

Por meio de análise de SDS-PAGE foi possível verificar a integridade de todas as amostras (ausência de padrões de degradação), e o perfil de bandas das réplicas biológicas também indicou boa reprodutibilidade das condições biológicas e do método de lise (figura 11). Foram observadas bandas de proteínas com abundância diferente ao longo do tempo de cultivo da condição controle (figura 11 A). Mais especificamente, dois grupos de bandas de alta massa, entre 180 kDa e 245 kDa, e um grupo de proteínas de baixa massa (abaixo de 11 kDa) apresentaram bandas mais intensas após 24 h de cultivo. Em contrapartida, foi observado um grupo de bandas próximo de 100 kDa e outro maior que 35 kDa com menor intensidade após 24 h de cultivo (figura 11 A). Curiosamente, alguns desses padrões foram observados comparando todas as condições após 24 h de cultivo (figura 11 B). As bandas de alta e de baixa massa identificadas com intensidades

diferentes na análise temporal foram observadas com maior tonalidade na condição controle em relação aos estímulos (T-40 e Pen-G) (figura 11 B). Além disso, dois grupos de bandas entre 75 kDa e 135 kDa apresentaram perfil que sugere aumento de intensidade nas condições com estímulo (figura 11 B). Congruente com a baixa intensidade da banda de baixa massa (abaixo de 11 kDa) nas condições estimuladas para a produção de glutamato após 24 h de cultivo (figura 11 B), não foi possível observar essa banda em nenhuma dos tempos de cultivo da condição Pen-G (figura 11 C) ou da condição T-40 (figura 11 D). Ademais, foi possível observar o aumento da intensidade de dois grupos de bandas após ambos os estímulos (figura 11 C, D).

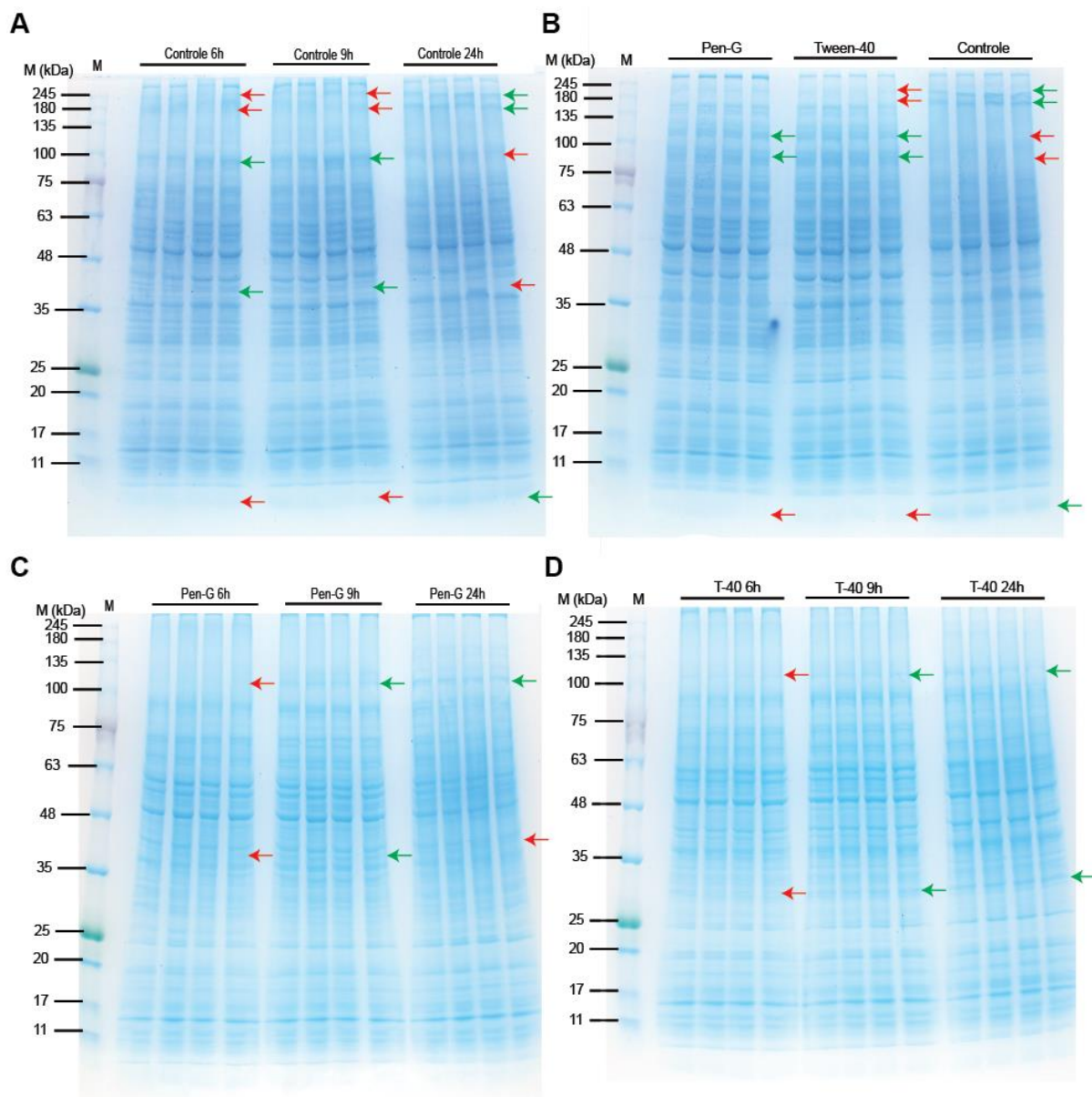


Figura 11. SDS-PAGE de quatro réplicas biológicas das diferentes condições analisadas. Foi realizada eletroforese em malha de poliacrilamida, utilizando gel concentrador de 5% e separador gradiente de 8% a 18%, utilizando o marcador de peso molecular pré-corado BLUeye (Sigma-Aldrich), indicado pela letra "M" com as massas moleculares descritas em kDa. Em cada poço foi adicionado as proteínas intracelulares extraídas de cada réplica biológica. Bandas com intensidade de cor diferente entre as condições estão indicadas pelas setas verdes nas condições que sugerem maior abundância, e vermelha nas condições com menor abundancia. A) Análise temporal da condição Controle. B) Análise das condições, Pen-G, T-40 e controle após 24 h de cultivo. C) Análise temporal da condição Pen-G. D) Análise temporal da condição T-40.

Análise proteômica

Foram identificadas 1029 proteínas com pelo menos 2 peptídeos únicos (Tabela suplementar 1). As medidas de mediana, quartis e limites de todas as amostras analisadas se encontram próximas (figura 12 A), sugerindo reprodutibilidade das amostras e normalização eficiente dos dados. A grande maioria das proteínas foi quantificada em todas as condições, aproximadamente 97 % (figura 12 B), indicando boas condições de cromatografia durante as análises por LC-MS/MS e competência do algoritmo de MBR (*match between runs*) para propagação da identificação das proteínas. Os parâmetros de MAD (*median absolute deviation*), assimetria e correlação das corridas foram utilizados para avaliar a presença de *outliers* extremos nas amostras por meio da escala RMD (*Robust Mahalanobis Distance*). Essa análise não apresentou nenhuma das amostras como *outlier* extremo, considerando p-valor de $1e-4$ (figura 12 C). A distribuição de abundância das proteínas apresentou distribuição com característica *bell-shaped* (figura 12 D), sugerindo aplicabilidade de testes paramétricos para as análises estatísticas.

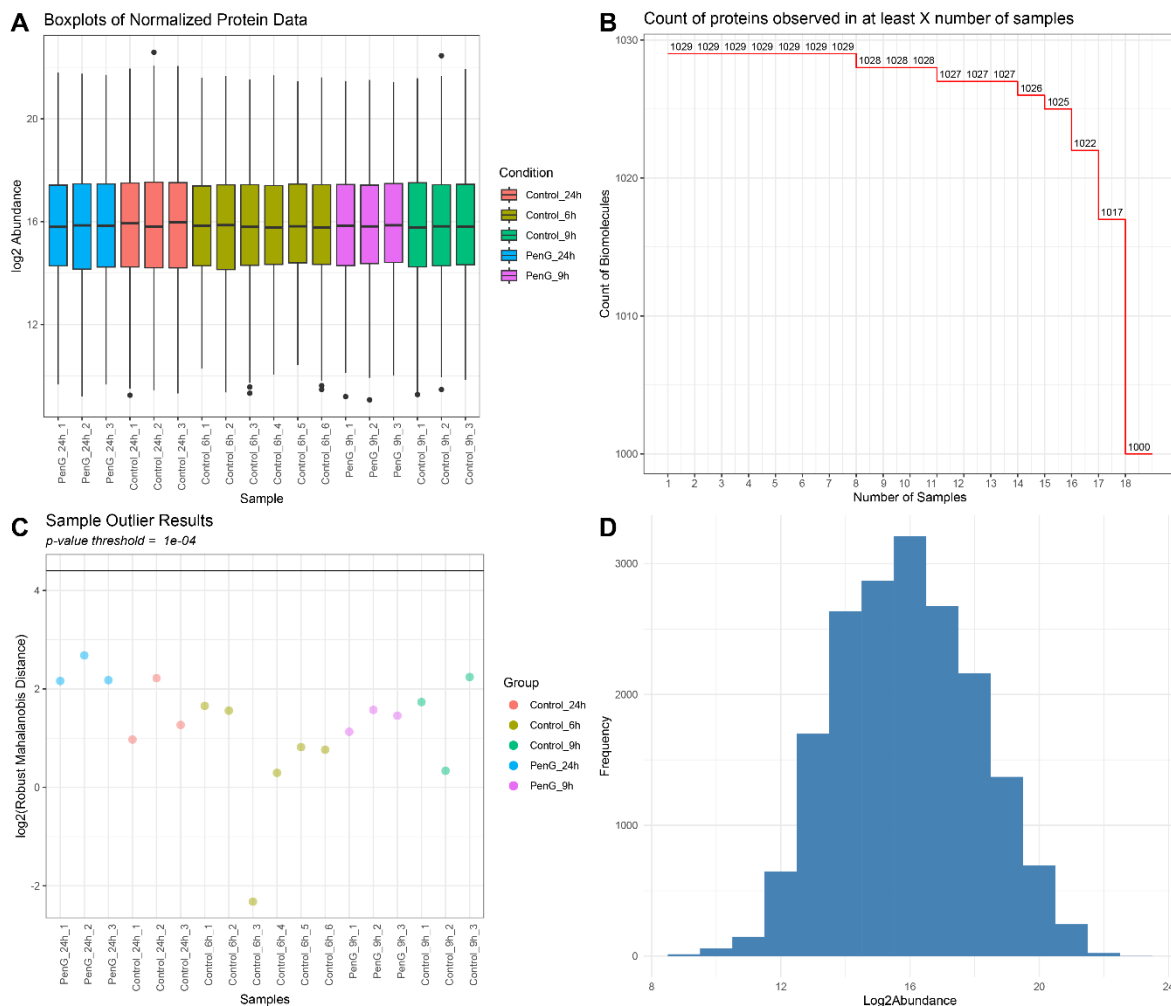


Figura 12. Análise de parâmetros globais de quantificação e identificação das proteínas. A) Boxplot da abundância das proteínas. B) Número de proteínas observadas em pelo menos x amostras. C) Análise de outliers por Robust Mahalanobis Distance (RMD). D) Distribuição das abundâncias de todas as proteínas quantificadas em todas as condições.

Considerando que a indução por penicilina-G ocorreu após 6,75 h de cultivo, o tempo 6h não apresenta diferença entre as condições, sendo agrupado todo como um controle temporal anterior ao estímulo. Sendo assim, foram feitas três comparações entre as condições analisadas: Controle_6h Vs PenG_9h; Controle 9h Vs PenG_9h; Controle_24h Vs PenG_24 h.

Em análise preliminar por PPCA (*probabilistic principal component analysis*) (figura 13 A), as condições de 6 h e 9 h apresentavam poucas diferenças, enquanto a diferença entre Controle e PenG após 24 h de cultivo foi mais significativa, como

evidenciado pela análise estatística por ANOVA, seguida de correção para múltiplos testes por FDR (figura 13 B). Essas análises sugerem que a maior parte das diferenças quantitativas causadas pela indução para produção de glutamato ocorrem após a fase logarítmica. No entanto, foram identificadas 3 proteínas reguladas na comparação PenG_9h vs Controle_9h, sugerindo modulação quantitativa no período inicial após a indução para produção de glutamato. Além disso, 25 proteínas estavam reguladas na comparação PenG_9h vs Controle_6h (figura 13 B).

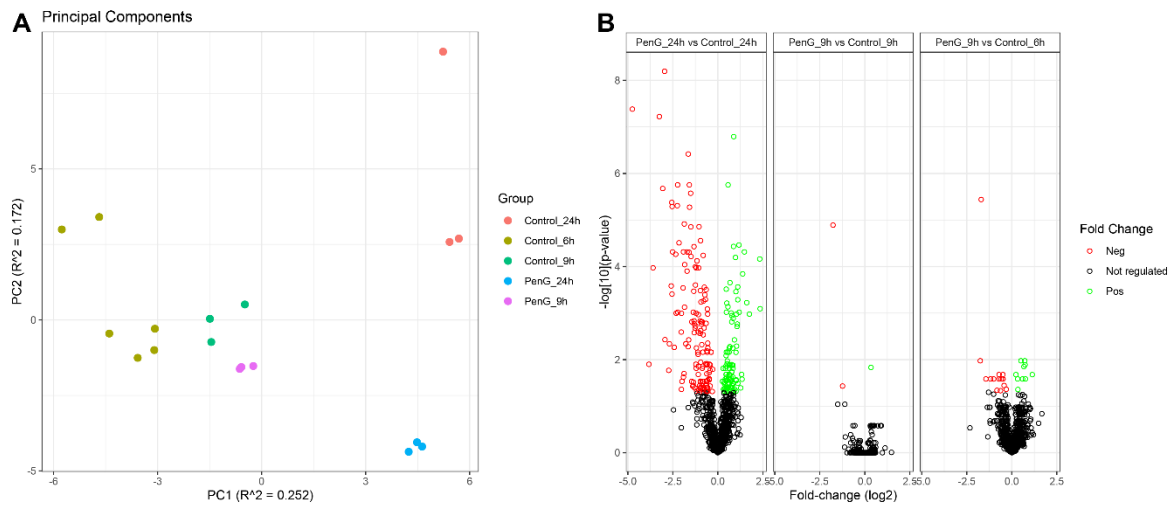


Figura 13. Análise das diferenças entre as condições analisadas. A) *Probabilistic Principal Component Analysis* (PPCA) das abundâncias e número de *missing values* das proteínas analisadas em cada réplica das diferentes condições. B) *Volcano plot* das comparações analisadas nesse estudo.

Apenas duas proteínas com menor abundância na comparação PenG_9h Vs Controle_9h (Q8NQ03/Cgl1651/cg1859 e Q8NSY0/Cgl0535/cg0625), também estão reguladas negativamente na condição PenG_9h comparada com Controle_6h, sugerindo função relevante ou influência no mecanismo inicial da indução de produção de glutamato por Penicilina G. Infelizmente, ambas são descritas como não caracterizadas, apesar disso, a proteína Q8NQ03 foi identificada como secretada e glicosilada, possivelmente por pmt (*protein mannosyltransferase*, Q8NRZ6/Cgl0890/cg1014) (Mahne, Tauch, Puhler, &

Kalinowski, 2006). Por sua vez, a proteína Q8NSY0 foi sugerida de forma elegante, utilizando cepas de mutadas e purificação por gradiente, como proteína da micomembrana de *C. glutamicum* (Marchand et al., 2012). Além disso, apesar de não terem sido mencionados no manuscrito, a expressão dos genes NCgl1588, correspondente ao cg1859/Q8NQ03, e NCgl0513, correspondente à proteína Q8NSY0/cg0625, apresentou tendência de redução ao longo de 360 min após indução para produção de glutamato por adição de penicilina-G (Hirasawa et al., 2018). Apesar de ambas as proteínas apresentarem pouca caracterização, as evidências propostas aqui sugerem que as duas podem ser influenciadas pelo processo de produção de glutamato induzido por penicilina-G.

Por outro lado, a comparação PenG_24h Vs Controle_24h apresentou grande diferença, 144 proteínas com menor abundância na condição PenG_24h e 103 com maior abundância (figura 14). Várias proteínas reguladas estavam relacionadas com processos metabólicos, sugerindo alterações drásticas nesses processos (figura 14). Além disso, proteínas reguladas positivamente na condição PenG_24h foram relacionadas à termos que sugerem aumento da tradução, como: "*Translation*", "*Peptide metabolic process*", "*cellular protein metabolic process*", "*Protein metabolic process*". Já as proteínas reguladas negativamente aparentam estar relacionadas com processos metabólicos para produção de vários aminoácidos, além de apresentar de maneira mais específica a biossíntese e metabolismo do triptofano (figura 14), sugerindo redução na produção de aminoácidos, principalmente o triptofano.

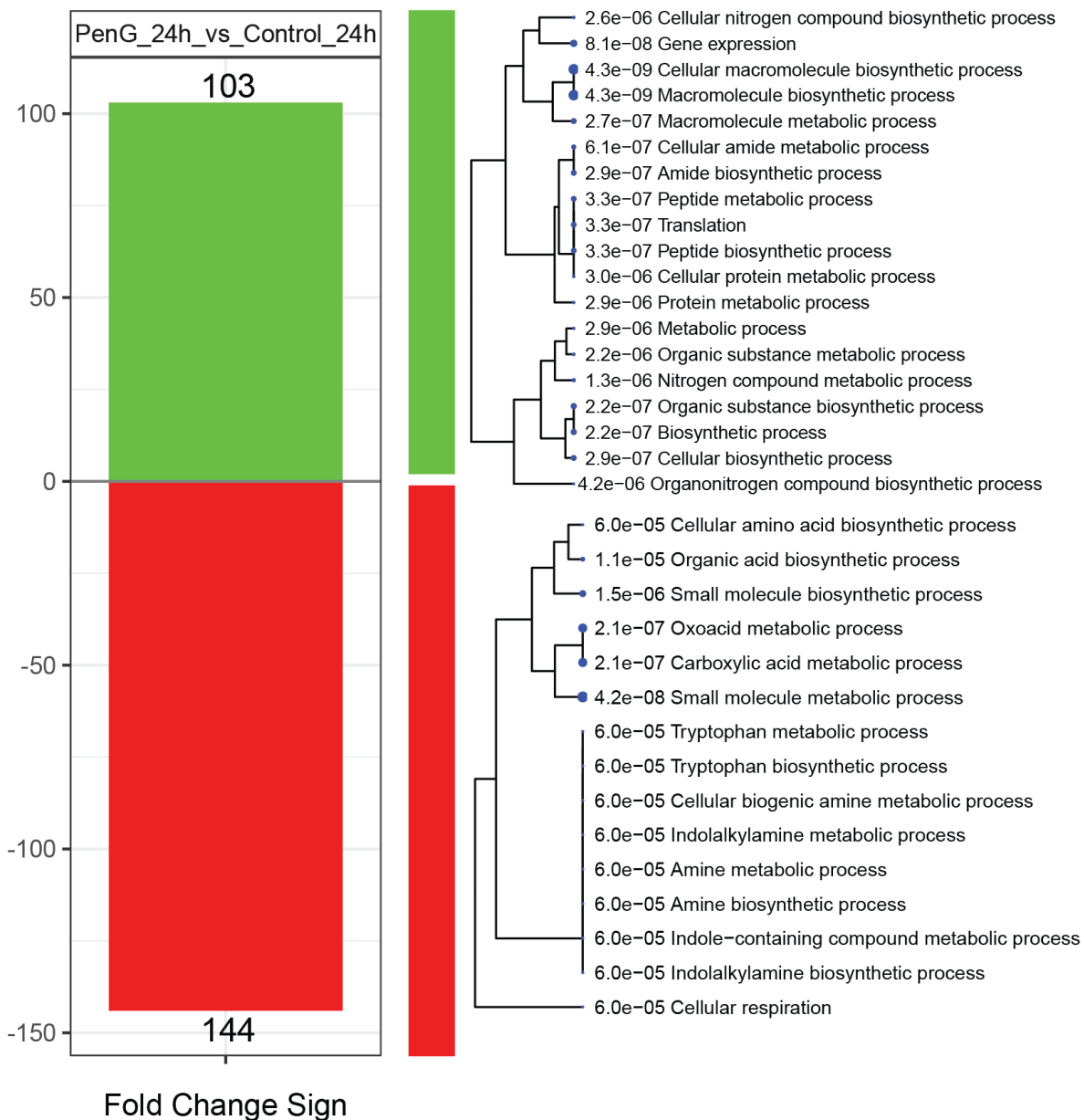


Figura 14. Proteínas reguladas na condição PenG_24h em comparação com Controle_24h e os processos biológicos relacionados. O número de proteínas reguladas positivamente na condição PenG_24h está apresentado em verde, enquanto as negativamente estão em vermelho. Os termos de *Gene Ontology* super-representados para os grupos de proteínas de acordo com sua regulação são apresentados ao lado de cada grupo. Análise de super-representação realizada com software shinyGO (Ge, Jung, & Yao, 2020) utilizando *cutoff* de p-valor = 0,05.

A alteração metabólica causada em diferentes vias, principalmente no metabolismo central do carbono e de aminoácidos foi relatada em nível de transcritos até 10 h após a indução por penicilina-G (Hirasawa et al., 2018). Mais

especificamente, os níveis dos transcritos dos genes *pfk* (NCgl1202), *pgk* (Q01655/NCgl1525), *gpmA* (Q8NTA5/NCgl0390), *eno* (Q8NRS1/NCgl0935) e *pyk* (Q46078/NCgl2008) estavam mais expressos após indução, enquanto os genes do ciclo do ácido cítrico (CAC) *gltA* (P42457/NCgl0795), *acn* (Q8NQ98/NCgl1482), *sucC* (Q8NMK7/NCgl2477), *sucD* (Q8NMK8/NCgl2466), *sdhA* (NCgl0360), *sdhB* (NCgl0361), e *sdhCD* (NCgl0359) estavam sendo menos expressos (Hirasawa et al., 2018). Oito proteínas correspondentes a esses genes descritos foram identificadas nesse estudo, no entanto, apenas a aconitato hidratase A (*acn*, Q8NQ98/NCgl1482) foi identificada com regulada nas comparações, com menor abundância na condição PenG em comparação com controle após 24 h de cultivo (Tabela suplementar 2). Por outro lado, a proteína fosfoenolpiruvato carboxiquinase [GTP] (*pckG*, Q9AEM1/NCgl2765) apresentou um padrão parecido com o observado no estudo transcriptômico, o qual descreveu aumento da expressão 1 h após a indução e redução para valores normais posteriormente. Essa proteína apresentou aumento de abundância 2,75 h após indução por penicilina, e redução na condição PenG_24 h em comparação com a Controle_24h. Uma possibilidade para as diferenças encontradas entre os estudos são as diferentes moléculas analisadas, proteínas e transcritos, mas a abordagem utilizada temporal utilizada no estudo transcriptômico também pode ter resultado nessas diferenças. Para analisar essa possibilidade, as comparações Controle_6h vs Controle_9h e Controle_6h Vs Controle 24h foram realizadas (Tabela suplementar 3), resultando na identificação de 193 proteínas reguladas positivamente e 201 proteínas reguladas negativamente somente pelo tempo de cultivo (figura 15). Congruente com essa hipótese, as proteínas *pgk*, *eno*, e *pyk* estão reguladas positivamente após 24 h de cultivo, enquanto as proteínas *SucC* e *SucD* estão reguladas negativamente (Tabela suplementar 3). De maneira contrária, a *acn* apresentou aumento da abundância após 24 h de cultivo.

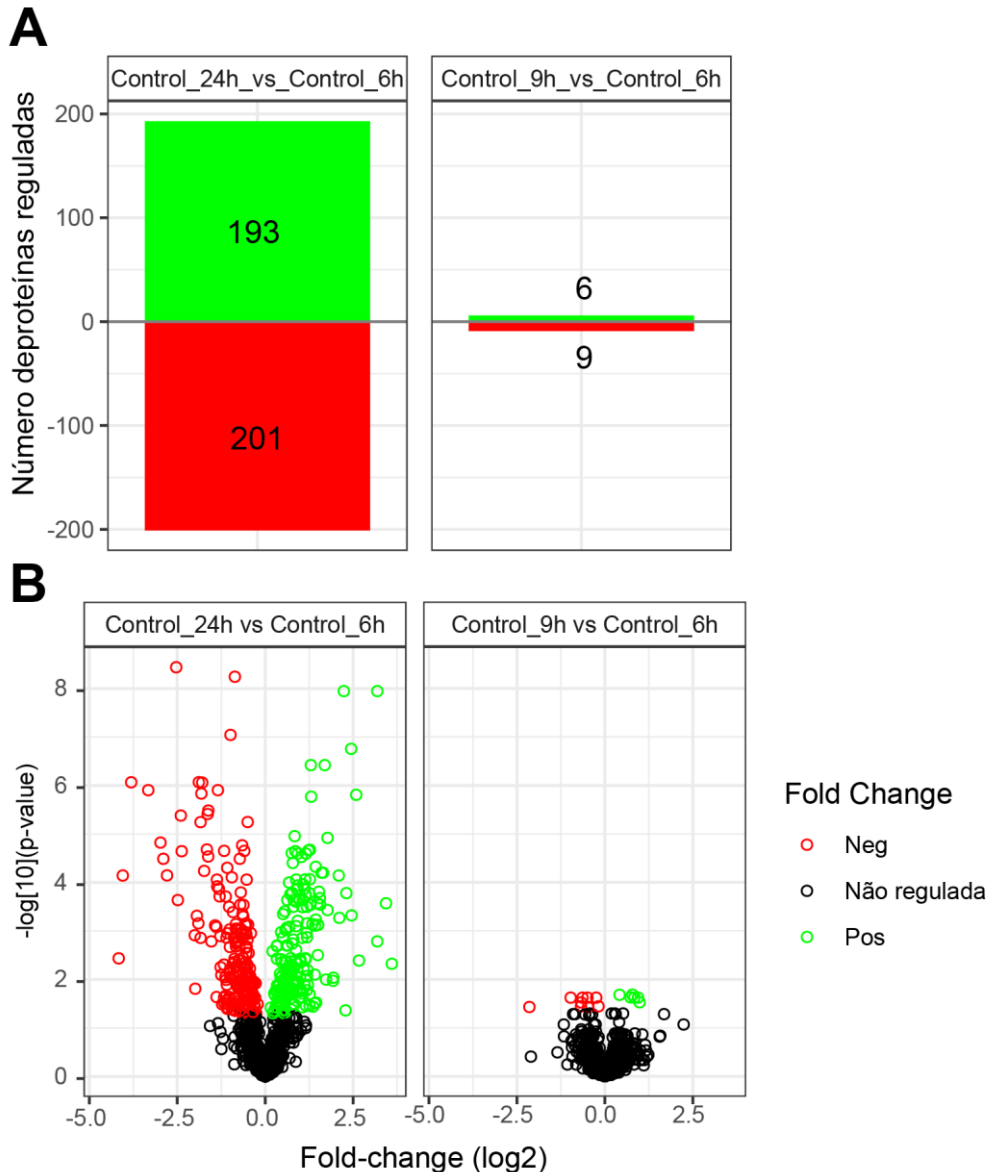


Figura 15. Análise da diferença temporal do cultivo de *C. glutamicum*, sem indução por penicilina. A) Número de proteínas reguladas. B) *Volcano plot* das proteínas analisadas.

Para avaliar o efeito da variação temporal na abundância das proteínas de *C. glutamicum* sem estímulo e sob estímulo para produção de glutamato por penicilina-G, foi realizada análise de clusterização sensível a variância (Schwämmle & Jensen, 2018). Foi estimado pelo método de distância mínima entre centroides que os valores ótimos de clusters para ambas as condições é de três (figura 16). Sendo assim, foi feita a clusterização, resultando na formação de dois clusters com

padrão semelhante entre as condições controle e PenG, clusters 1 e 2 (figura 16). A maior diferença observada foi no cluster 3, exibindo aumento de abundância no tempo 9 h, seguido de redução em 24 h na condição controle (figura 16 A), enquanto na condição PenG, a abundância dessas proteínas apresentou redução no tempo 9 h, seguido de aumento no tempo 24 h (figura 16 B).

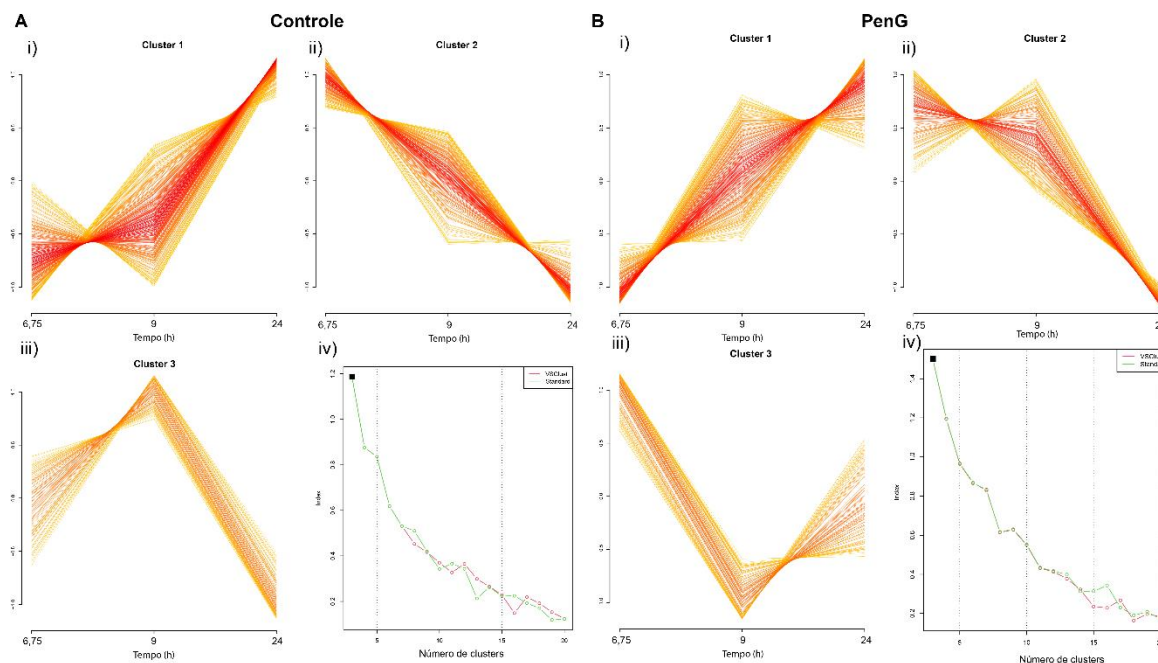
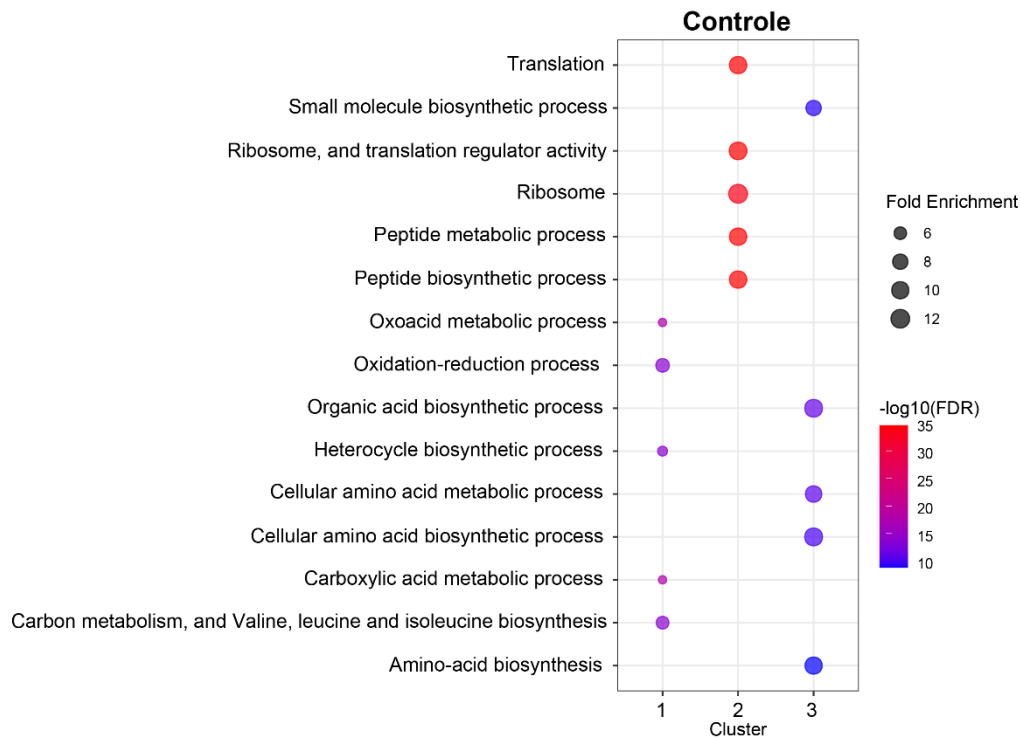


Figura 16. Análise de cluster sensível a variância. Clusterização das abundâncias das proteínas sem *missing values* em três clusters como determinado pela distância mínima entre os centroides (quadro iv). Cada condição foi analisada separadamente: controle (A); PenG (B). Os clusters formados foram nomeados em ordem de acordo com o número de proteínas agrupadas, e são apresentados nos diferentes quadros (i, ii, iii) de cada condição. Análise realizada pelo software VSclust (Schwämmle & Jensen, 2018) com os dados normalizados pelo MetaMorpheus e transformados por log2.

A diferença apresentada entre os perfis de variação, principalmente do cluster 3 (figura 16, quadro iii) sugerem que variações temporais no cultivo podem ser confundidas com efeitos causados pela indução por glutamato, como sugerido anteriormente. A função biológica dos clusters de cada condição também apresentou diferenças (figura 17). O cluster 3 da condição controle apresentou termos relacionados ao metabolismo de aminoácidos e biossíntese de outras moléculas (figura 17 A), sugerindo que ocorre um aumento dessas funções ao longo

da fase exponencial do crescimento bacteriano (6 h e 9 h), seguida de decréscimo drástico na fase estacionária (24 h). Na condição PenG, o cluster 3 apresentou perfil diferente, com termos super-representados relacionados à maquinaria de tradução e metabolismo de amida (figura 17 B), sugerindo que proteínas relacionadas à essas funções diminuíam em abundância após o estímulo por Penicilina G, mas que 17,25 h após a indução (24 h de cultivo) aumentem em quantidade. Ademais, os termos “*Mixed, incl. carbon metabolism, and valine, leucine and isoleucine biosynthesis*”, “*Carbon metabolism, and Valine, leucine and isoleucine biosynthesis*”, “*Carboxylic acid metabolic process*”, e “*Oxoacid metabolic process*” foram super-representados no cluster 1 de ambas as condições, sugerindo que a variação apresentada de algumas ou todas as proteínas possam ocorrer decorrente do tempo, e serem confundidas com variações causadas pelo estímulo. Em relação ao cluster 2, os termos super-representados nas duas condições apresentaram grandes diferenças, estando muito relacionados à tradução na condição controle (figura 17 A), enquanto na condição PenG, estão mais relacionados a vias metabólicas de biossíntese (figura 17 B), sugerindo diferenças consideráveis nas proteínas dos dois clusters, apesar de apresentarem padrão de variação temporal semelhante (figura 17, quadro ii).

A



B

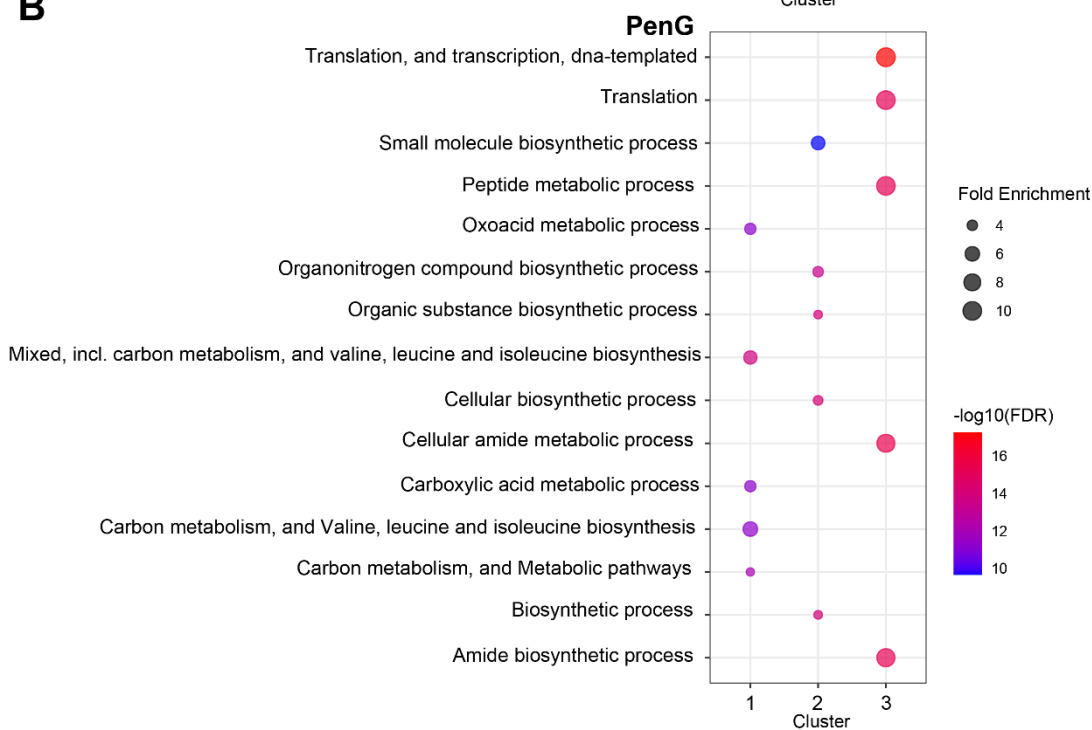


Figura 17. Análise de super-representação das proteínas de cada cluster obtido pela clusterização sensível à variância. As proteínas de cada cluster foram submetidas separadamente a análise de super-representação em ShinyGO (Ge et al., 2020) e os 5 termos mais enriquecidos (com maior *fold enrichment*) foram selecionados e plotados em forma de gráfico de bolhas. A) Termos super-representados nos clusters da condição controle. B) Termos super-representados nos clusters da condição PenG.

Por apresentarem perfis tão diferentes, é possível que as proteínas presentes no cluster 3 da condição PenG, relacionadas à processos de tradução, estejam associadas à indução para produção de glutamato por penicilina G (figura 17, quadro iii). Ademais, termos relacionados à tradução na condição controle foram super-representados no cluster 2, apresentando redução de 6 h para 9 h e de 9 h para 24 h (figura 17 A). Essas diferenças sugerem uma mudança no padrão de proteínas relacionadas a tradução decorrente da indução por penicilina G, causando diminuição dessas proteínas após 9 h, seguida de aumento após 24 h (figura 16 B). Em concordância com essa observação, as proteínas com maior abundância na condição PenG_24h em comparação com a Controle_24h estavam super-representadas para termos relacionados à tradução (figura 14). A regulação negativa da tradução também foi constatada em nível de transcritos após indução por penicilina G, com diminuição após 1 h de indução, até pelo menos 10 h (Hirasawa et al., 2018). Em nossa análise com o controle temporal, no entanto, sugere que a abundância das proteínas relacionadas a tradução diminui logo após o estímulo por penicilina (9 h de cultivo), mas aumenta após 24 h. Análise proteômica semi-quantitativa da indução para produção de glutamato por tween-40 também não observou a redução na abundância de proteínas ribossomais (Mizuno et al., 2016). Esse estudo avaliou a variação de proteínas somente no tempo 9 h, aproximadamente duas horas após a adição de tween-40, em relação a um controle, no entanto só foram consideradas proteínas reguladas as que estivessem presentes em uma condição e ausentes na outra, ou com pelo menos o dobro da abundância. Isso pode ter dificultado a detecção da variação na abundância das proteínas ribossomais, congruente com o ANOVA realizado nesse estudo, onde poucas proteínas foram consideradas reguladas entre as condições controle_9h e PenG_9h (figura 13 B).

As características temporais observadas nas condições controle e PenG sugerem a relevância da utilização de controles temporais no estudo de estímulos para produção de glutamato em *C. glutamicum*.

Grandes diferenças foram observadas entre as condições Controle_24h e PenG_24h, mais especificamente, foram identificadas 23 proteínas com pelo menos o dobro da abundância ($\log_2FC > 1$), e 77 proteínas com menos da metade da abundância ($\log_2FC < -1$) na condição PenG (figura 18). Dentre as proteínas reguladas negativamente, estão pelo menos três grupos associados diretamente ao metabolismo de aminoácidos: “*Glutamate biosynthetic process*”, “*Aromatic amino acid biosynthesis*”, “*Cysteine and methionine metabolism*”. As duas proteínas pertencentes ao processo de biossíntese de glutamato são as subunidades pequena (gltD) e grande (gltB) da glutamina 2-oxoglutarato aminotransferase. Apesar de estarem envolvidas no metabolismo de glutamato, o complexo da glutamato sintase (GOGAT), formado por gltD e gltB, não é essencial para a produção de glutamato em *C. glutamicum* (Beckers, Burkovski, & Nolden, 2001). Mais especificamente, a enzima da glutamato desidrogenase (GDH) foi descrita como responsável pela aminação de 2-oxoglutarato em condições de superprodução de glutamato em *C. glutamicum* (Hashimoto, 2016), o que está de acordo com o papel da GDH na assimilação de amônio, quando esse está presente em alta concentração (Burkovski, 2003). Apesar disso, a GDH (P31026/Cgl2079/cg2280/NCgl1999) não apresentou regulação em nenhuma das comparações analisadas nesse estudo. Em estudo de transcritos, a expressão de *gdh* estava aumentada 2 h após a indução por penicilina (Hirasawa et al., 2018). No entanto, como já mencionado anteriormente, esse estudo não apresentou controle temporal. Na nossa análise, a GDH apresentou regulada negativamente na comparação Controle_24h Vs Controle_6h (Tabela suplementar 3), sugerindo que a regulação negativa dessa enzima descrita anteriormente em nível de transcrito possa ser ocasionada pelo tempo de cultivo.

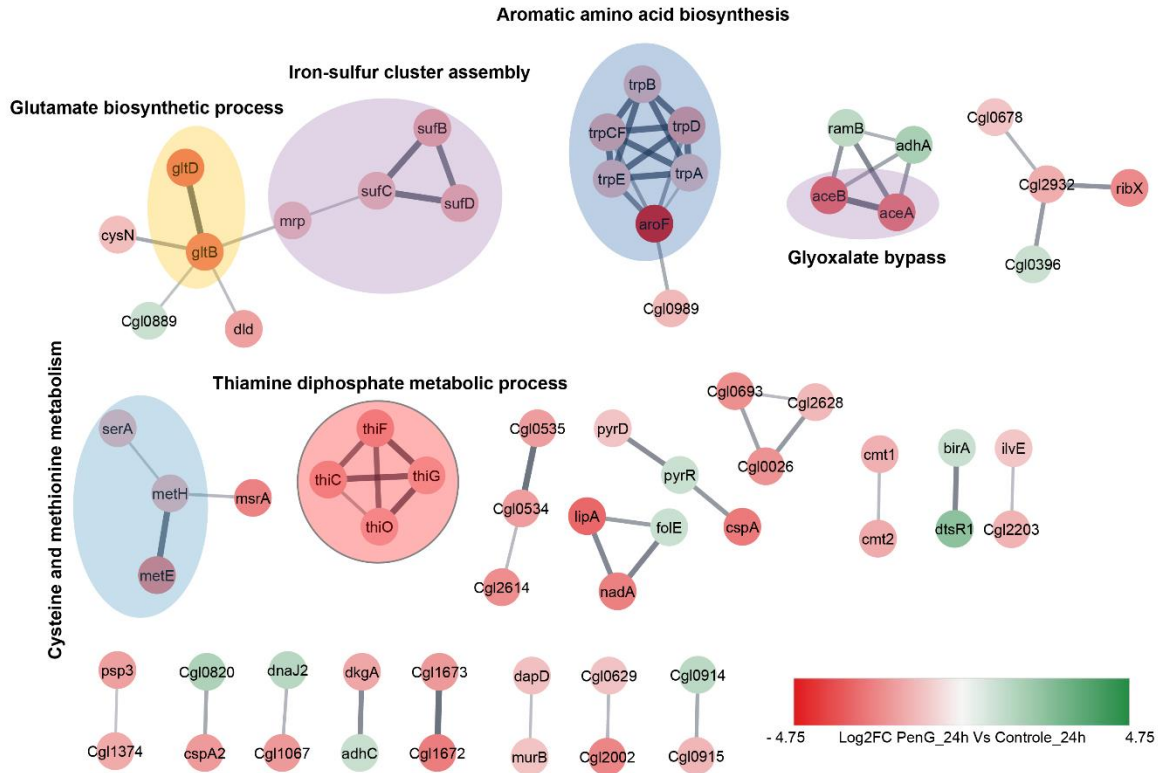


Figura 18. Interações entre proteínas reguladas com $\log_2FC > 1$ ou $\log_2FC < -1$. As interações das proteínas reguladas foram obtidas por meio do Cytoscape stringapp (N.T. Doncheva, Morris, Gorodkin, & Jensen, 2019), em seguida foram clusterizadas de acordo com suas interações e os grupos formados foram submetidos a análise de super-representação para identificar funções biológicas relacionadas aos grupos. As funções biológicas super-representadas e as proteínas pertencentes dos grupos estão destacadas em diferentes cores.

Um grupo de proteínas relacionadas a “*iron-sulfur cluster*” (Fe-S) (sufC, sufB, sufD) apresentou interação com as proteínas da biossíntese de glutamato por meio da proteína mrp (figura 18). Todas essas proteínas apresentaram regulação negativa na condição PenG. As proteínas de cluster Fe-S são responsáveis pela síntese do cluster ferro-enxofre e pela transferência desse para outras proteínas, formando holoproteínas (Esquilin-Lebron, Dubrac, Barras, & Boyd, 2021). Essas proteínas também estão comumente relacionadas à proteção contra estresses oxidativos (Nakunst et al., 2007). Em *C. glutamicum*, proteínas do operon *suf* foram relacionadas à estresse de dissulfeto, tendo a expressão regulada por SigM (Nakunst et al., 2007). A redução na abundância das proteínas do cluster Fe-

S na condição PenG sugerem que a condição controle de *C. glutamicum* pode estar sobre estresses oxidativos.

A única proteína regulada positivamente na condição PenG presente no cluster de biossíntese de glutamato e cluster de Fe-S foi uma glicosidase (Q8NRZ7/Cgl0889), da família Glicosil hidrolase 13 (GH13), mais especificamente, ela está anotada com função molecular de atividade de ciclomaltodextrinase (EC:3.2.1.54), provavelmente sendo classificada na subfamília GH13_20. A família GH13 é a maior entre as glicosil hidrolases, sendo dividida em subfamílias (Stam, Danchin, Rancurel, Coutinho, & Henrissat, 2006).

Outros dois grupos relacionados ao metabolismo de aminoácidos apresentaram regulação negativa na condição PenG_24h, o metabolismo de cisteína e metionina, e o metabolismo de aminoácidos aromáticos (figura 18). A expressão dos genes desses dois grupos metabólicos foi regulada negativamente após a indução por penicilina-G (Hirasawa et al., 2018). Além disso, a biossíntese de fenilalanina, tirosina e triptofano foi sugerida como regulada negativamente pela indução por tween 40 para produção de glutamato em *C. glutamicum* (Mizuno et al., 2016).

O metabolismo de tiamina também foi muito regulado negativamente na condição PenG_24h (figura 18). Essa regulação negativa também foi observada em indução por tween 40 (Mizuno et al., 2016). Ademais, a reação de decarboxilação da 2-oxoglutarato pela ODHC, mais especificamente pela subunidade OdhA, ocorre de maneira dependente da tiamina pirofosfato. Considerando o papel fundamental da atividade de ODHC na produção de glutamato (J. Kim et al., 2011; Kimura et al., 1999; Shimizu et al., 2003), o metabolismo de tiamina pirofosfato pode estar influenciando a produção de glutamato induzida por penicilina-G.

Duas proteínas do ciclo do glioxilato, a malato sintase (aceB/glcB/P42450) e a isocitrato liase (aceA/P42449) estavam com menor abundância na condição PenG_24h (figura 18), essa regulação é congruente com dados de transcritos descritos anteriormente, onde genes estavam regulados negativamente após indução por penicilina-G (Hirasawa et al., 2018). Apesar disso, análise metabólica

indica que a indução para produção de glutamato não altera o fluxo por essa via (Shirai et al., 2007).

4.2.4 Conclusões

Nesse estudo foram analisadas a secreção de aminoácidos e a resposta em nível de proteínas da bactéria *C. glutamicum* sob estímulo para produção de glutamato por penicilina-G em diferentes tempos, correspondendo a condição em fase de crescimento logarítmica (6h e 9h) e estacionária (24h). Como esperado, foi observado grande produção de glutamato na condição estimulada, enquanto na condição controle não foi detectado em nenhum dos tempos analisados. Congruente com o aumento da produção desse aminoácido, foi descrita redução significativa da produção de arginina na condição com estímulo por penicilina, sugerindo relevância da biossíntese desse aminoácido para a produção de glutamato.

A análise proteômica considerando variações temporais resultou na detecção de efeitos já descritos anteriormente por diferentes mecanismos de indução da produção de glutamato, como diminuição da abundância de proteínas da biossíntese de aminoácidos aromáticos, inibição do metabolismo de cisteína e metionina, diminuição do ciclo do glioxilato, e diminuição do metabolismo da tiamina (Hirasawa et al., 2018; Mizuno et al., 2016). Além disso, essa análise revelou diferenças consideráveis entre as análises anteriores, indicando possível influência do tempo nesse mecanismo, ou regulação pós-traducional não observada no nível de transcrito. Por exemplo, as proteínas Q8NQ03/cg1859 e Q8NSY0/cg0625, proteína secreta e proteína de micomembrana, foram observadas como reguladas negativamente pela adição de penicilina. Mudanças globais no perfil de variação das proteínas também foram notadas entre as condições Controle e sob estímulo por penicilina-G, sugerindo que alterações metabólicas e na síntese de proteínas previamente descritas como causadas por indução podem ser decorrentes de efeitos do crescimento bacteriano.

4.3 Integração de proteômica *bottom-up* multi-enzimática e *top-down* para caracterização de proteoformas da bactéria *C. glutamicum*

4.3.1 Objetivos

O estudo de *C. glutamicum* por abordagens proteômicas *top-down* e *bottom-up* foi realizado com a finalidade de caracterização das proteoformas dessa bactéria, principalmente aquelas relacionadas ao metabolismo de aminoácidos, mais especificamente o metabolismo de glutamato. Foram analisadas proteínas intracelulares de *C. glutamicum* obtidas após 9 h de cultivo em condição controle e com estímulo para produção de glutamato por Penicilina-G. A proteômica *bottom-up* identificou modificações pós-traducionais (PTMs) presentes nessas condições e foi usada para a criação de bancos de dados com essas PTMs (Solntsev et al., 2018). Por outro lado, a proteômica *top-down* fez a identificação de proteoformas, inicialmente utilizando a abordagem tolerante a erro (Kou et al., 2016), em seguida foi utilizado o banco de dados com as PTMs identificadas pela abordagem *bottom-up* para a identificação de proteoformas, com tolerância restrita (Yuling Dai, Millikin, Rolfs, Shortreed, & Smith, 2022; Schaffer et al., 2020) (figura 19).

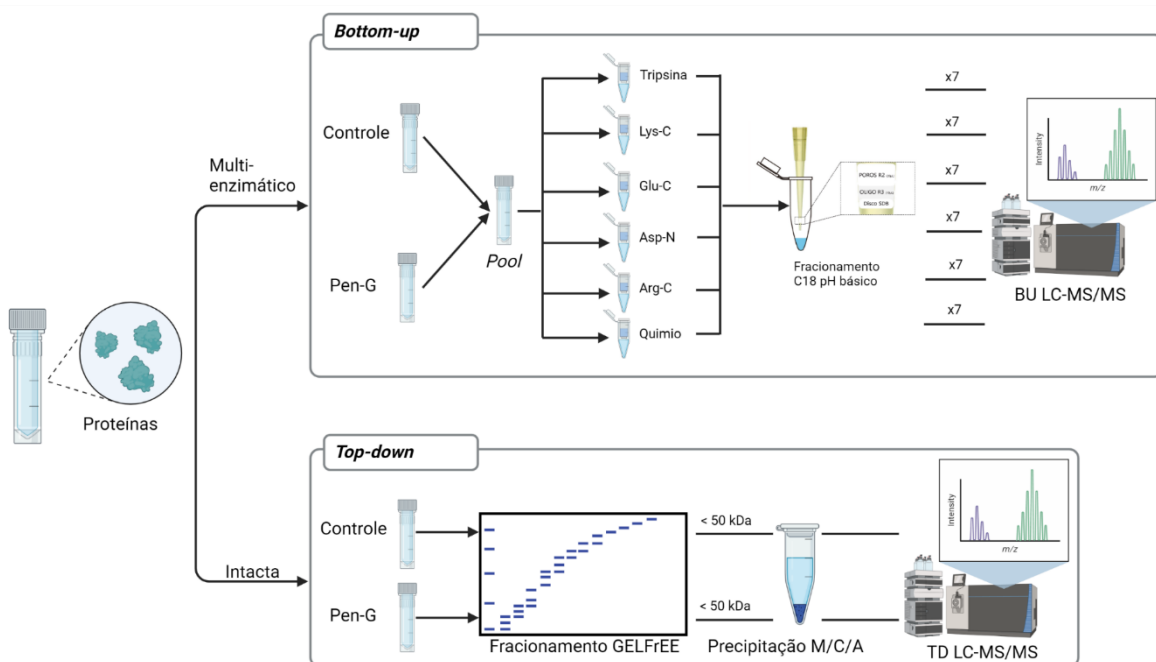


Figura 19. Fluxograma do preparo de amostra para análises proteômicas *bottom-up* e *top-down* das amostras de *C. glutamicum*.

4.3.2 Métodos

Cultivo e extração de proteínas

As amostras analisadas para integração de proteômica *bottom-up* e *top-down* foram obtidas a partir do cultivo e lise celular descritos no estudo da sessão 4.3 dese capítulo “Análise *bottom-up* quantitativa comparando condição controle vs condição produtora de glutamato”. No entanto, foram selecionadas somente as condições controle e Pen-G, 9 h após o inóculo, para observar mudanças iniciais do estímulo para produção de glutamato.

Análises proteômicas *bottom-up* e *top-down*, e integração dos dados

Todos os métodos para análises proteômicas, assim como a integração dos dados realizados nesse estudo foram como os descritos no capítulo III (Metodologia Comum).

4.3.3 Resultados

Análise global da proteômica *top-down*

A análise proteômica *top-down* foi realizada utilizando as amostras provenientes da condição controle e Pen-G cultivadas por 9 h. Foi feito um *pool* de cada condição, e a partir de 400 µg de cada *pool* foi realizado fracionamento por massa, GELFrEE. Cada fração foi analisada por SDS-PAGE para avaliar a eficiência do fracionamento e quais frações seriam analisadas por LC-MS/MS. Foi evidenciado a competência desse fracionamento em ambas as amostras no perfil apresentado em SDS-PAGE (figura 20). As frações 0, 1, 2, 3, 4, 5, 7, 9, 11, 13, 15 e 20 foram selecionadas para análise em LC-MS/MS por possuírem proteínas abaixo de 50 kDa (figura 20).

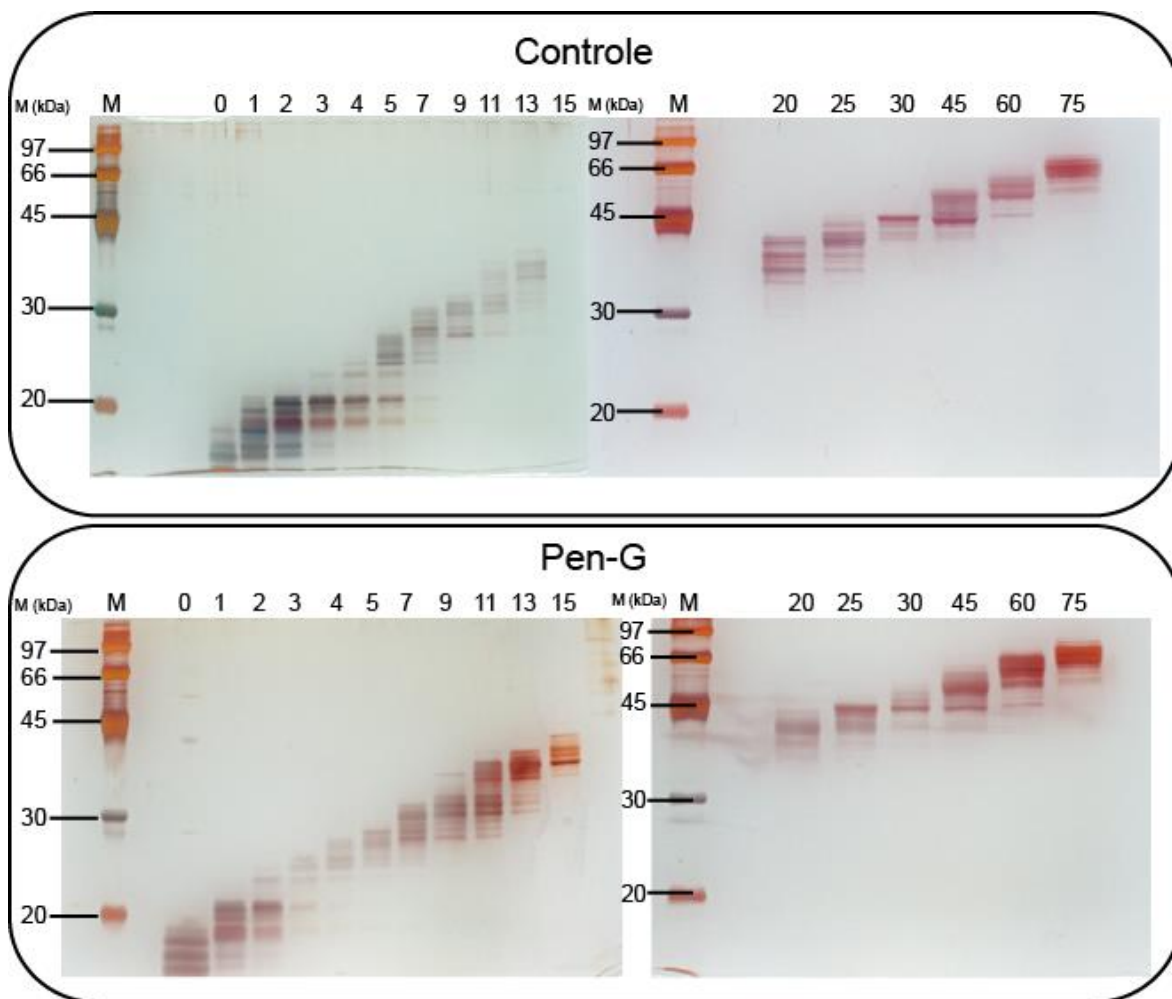


Figura 20. SDS-PAGE das frações de GELFrEE obtidas a partir das amostras das condições controle e Pen-G. As frações são representadas por números que indicam o tempo em minutos após a saída do azul de bromofenol durante o fracionamento por GELFrEE.

As frações selecionadas para LC-MS/MS foram então precipitadas por metanol/clorofórmio/água e ressuspensas em solução de injeção. A análise dos espectros deconvoluídos de MS1 demonstrou que a grande maioria das proteoformas detectadas apresentaram massa menor de 30 kDa (figura 21). No entanto, algumas proteoformas foram maiores que 30 kDa foram detectadas, principalmente nas frações 13, 15, 20 da condição controle (figura 21 B), enquanto na condição Pen-G foi possível detectar ainda mais proteoformas acima de 30 kDa, e até mesmo algumas maiores que 40 kDa (figura 21 A). Isso sugere um

fracionamento diferente entre as condições, no entanto, ambos foram eficientes na separação das proteoformas abaixo de 50 kDa das demais.

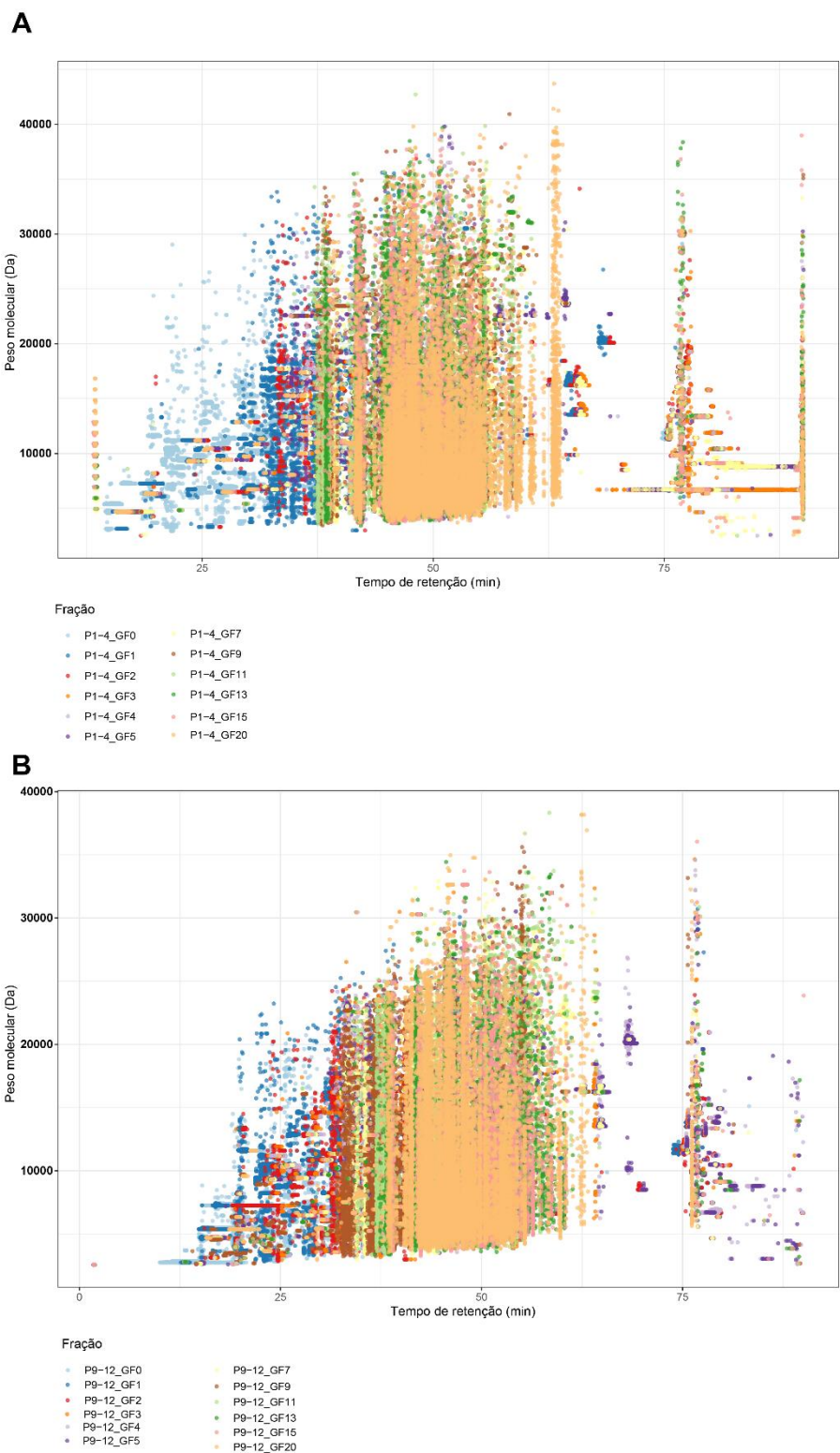


Figura 21. Massa dos *features* detectados e deconvoluídos de acordo com o tempo de eluição de cada fração dos *pools* analisados. Os espectros adquiridos de cada fração foram deconvoluídos pelo programa TopFD (Basharat et al., 2023) e representados em imagem com auxílio do programa VisioProt-MS (Locard-Paulet et al., 2019). A: *Pool* da condição Pen-G; B: *Pool* da condição controle.

Os eventos detectados da fração GF0 da condição controle e Pen-G foram analisados quanto a diferença de massa (Δm - *mass shift*) entre outros *features* utilizando o programa MStopdiff, revelando possíveis modificações frequentes na amostra (Kaulich et al., 2022). Essas modificações são inferidas baseado na subtração entre todos os *features* detectados da corrida em determinados tempos de retenção e similaridade de carga. As diferenças relacionadas a modificações mais comuns vão se sobressair em relação ao ruído de diferenças aleatórias entre proteínas diferentes (Kaulich et al., 2022). O padrão de Δm entre as duas condições apresentou algumas diferenças, mas Δm mais intensas, como 16 Da; 52,91 Da; 99,93 Da foram observadas em ambos (figura 22). Além das Δm observadas é importante notar a ausência ou baixa intensidade de Δm próxima de 266 Da, que normalmente está relacionada a adutos de SDS (Kaulich et al., 2022), possivelmente decorrente do fracionamento por GELFrEE, sugerindo que a limpeza da amostra por metanol/clorofórmio foi eficiente. Os histogramas também apresentaram baixa intensidade de íons com variação de massa de 57 Da (figura 22), isso sugere que a alquilação por iodoacetoamida ocorreu de maneira eficiente.

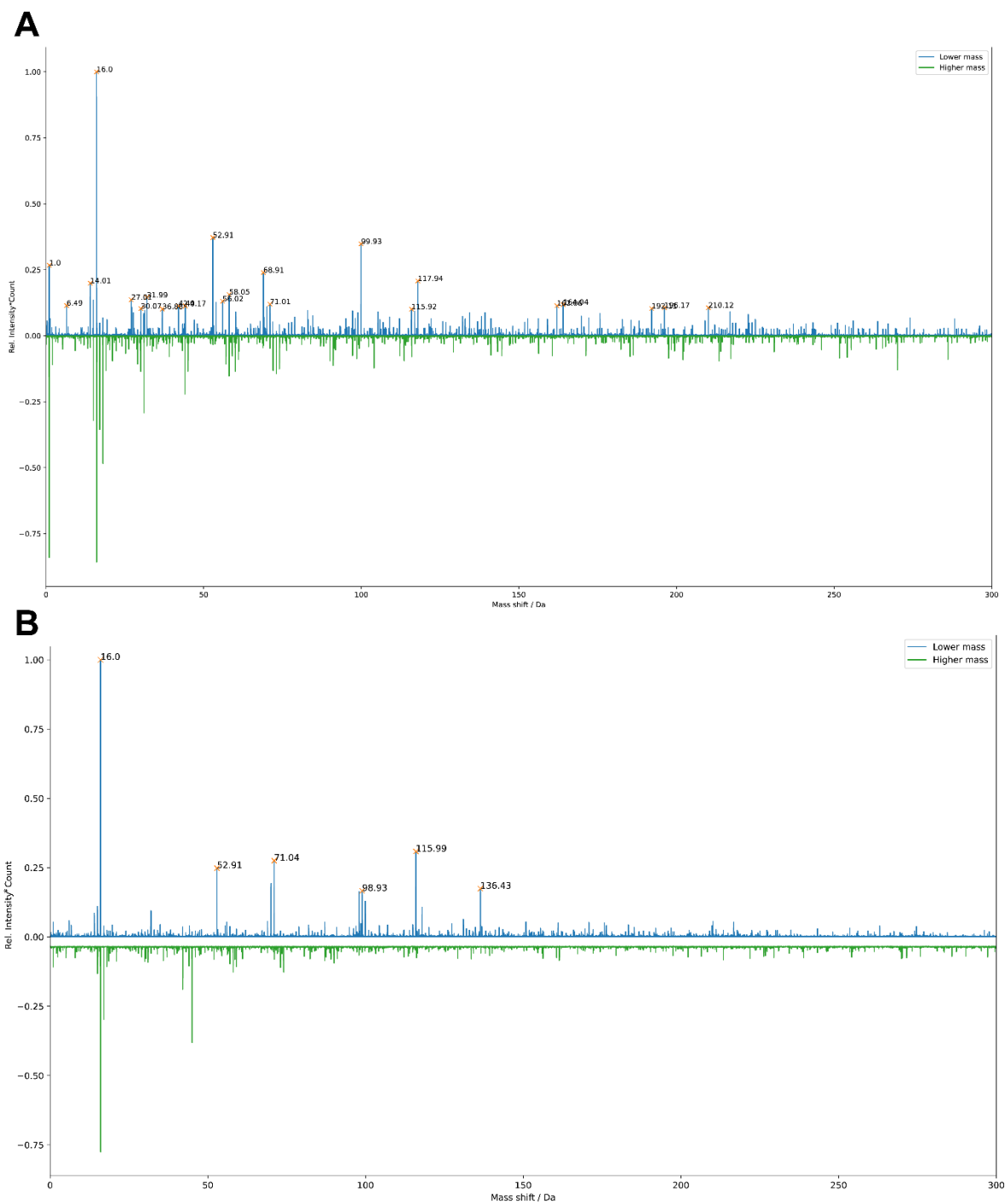


Figura 22. Histograma com a diferença de massa (Δm) calculada entre todos os *features* e relação de intensidade e contagem dessas diferenças da fração GF0. Histograma construído com *binwidth* = 0,01 pelo programa MStoDiff (Kaulich et al., 2022). A: Condição controle. B: Condição Pen-G.

A variação de massa predominante nos histogramas foi de 16 Da, comumente relacionado à presença de oxidação (Unimod AC: 35) (Kaulich et al., 2022). Outra Δm intensa observada foi a de 52,91 Da, essa variação de massa pode estar relacionada a substituição de três prótons por ferro (Unimod AC: 1870), sugerindo que em alguma parte do processamento, as amostras podem estar sujeitas à presença indesejada de metais. A variação de massa de aproximadamente 71 Da foi mais notável no P1-4 (figura 23 B), mas ainda foi detectável no P9-12 (figura 23 A). Essa Δm sugere a modificação de proteínas por acrilamida, possivelmente decorrente do processo de fracionamento por GELFrEE. Essa reação ocorre a partir da acrilamida na matriz do gel interagir com as cisteínas das proteínas, formando o grupamento cisteínil-S- β -propionamida (Xing, Zhang, Chen, & Zhao, 2008). Por meio desses gráficos também é possível observar a relação de intensidade entre a forma de maior massa e a de menor massa. Por exemplo, a intensidade dos íons com Δm de 16 Da de menor massa está próximo da intensidade 1,00, enquanto os de maior massa estão próximo de 0,75, nos dois *pools* analisados (figura 23). Isso sugere que os íons de menor massa com Δm de 16 Da normalmente estão mais intensos que os de maior massa. Por outro lado, os íons com Δm de 52,91 Da praticamente não apresentam intensidade quanto aos íons de maior massa, sugerindo que os eventos com Δm de 52,91 apresentam intensidade muito menor (figura 22). A diferença na intensidade sugere que apesar de ocorrer a provável substituição de três prótons por ferro, a formação desses adutos ocorre em baixa intensidade.

Além das possíveis modificações mais abundantes, a região de Δm até 50 Da apresentou variações que sugerem a presença das PTMs: metilação ($\Delta m = 14,01$ Da; Unimod AC: 34); duas oxidações ($\Delta m = 31,99$ Da; Unimod AC: 425); e acetilação ($\Delta m = 42,01$ Da; Unimod AC: 1) (figura 22 A), todas Δm com maior intensidade no íon de menor massa. Por outro lado, as variações de massa de 17 Da e 18 Da, apresentaram o íon com maior massa sendo o mais intenso (figura 22 A). A perda de 17 Da e 18 Da em proteínas pode estar relacionado às perdas neutras de amidação e desidratação, respectivamente (C. W. Brown et al., 2017). Essa perda de massa em algumas moléculas explicaria os íons de maior

massa estarem mais intensos. Também foram analisadas as frações com proteínas de maior massa, sendo possível observar um padrão de modificações com $\Delta m = 74,02$ Da (figura 23 B). Além disso, diferenças maiores de massa sugeriram diferentes quantidades da mesma modificação de 74,02 Da, por exemplo: 148,04 Da (2x Δm de 74,02 Da); 222,06 Da (3x 74,02 Da); e 296,08 Da (4x 74,02 Da). A intensidade das modificações putativas indica que os íons de menor massa e os de maior massa possuem quantidade parecidas (figura 23 B). A característica de histograma com Δm de 74 Da foi observada em outras frações com proteínas de maiores massas nas duas condições. A variação de massa de 74 Da já foi correlacionada a uma modificação *in-vitro*, causada pela esterificação de glutamato e aspartato por glicerol (Xing et al., 2008).

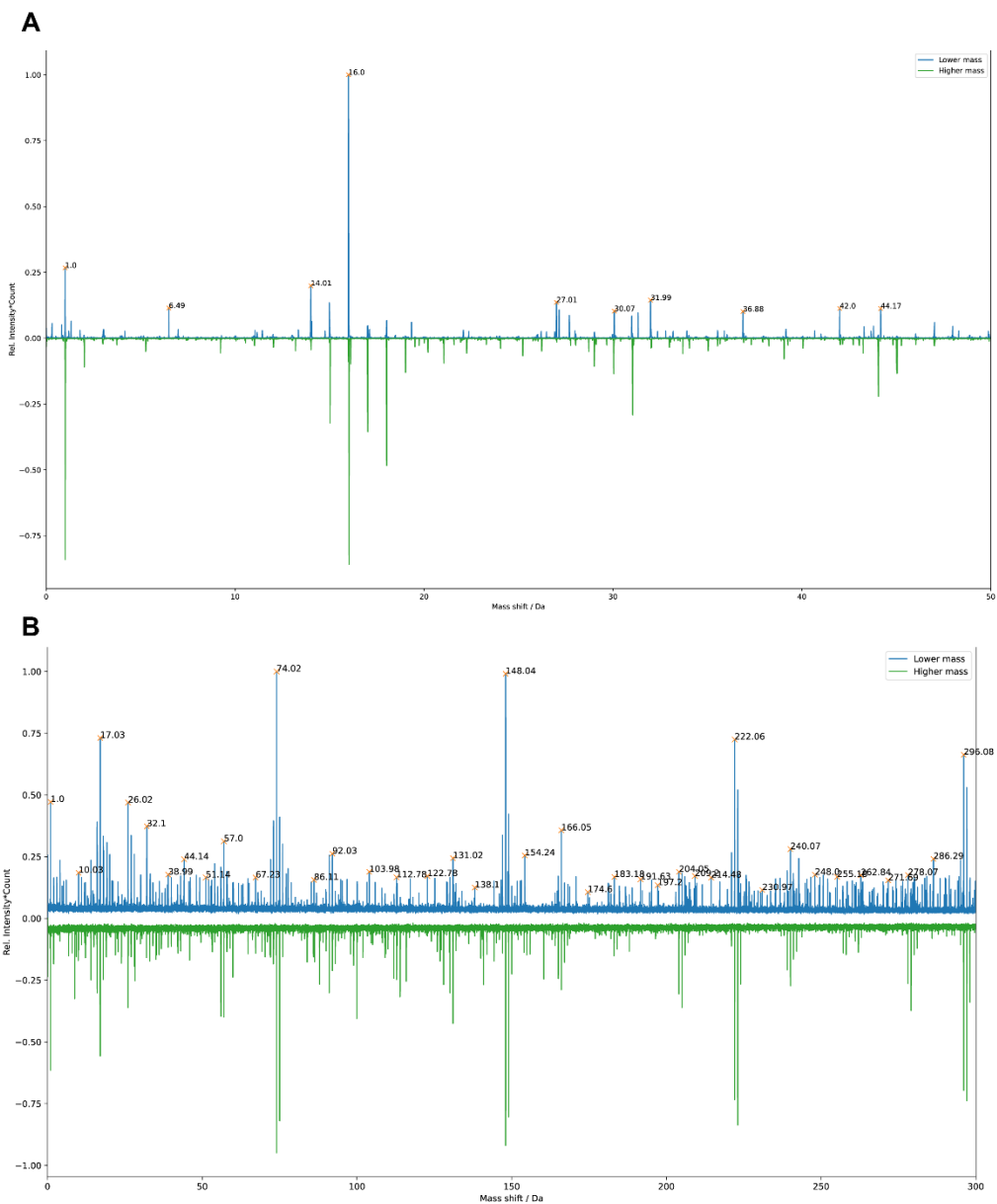


Figura 23. Histogramas com cálculo de diferença de massas entre *features* (Δm). A: Aproximação na região de baixa massa (0 a 50 Da) do histograma da fração GF0, condição controle. B: Histograma da fração GF15 da condição controle.

A identificação de indícios de modificações e até mesmo artefatos ocasionados pelo processamento de amostras podem ser importantes na identificação, possibilitando identificar proteoformas que só seriam reconhecidas com modificações e discernir identificações erradas decorrentes de interpretação equivocada da massa geradas por artefatos. Exemplos dessa relevância já foram

demonstrados em análises proteômicas *top-down* de *Escherichia coli* (Yunxiang Dai et al., 2017) e células humanas do tipo Jurkat (Yunxiang Dai et al., 2019). Sendo assim, as diferenças de massas detectadas pelo *software* MStopDiff devem ser levadas em consideração na identificação de modificações desconhecidas em *C. glutamicum*, e podem ser prospectadas para utilização como modificações variáveis na identificação de proteoformas.

A análise proteômica *top-down* de todas as frações e condições de *C. glutamicum* resultou na identificação de 7351 PrSMs, 1363 proteoformas e 382 proteínas, pelo *software* TopPic (Tabela suplementar 4). As proteoformas identificadas variaram de 1.043,477 Da a 38.182,57 Da (Tabela suplementar 4), e a maioria dos PrSMs identificados possui massa entre 10.000 Da e 20.000 Da (figura 24 B). Atualmente, limitações como o decaimento de proteínas por colisão ou diluição de sinal dificultam a identificação de proteoformas maiores que 30 kDa de maneira global (Fornelli & Toby, 2022). Considerando tal limitação, o número de proteínas com massa teórica menor do que 30 kDa no proteoma de referência de *C. glutamicum* é de 1616, e menor do que 50 kDa é de 2564 (figura 24 A). Levando em consideração as 382 proteínas identificadas e que o número de proteínas com massa teórica menor do que 30 kDa no proteoma de referência de *C. glutamicum* é de 1616 (figura 24 A), foi possível identificar aproximadamente 24% das proteínas possíveis presentes no proteoma de referência.

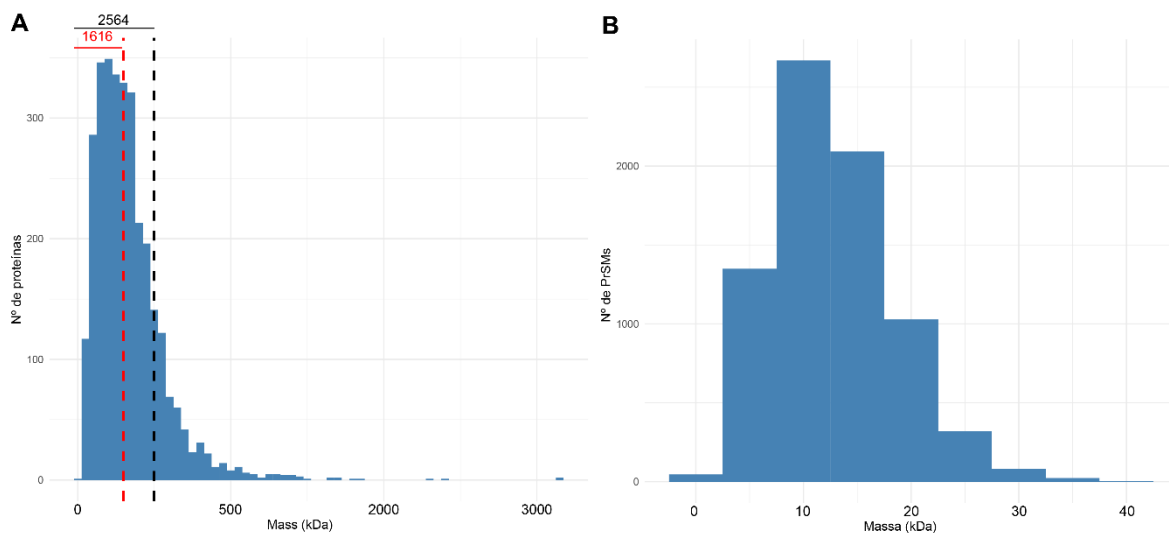


Figura 24. Análise de massa de todas as proteínas PrSMs de *C. glutamicum*. A: Histograma da massa de todas as proteínas do proteoma de referência de *C. glutamicum* disponibilizado no Uniprot. B: Histograma de massa de todos os PrSMs de *C. glutamicum* identificados nesse estudo. Em ambos os histogramas foi utilizado *binwidth* de 5 kDa.

As frações de proteínas de menor massa foram as que apresentaram maior número de PrSMs identificados, enquanto o número de PrSMs reduziu nas últimas frações (figura 25 A). Em relação à massa das proteoformas identificadas em cada fração, foi possível observar o aumento da tendência central (representado pela mediana) de maneira gradual a partir da fração GF0 até a GF7, apresentando medianas próximas de 10.000 Da na fração GF0 e próximas de 18.000 Da na fração GF5 e GF7 (figura 25 B). Além disso, a densidade apresentou tendência de aumento até a fração GF5, enquanto, na fração GF7, essa densidade ficou distribuída entre regiões de massa próximas dos 18.000 Da e abaixo de 10.000 Da (figura 25 B). A partir da fração GF9 a densidade e a mediana reduziram para valores próximos ou abaixo de 10.000 Da, no entanto, proteoformas com valores acima de 20.000 Da aumentaram a densidade acima dessa região (figura 25 B), sugerindo que apesar de apresentar muitas proteoformas de baixa massa, algumas proteoformas de alta massa só foram identificadas nessas frações. A análise proteômica *top-down* de proteoformas grandes (> 30 kDa) é uma dificuldade atual reconhecida pela comunidade científica (Fornelli & Toby, 2022). A limitação dessas proteínas está frequentemente relacionada a diluição de sinal por muitos canais de carga e canais

isotópicos. Outros problemas estão relacionados a congestão de MS2, causada por diversos envelopes isotópicos e diferentes cargas, ocasionando em dificuldade de correlação com espectros teóricos (Fornelli & Toby, 2022).

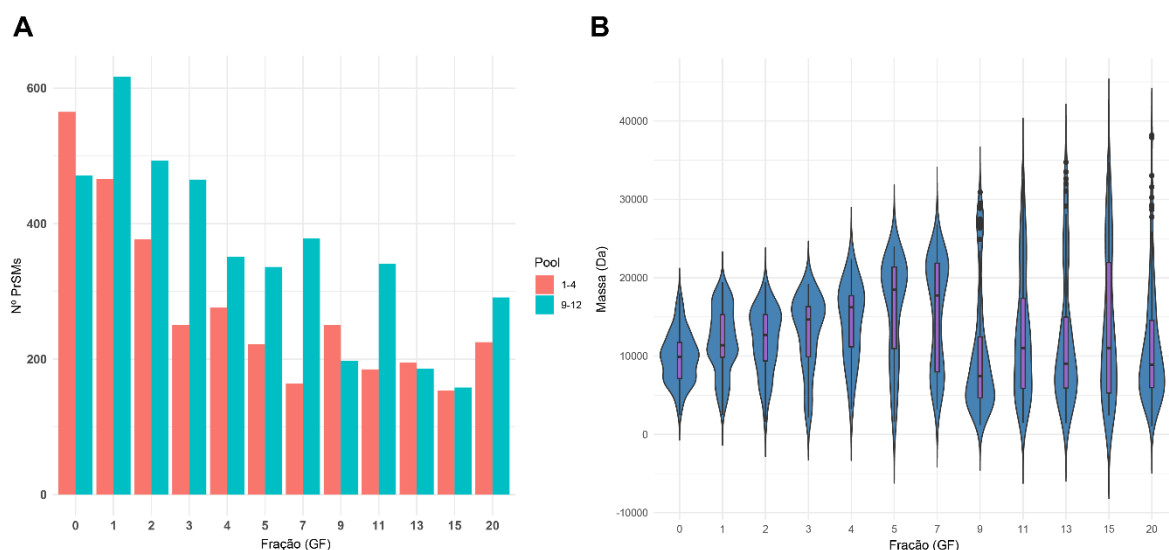


Figura 25. Análise do número de identificações e massa das proteoformas identificadas nas frações de GELFrEE. A: Número de PrSMs identificados em cada fração a partir dos *pools* das condições analisadas. B: *Violin plot* das massas das proteoformas identificadas nas frações de GELFrEE analisadas.

Análise de proteínas com variação de massa (Δm) – Identificações por TopPic

Foram identificadas 279 proteínas com Δm , gerando 998 proteoformas diferentes com FDR < 5 % em nível de proteoforma, o que sugere a presença de proteoformas não anotadas no banco de dados de *C. glutamicum*. A análise funcional dessas proteínas possibilitou a identificação de diversos grupos com diferentes funções, como: grupo de proteínas ribossomais; proteínas da glicólise/gliconeogênese; canal de prótons e atividade de endonuclease NucS; metabolismo de histidina; tioredoxina; metabolismo de tiamina; metabolismo de nucleotídeos; e atividade de NADH desidrogenase (figura 26).

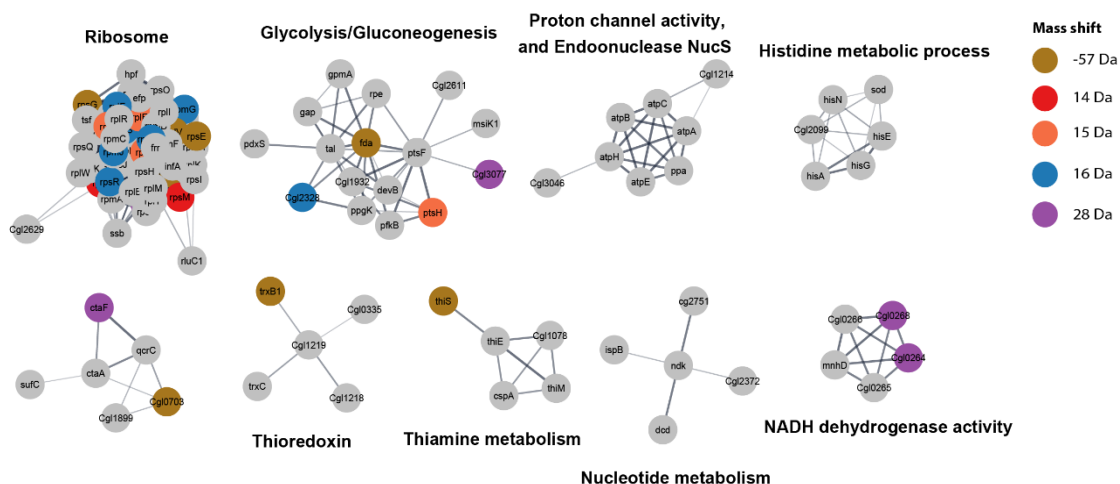


Figura 26. Interação entre proteínas identificadas com variação de massa e representação das variações mais comuns por cores. Interações entre proteínas foram mantidas após a clusterização com granularidade de 4, apresentando os termos super-representados de cada grupo.

Um grande número dessas Δm estavam próximos de ± 2 Da (figura 27 A). Variações de massas dentro desse espaço são comumente causadas por determinação incorreta do pico monoisotópico na análise de proteínas intactas (Yunxiang Dai et al., 2017; Gallagher et al., 2020). Apesar disso, a avaliação do espectro, e observação do sítio modificado da proteína pode sugerir a presença de pontes dissulfeto (Melo et al., 2023). Devido à grande presença de variações de massas próximas a 2 Da, foi feita a remoção dos proteoformas com $-3 < \Delta m < 3$ para remover interpretações de proteoformas causadas por essa determinação errada de íon monoisotópico (figura 27 B). Foram identificados várias proteoformas com Δm de 16 Da e 28 Da, sugerindo a presença de modificações por oxidação (UnimodAC: 35, $\Delta m = 15.994915$) e formilação (UnimodAC: 122, $\Delta m = 27.994915$), respectivamente. Também chamou atenção proteoformas com variações de massas próximas de 14 Da e 15 Da, sugerindo a presença de metilação (UnimodAC: 34, $\Delta m = 14.015650$) e deamidação seguida de metilação (UnimodAC: 528, $\Delta m = 14.999666$), respectivamente.

Além dessas modificações, chamaram atenção Δm com valores negativos, como -57 Da e -116 Da. A primeira pode ser causada por alquilação insuficiente

durante o preparo da amostra, causando cisteínas livres, enquanto a segunda não apresenta variação em banco de dados, mas a perda de 115 Da pode ser relacionada à substituição de triptofano por alanina (UnimodAC: 1224, $\Delta m = -115.042199$).

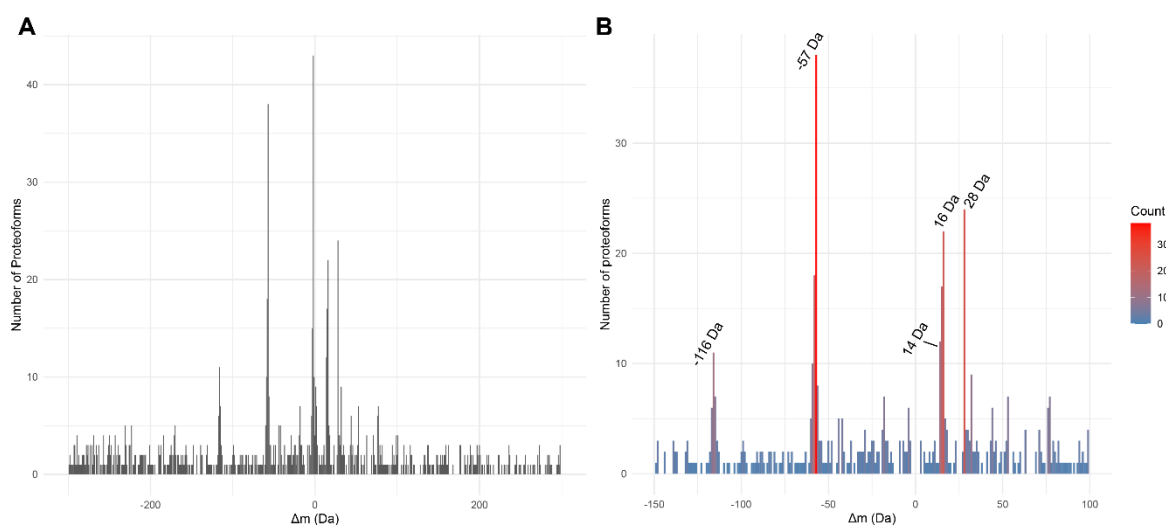


Figura 27. Histogramas da distribuição de variações de massas (Δm) identificadas em *C. glutamicum*. A) Histograma com todas as variações de massas identificadas. B) Histograma com as variações de massas após exclusão daquelas entre -2,5 Da e 2,5 Da. O *binwidth* dos dois histogramas foi de 1 Da.

A grande quantidade de PrSMs com Δm de 16 Da, 15 Da, e 28 Da também foi observada em um estudo proteômico *top-down* de *C. glutamicum* anterior (Melo et al., 2023). Ademais, foi identificado a variação de massa de -115 Da, próximo do -116 Da observado no nosso estudo.

A oxidação é uma PTM comum em organismos com possíveis implicações biológicas, no entanto, é comumente causada como artefato do processamento de amostras para espectrometria de massas (Lennicke et al., 2016). A formilmetionina é o primeiro resíduo de aminoácido utilizado na síntese de proteínas bacterianas. No entanto, a maioria das proteínas produzidas são deformiladas por enzimas deformilase (PDF) durante a tradução, e em seguida tem a metionina do N-terminal clivada por metionina aminopeptidases (MAP), processos de modificação co-

traducional (Giglione, Fieulaine, & Meinel, 2009). A formilação de metionina N-terminal foi relacionada como um mecanismo para controle de qualidade da tradução por meio da sinalização para vias de degradação, fMet/N-degron (Piatkov, Vu, Hwang, & Varshavsky, 2015; Ranjan, Mercier, Bhatt, & Wintermeyer, 2017). Apesar disso, foi observado em outro estudo que proteínas intra-membranares de *E. coli* com N-terminal extracelular podem manter a formilmetionina (Ranjan et al., 2017). Por outro lado, a formilação é uma modificação que também já foi relacionada a presença de ácido fórmico em alta concentração, apesar de que em baixas concentrações não é esperado a indução dessa modificação (Zheng & Doucette, 2016).

Todas as proteoformas identificadas com a variação de massa de 28 Da apresentaram sítio de modificação próximo ao N-terminal, e 23 das 24 proteoformas identificadas apresentavam a metionina no N-terminal (Tabela suplementar 5). A maioria das proteínas identificadas com a metionina formilada no N-terminal são membranares, o que pode ser observado pela análise de super-representação dessas proteínas em relação ao genoma (figura 28 A). Recentemente foi demonstrado que a presença de domínios transmembrana ou regiões hidrofóbicas próximo ao N-terminal, nos 12 primeiros resíduos, inibem a atividade das enzimas PDF e MAP (Yang, Kim, & Shan, 2022). Dentre as proteínas identificadas com N-terminal formilado ($\Delta m = 28$ Da), 14 apresentaram predição de domínio transmembrana iniciando nos 12 primeiros resíduos de aminoácidos. As 9 proteínas sem domínio transmembrana predito tiveram as sequências de 12 resíduos iniciais submetidas a análise de hidrofobicidade, e 8 delas apresentaram GRAVY > 0, sugerindo região próxima ao N-terminal mais hidrofóbica. Tais evidências sugerem fortemente que essas proteínas apresentam proteoformas com metionina no N-terminal formilado. Apesar dessas evidências de identificação sugerirem a inibição do processamento co-traducional por PDF, mantendo a formilação, também foram identificadas outras proteoformas em 16 dessas proteínas, sendo que sete delas apresentaram apenas a proteoforma formilada. Isso indica que apesar inibição pela região transmembrana ou hidrofóbica próxima ao N-terminal, essas proteínas também são encontradas sem essa modificação.

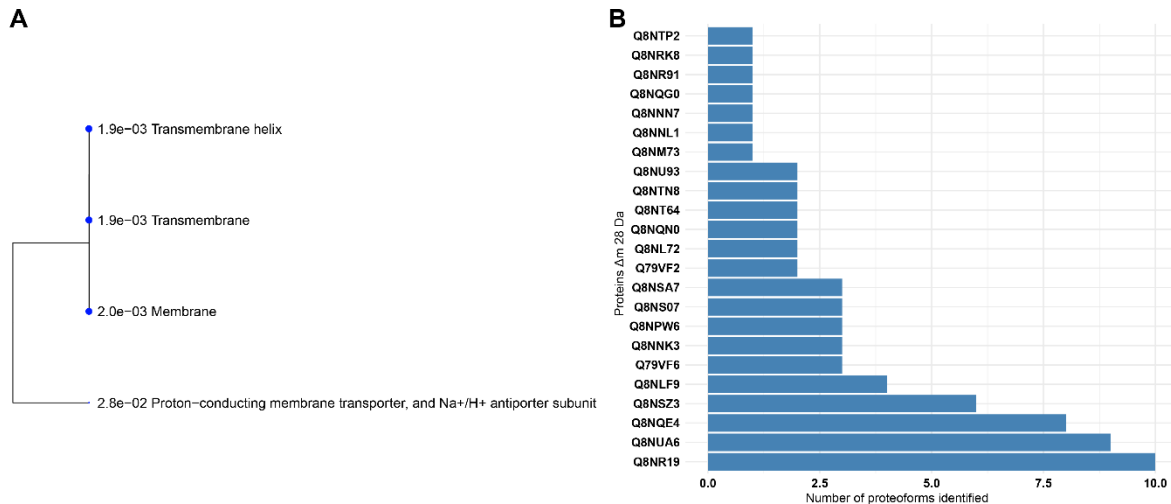


Figura 28. Análise das proteínas formiladas e suas demais proteoformas. A) Termos super-representados nas proteínas identificadas com o N-terminal formilado em relação ao genoma de *C. glutamicum*. B) Número total de proteoformas identificadas em cada proteína com N-terminal formilado.

A identificação das proteoformas com metionina no N-terminal formilado em *C. glutamicum* dão suporte à manutenção da formilmetionina no N-terminal de proteínas transmembranares e/ou com região hidrofóbica próxima ao N-terminal. Além disso, esses resultados demonstram o potencial da abordagem *top-down* para o estudo sobre mecanismos e funções do sistema de modificações co-traducionais por MAPs e PDFs.

Análise comparativa entre controle e Pen-G

Foram identificadas 375 proteínas na condição controle e 265 na condição Pen-G, sendo 222 proteínas em comum entre as duas condições, 153 exclusivas da condição controle, e 43 únicas da condição Pen-G (figura 29 A). Para análise comparativa de proteoformas entre as condições foram consideradas as variações de massa arredondadas (números inteiros), juntamente com o código da proteína para definir proteoformas redundantes. Usando essa estratégia foram identificadas 239 proteoformas similares nas duas condições, 842 proteoformas exclusivas da condição controle e 526 proteoformas identificadas somente na condição Pen-G

(figura 29 B). As proteoformas exclusivas aparentemente diferenciam melhor as condições do que as proteínas.

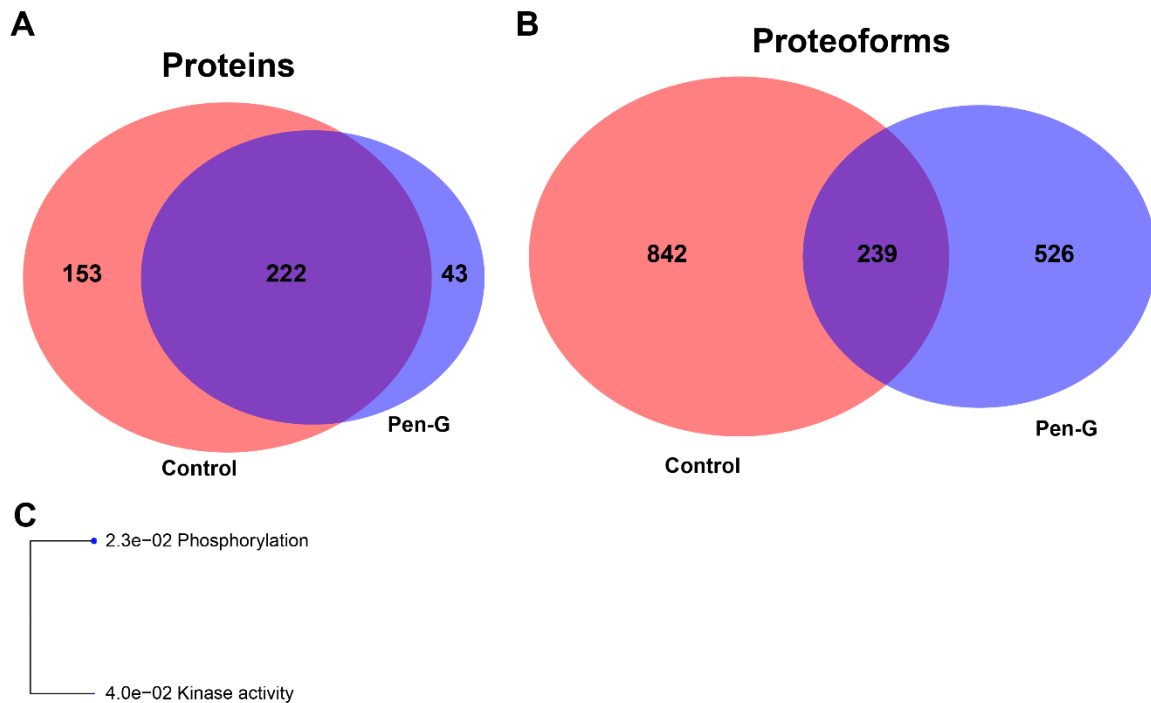


Figura 29. Diagrama de Venn das identificações das condições controle e penicilina-G (Pen-G). A) Comparação na identificação de proteínas. B) Comparação da identificação de proteoformas. C) Termos super-representados para as proteínas exclusivas da condição controle.

As proteínas exclusivas da condição controle foram super-representadas para os termos fosforilação e atividade de quinase, enquanto as exclusivas da condição Pen-G não apresentaram super-representação significativa (figura 29 C). A proteína inibidora da oxoglutarato desidrogenase (*Oxoglutarate dehydrogenase inhibitor* -Odhl) é fundamental para a produção de glutamato em *C. glutamicum* (Niebisch et al., 2006), sendo regulada fosforilação de treonina, quando está fosforilada, reduz a inibição do complexo da oxoglutarato desidrogenase (ODHC), reduzindo a produção de glutamato (J. Kim et al., 2011). Sendo assim, a super-representação de atividade quinase e fosforilação na condição controle está de acordo com a redução da inibição da ODHC e menor produção de glutamato. A principal proteína responsável pela fosforilação de Odhl é a PknG (*serine/threonine*

protein kinase), no entanto ela não foi identificada em nenhuma das condições, possivelmente por sua ter massa superior a 30 kDa, aproximadamente 83,345 kDa (UniprotAC: Q8NM29), o que dificulta a análise por proteômica *top-down* (Fornelli & Toby, 2022). Por outro lado, outras quinases como Dephospho-CoA quinase (UniprotAC: P58897), Aspartoquinase (UniprotAC: P26512), polifosfatoquinase (UnimodAC: Q8NM65), entre outras, só foram identificadas na condição controle (Tabela suplementar 6).

As 222 proteínas identificadas nas duas condições (figura 29 A) foram descritas de maneira diferente quando analisadas em nível de proteoforma, foram mantidas 239 proteoformas em comum entre as duas condições (figura 29 B), correspondendo a 167 proteínas (Tabela suplementar 7). Essa diferença se deve à diferenciação de proteínas apenas em nível de proteoforma. Por exemplo, a proteína reguladora de nitrogênio PII (*Nitrogen regulatory protein PII*, UniprotAC: Q79VF2) foi identificada pela proteoforma sem modificações pós-traducionais (figura 30 A) na condição controle e Pen-G, no entanto, a proteoforma com resíduo próximo ao N-terminal com $\Delta m = 28$ Da só foi identificada na condição controle (figura 30 B). A análise dos espectros dessas proteoformas evidencia a confiabilidade da identificação da proteoforma com a possível formilação do N-terminal, não apresentando envelopes isotópicos na janela de exclusão do precursor (figura 30 B, meio), e apresentando grande acurácia dos fragmentos identificadores da modificação no N-terminal (*b8, b15, b16, b19, b20, b21, b22, b23, b24*) (figura 30 B, inferior). O espectro da proteoforma sugerida como canônica identificada em ambas as condições (figura 30 A) apresentou possível variação de massa menor que 1 Da, no entanto, essa variação pode ter ocorrido devido a um erro de determinação do íon monoisotópicos.

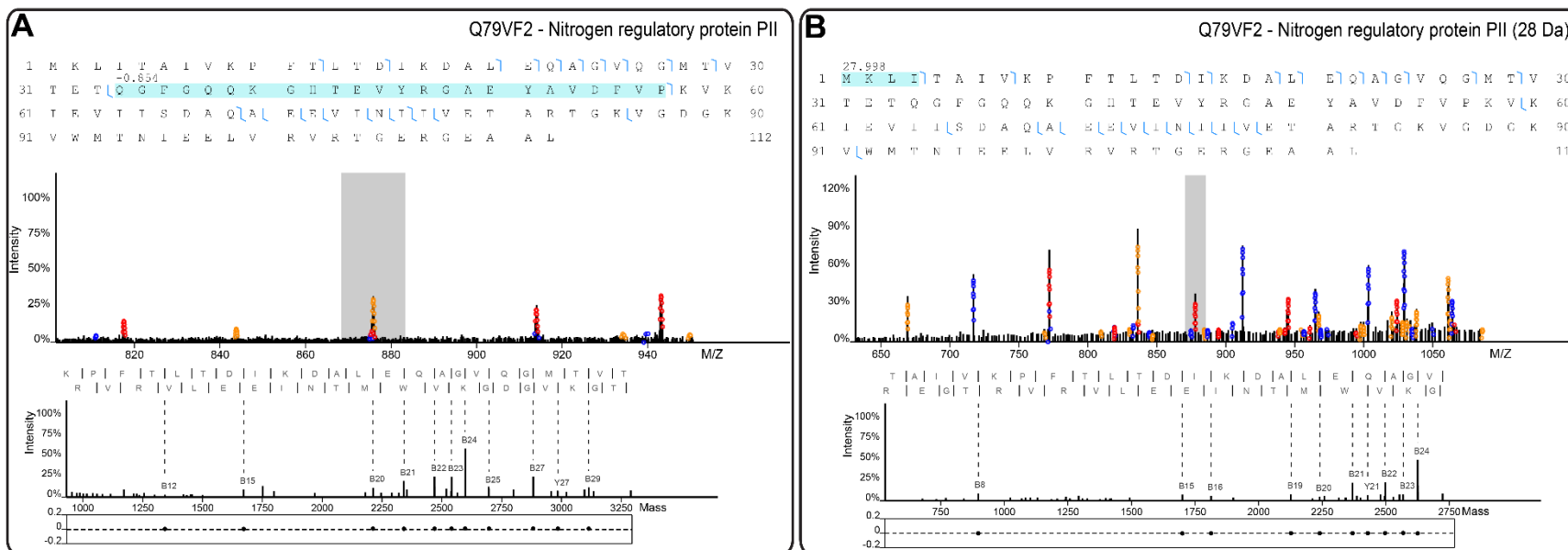


Figura 30. Análise das proteoformas da proteína reguladora do nitrogênio PII (UniprotAC: Q79VF2). São apresentadas duas proteoformas da proteína PII, com as evidências de espectro e sequência de cada uma. No topo de cada figura está apresentada a sequência da proteoforma identificada, com a região modificada preenchida em azul e os fragmentos identificados são representados por traços em azul. No meio das figuras é apresentada o MS1, apresentando o precursor e a janela de isolamento de cada proteoforma. Na parte inferior da figura é apresentado o espectro de massas do MS2, demonstrando os fragmentos identificados de cada proteoforma com o quadro de erro abaixo do espectro. A) Proteoforma com $\Delta m < 1$ Da, provavelmente decorrente de determinação errada de isótopo identificada na condição Pen-G. B) Proteoforma com Δm próximo ao N-terminal com aproximadamente 28 Da, identificada na condição controle.

Como apresentado anteriormente, a modificação de 28 Da próximo a N-terminal ligada a um resíduo de metionina pode estar relacionada com a via de degradação de proteínas para controle de qualidade de tradução (Piatkov et al., 2015), mas também já foi encontrada em proteínas direcionadas a membrana e com região N-terminal muito hidrofóbica (Yang et al., 2022). A sequência da proteína reguladora do nitrogênio PII não apresentou predição para domínio transmembrana (figura 31 A). No entanto, de acordo com o índice Kyte & Doolittle, ela possui aminoácidos hidrofóbicos na região do N-terminal (figura 31 B).

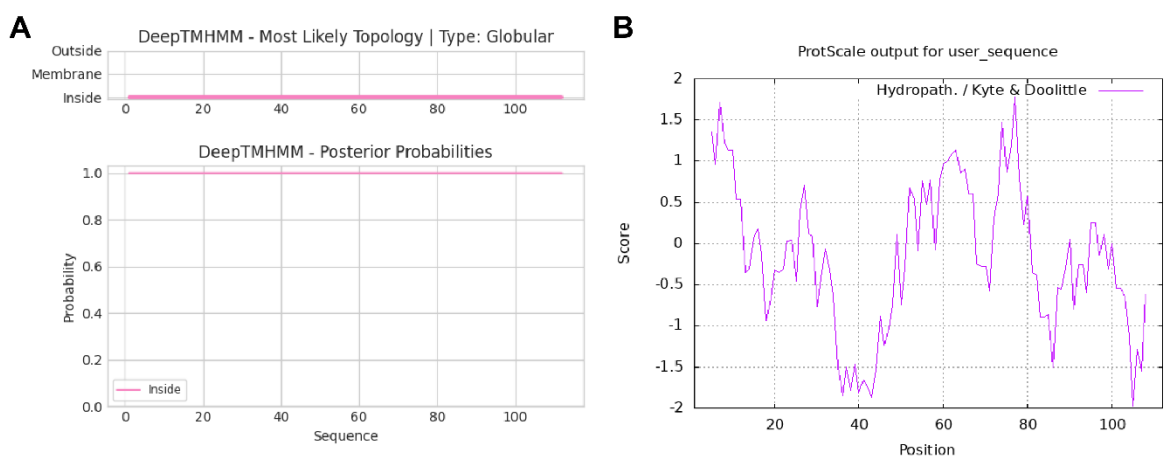


Figura 31. Análise das características bioquímicas preditas da proteína reguladora do nitrogênio PII de *C. glutamicum*. A) Predição de domínio transmembrana por *deepTMHMM* (Hallgren et al., 2022). B) Índice de hidrofobicidade, de acordo com Kyte & Doolittle (Kyte & Doolittle, 1982).

Proteínas PII estão envolvidas na regulação de diversos processos biológicos, como fatores de transcrição, transporte membranar, além de enzimas, sendo fundamentais na regulação de nitrogênio e tem como um de seus principais reguladores o 2-oxoglutarato (Huergo, Chandra, & Merrick, 2013). Em *C. glutamicum* foi recentemente descrito que a proteína PII (CgGlnK) causa o aumento do rendimento da produção de L-arginina, por meio do aumento da expressão dos genes da via de biossíntese desse aminoácido, além de influenciar a produção de ácido glutâmico (M. Xu et al., 2020). O 2-oxoglutarato é substrato para a produção de ácido glutâmico, sua concentração está relacionada à atividade de ODHC, que

normalmente se encontra regulada pelo estado de fosforilação da proteína OdhI (Boulahya et al., 2010; Krawczyk et al., 2010; Niebisch et al., 2006). A identificação dessa proteína PII de *C. glutamicum* com N-terminal formilado indica suscetibilidade ao mecanismo de degradação fN-degron (Piatkov et al., 2015), ou possível endereçamento para região membranaar (Yang et al., 2022). Todavia, essa proteoforma foi identificada como exclusiva da condição controle, sugerindo que a sua formilação pode estar relacionada a produção de glutamato.

A subunidade pequena da acetolactato sintase (uniprotAC: Q79VG4) também só foi diferenciada no nível de proteoforma. Ela foi identificada pela proteoforma sem modificações nas duas condições, enquanto uma proteoforma com Δm de 14 Da foi identificada somente na condição estimulada por penicilina-G (figura 32). O espectro dos precursores identificados (figura 32 A e B, meio) corrobora a identificação da proteoforma na condição Pen-G (figura 32 B), onde é possível observar, dentro da janela de isolamento dois envelopes isotópicos próximos, um de maior massa, identificado como o precursor identificador da proteoforma com a modificação de 14 Da (indicado pela seta em vermelho), e um de menor massa (indicado pela seta azul), com perda de 14 Da em relação ao de maior massa, sugerindo a presença da proteoforma canônica (figura 32 B, meio). Consoante com a exclusividade da proteoforma na condição Pen-G, a presença de envelope isotópico próximo ao canônico não foi observada na proteoforma identificada na condição controle (figura 32 A, meio). A variação de massa de 14 Da normalmente está relacionada a modificação por metilação (UnimodAC: 34, $\Delta m = 14.015650$).

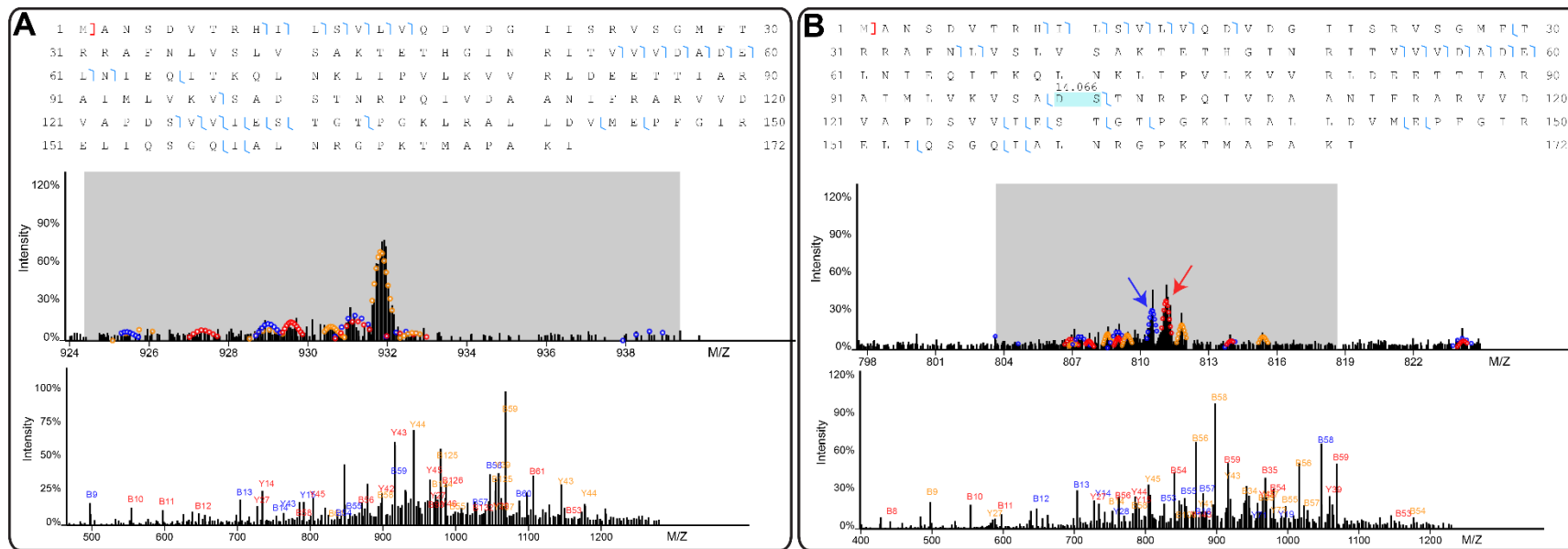


Figura 32. Análise dos espectros das proteoformas da subunidade pequena da acetolactato sintase (uniprotAC: Q79VG4). A sequência de resíduos de aminoácidos das duas proteoformas subunidade pequena da acetolactato sintase são apresentadas no topo de cada figura, com a região modificada preenchida em azul e os fragmentos identificados são representados por traços da mesma cor. No meio das figuras é apresentado o MS1, com o precursor e a janela de isolamento de cada proteoforma. Na parte inferior da figura é apresentado o espectro de MS2, indicando os íons fragmentos identificadores de cada proteoforma. A) Proteoforma canônica. B) Proteoforma com Δm de aproximadamente 14 Da.

A acetolactato sintase, também conhecida como *aceto-hydroxyacid synthase* (AHAS, E.C 2.2.1.6) é uma enzima chave na biossíntese de aminoácidos de cadeias ramificadas (*branched chain amino acids* – BCAA), catalisando a primeira reação para a biossíntese de L-leucina, L-valina, e L-isoleucina (Yadi Liu, Li, & Wang, 2016; Pue & Guddat, 2014). A AHAS é composta de duas subunidades, uma pequena e uma grande, capazes de sintetizar acetolactato a partir da condensação de dois piruvatos, ou acetohidroxibutirato, a partir da condensação de 2-cetobutirato e piruvato (Yadi Liu et al., 2016). O acetolactato é utilizado na biossíntese de L-valina e L-leucina, já o acetohidroxibutirato é utilizado síntese de L-isoleucina (Yadi Liu et al., 2016). O uso de piruvato como precursor pode influenciar o ciclo do ácido cítrico, considerando o desvio do fluxo metabólico da formação de acetil-CoA, por meio do complexo da piruvato desidrogenase (PDHC), para a formação de acetolactato ou acetohidroxibutirato. A via metabólica do ciclo do ácido cítrico é muito relevante na produção de glutamato, já que seu precursor é 2-oxoglutarato, intermediário dessa via. A identificação da proteoforma metilada da subunidade pequena da AHAS somente na condição Pen-G, junto a relevância dessa enzima na produção de outros aminoácidos e o seu papel de desvio do fluxo metabólico para o ciclo do ácido cítrico, sugere que essa modificação pode ser relevante na regulação do ciclo metabólico para a produção de aminoácidos de cadeia ramificadas e para a produção de glutamato. Mais especificamente, considerando o aumento de produção de glutamato na condição Pen-G em relação ao controle, e a ausência da modificação por possível metilação na segunda condição, essa PTM pode ser responsável por uma redução da atividade da enzima AHAS.

A proteína da divisão celular SepF (UniprotAC: Q8NNN6) foi identificada por duas proteoformas diferentes, com acetilação e canônica (figura 33). Essa proteína foi identificada nas duas condições, no entanto, a proteoforma canônica só foi identificada na condição Pen-G (figura 33 A). O precursor da proteoforma canônica identificada na condição Pen-G demonstra isolamento de outras proteoformas, sem a presença da forma acetilada (figura 33 A, meio). Entretanto, a proteoforma acetilada também foi identificada na condição Pen-G e no espectro de seu precursor é possível observar a presença de dois envelopes isotópicos com diferença de

massa de aproximadamente 42 Da, corroborando a presença de ambas proteoformas de SepF na condição Pen-G. Por outro lado, na condição controle, só foi observada e identificada a proteoforma acetilada (figura 33 B, meio). Na identificação de ambas proteoformas foram detectados fragmentos que expliquem as diferentes proteoformas próximos ao N-terminal, determinando que a acetilação ocorra próxima dessa região (figura 33, inferior).

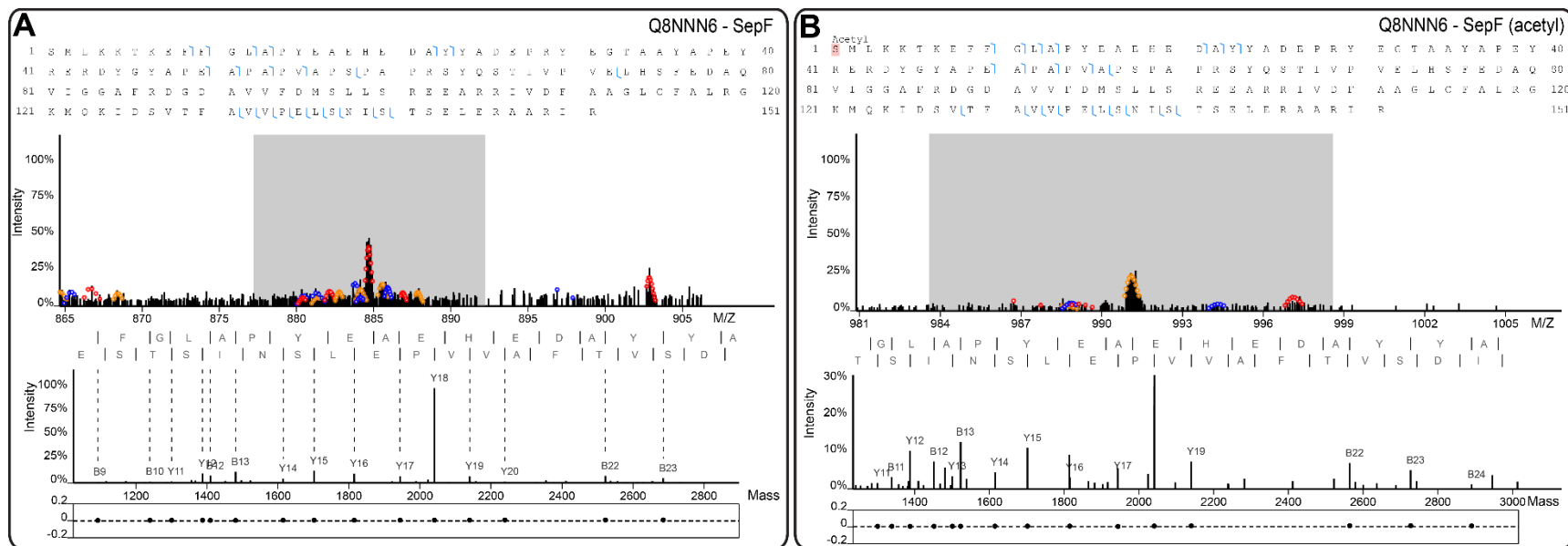


Figura 33. Análise dos espectros e seqüências das proteoformas de SepF (Q8NNN6). A seqüência de resíduos de aminoácidos das duas proteoformas de SepF são apresentadas no topo de cada figura, com a região modificada preenchida em vermelho e os fragmentos identificados são representados por traços em azul. No meio das figuras é apresentado o MS1, com o precursor e a janela de isolamento de cada proteoforma. Na parte inferior da figura é apresentado o espectro de massas, indicando os fragmentos identificadores de cada proteoforma. A) Proteoforma canônica. B) Proteoforma N-terminal acetilado.

C. glutamicum não possui diversos genes reguladores da divisão celular presentes em outras bactérias como *E. coli* e *B. subtilis*. Por exemplo, no seu genoma não são encontrados genes homólogos a *mreB*, *ftsA*, *zipA*, *zapA*, entre outros. Sendo assim, o cluster de genes de divisão celular (*dcw - division cell wall*) é considerado como uma versão simplificada do presente em outras bactérias (Letek et al., 2008). A proteína SepF é considerada como não essencial em organismos modelo, como *B. subtilis* (Gola, Munder, Casonato, Manganeli, & Vicente, 2015) mas foi descrita recentemente como fundamental na divisão celular de *C. glutamicum*, agindo como um homodímero que interage com o domínio C-terminal de FtsZ, e essas duas proteínas funcionam de maneira interdependente na formação de *z-ring* funcional (Sogues et al., 2020). A proteoforma canônica só foi identificada na condição estimulada por penicilina-G neste estudo (figura 33 A). Já foi descrita a alteração no padrão de acetilação e succinilação de proteínas de *C. glutamicum*, principalmente relacionadas ao metabolismo central, durante estímulo para produção de glutamato por tween-40 (Mizuno et al., 2016). Nesse estudo foi demonstrado a diminuição da acetilação em condição de superprodução de glutamato, possivelmente ocasionado pela atenuação no fluxo de vias produtoras de ácidos, gerando poucos substratos para acetilação (Mizuno et al., 2016). A acetilação de enzimas como a fosfoenolpiruvato carboxilase (Nagano-Shoji et al., 2017) e o supercomplexo atípico de oxoglutarato desidrogenase e piruvato desidrogenase (ODH-PDH) (Komine-Abe et al., 2021) presentes em *C. glutamicum* influenciam a sua produção de glutamato. Sendo assim, a exclusividade da proteoforma não acetilada de SepF na condição Pen-G está de acordo com o esperado pelo aumento de produção de glutamato em *C. glutamicum*. Além disso, o crescimento celular de *C. glutamicum* é gravemente impactado por estímulos para a produção de glutamato, como apresentado no estudo “Análise *bottom-up* quantitativa comparando condição controle vs condição produtora de glutamato” desse capítulo e em estudos anteriores (J. Kim et al., 2011; Niebisch et al., 2006). Considerando a influência da acetilação no metabolismo de *C. glutamicum*, a influência das diferentes proteoformas de SepF na divisão celular precisa ser melhor investigada.

Além das proteínas já apresentadas que tiveram seus espectros cuidadosamente avaliados para verificação das proteoformas, outras proteínas só foram consideradas diferentes entre as condições quando analisadas em nível de proteoformas. Por exemplo, três proteoformas foram identificadas como exclusivas da condição Pen-G e outras três da condição controle (tabela 4). Apesar de várias outras proteoformas apresentarem potencial de serem exclusivas entre as condições, a complexidade das variações de massas dificultou a interpretação desses resultados (Tabela suplementar 8). Uma alternativa para a dificuldade na interpretação das proteoformas é a utilização de programa capaz de identificar PTMs a partir do banco de dados fornecido, como o MetaMorpheus (Yuling Dai et al., 2022; Solntsev et al., 2018).

Tabela 4. Proteoformas exclusivas de cada condição.

Condição exclusiva	Proteoforma (UniprotAC_Δm)	Descrição
Pen-G	Q8NUA5_14 Da	Peptidyl-prolyl cis-trans isomerase
Pen-G	Q8NRX5_76 Da	Uncharacterized ACR
Pen-G	Q8NRQ9_14 Da	Transcription elongation factor GreA
Controle	Q8NQNO_28 Da	Secreted protein
Controle	Q8NP7_158 Da	HPr domain-containing protein
Controle	Q8NP7_176 Da	HPr domain-containing protein

Análise proteômica *top-down* pelo software MetaMorpheus

Usando a abordagem de identificação de proteoformas fornecida pelo MetaMorpheus foram identificados 1619 PrSMs, 129 proteoformas e 125 proteínas (Tabela suplementar 9). Essa redução no número de identificações pelo programa MetaMorpheus utilizando o proteoma de referência de *C. glutamicum* como banco de dados provavelmente foi ocasionada pelo baixo número de PTMs anotadas nesse banco (figura 34). Analisando e as proteínas anotadas com PTMs no proteoma de referência de *C. glutamicum* e *E. coli* pode-se observar que nenhum dos dois apresenta proteínas glicosiladas, além de que o número de proteínas com

PTMs comuns (Mod_res) e lipidação de *C. glutamicum* são muito inferiores aos de *E. coli* (figura 34)

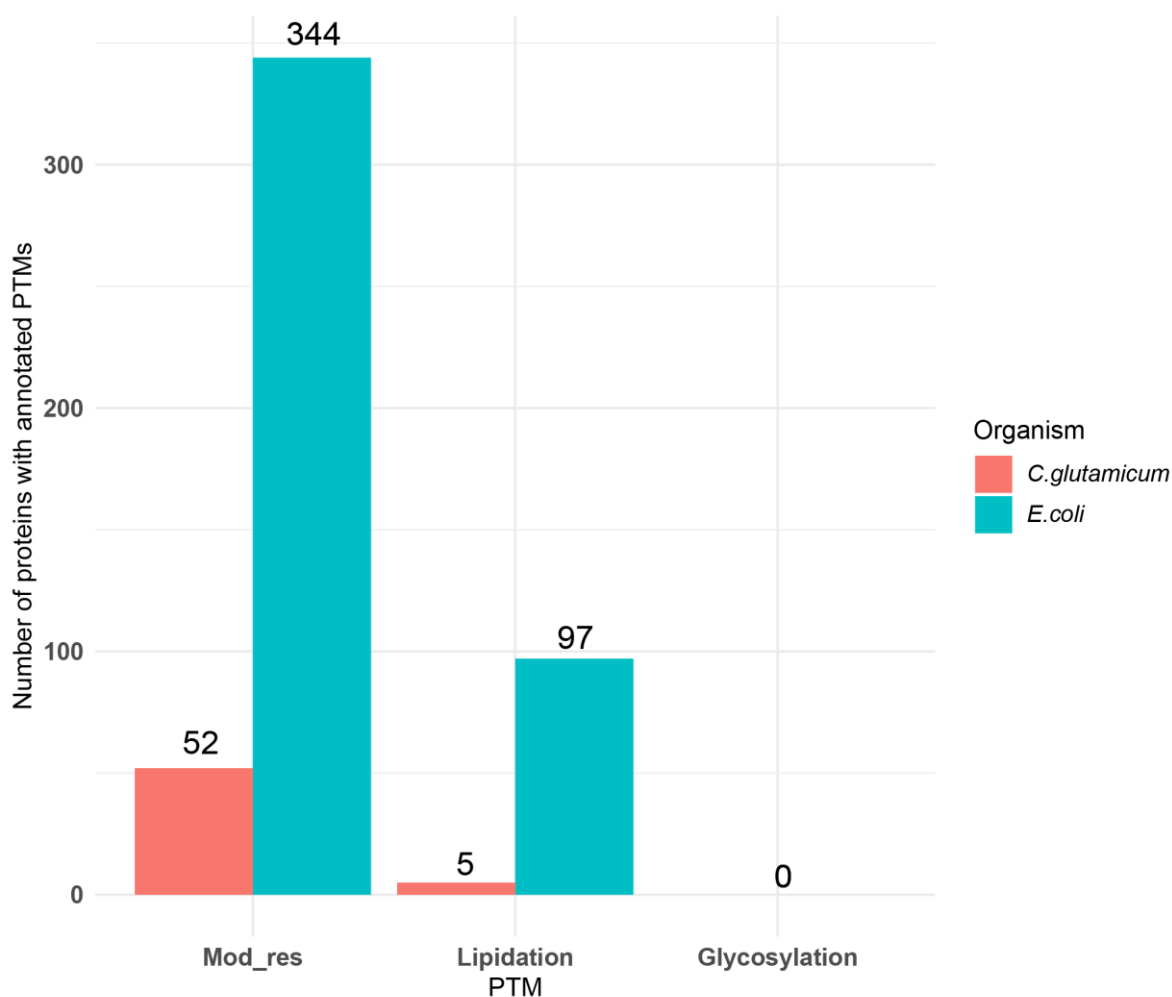


Figura 34. Proteínas dos proteomas de referência de *C. glutamicum* (UP000000582) e *E. coli* (UP000000625) anotadas com modificações pós-traducionais. As informações foram extraídas e interpretadas da base de dados UniprotKB (Bateman et al., 2021). O campo Mod_res se refere às informações armazenadas no campo *modified residue* do Uniprot, que engloba todas as PTMs, menos lipidações e glicosilações.

Esses resultados sugerem que existe um grande número de proteoformas de *C. glutamicum* que foram identificadas, mas não estão anotadas no banco de dados de referência do Uniprot. Essa característica indica que uma abordagem integrativa com a criação de um banco de dados específico anotado por meio de proteômica

bottom-up multi-enzimática com auxílio de G-PTM-D favoreceria a identificação de proteoformas ausentes no proteoma de referência do Uniprot. Essa abordagem já demonstrou seu potencial na identificação de PTMs, identificando mais de 8000 PTMs em células Jurkat (Miller et al., 2019), além de já ter sido utilizada para a construção de bancos de dados para identificação de proteoformas por análise de massa intacta (Yunxiang Dai et al., 2019) e para construção de bibliotecas espectrais para identificação por proteômica *top-down* e *bottom-up* (Yuling Dai et al., 2022).

Abordagem *bottom-up* multi-enzimática

Foram utilizadas seis enzimas com sítios de clivagens distintos para a identificação de PTMs e construção de um banco de dados. Especificamente, foram utilizadas as enzimas tripsina, Lys-C, Glu-C, quimiotripsina, Asp-N, e Arg-C. A utilização dessas enzimas para a identificação de PTMs apresentou complementariedade notória em análise de células humanas (Miller et al., 2019). A recuperação de peptídeos digeridos foi avaliada, usando o digesto de tripsina como referência, e todas as digestões obtiveram recuperação equivalente ou maior à de tripsina (tabela 5).

Tabela 5. Análise de quantidade de peptídeos obtidos após digestão multi-enzimática utilizando FASP.

Enzima	Quantidade de peptídeos (μ g)	Proporção em relação à tripsina
Tripsina	22.7	1.00
Lys-C	32.6	1.44
Quimiotripsina	29.3	1.29
Glu-C	24.2	1.07
Asp-N	23.6	1.04
Arg-C	29	1.28

Os digestos também foram analisados em nível de espectros utilizando MALDI-TOF. Assim como esperado, os espectros dos peptídeos obtidos apresentaram grande complexidade, sugerindo digestão eficiente de todas as

enzimas, além de ser observada grande diferença do perfil de peptídeos entre as enzimas (figura 35).

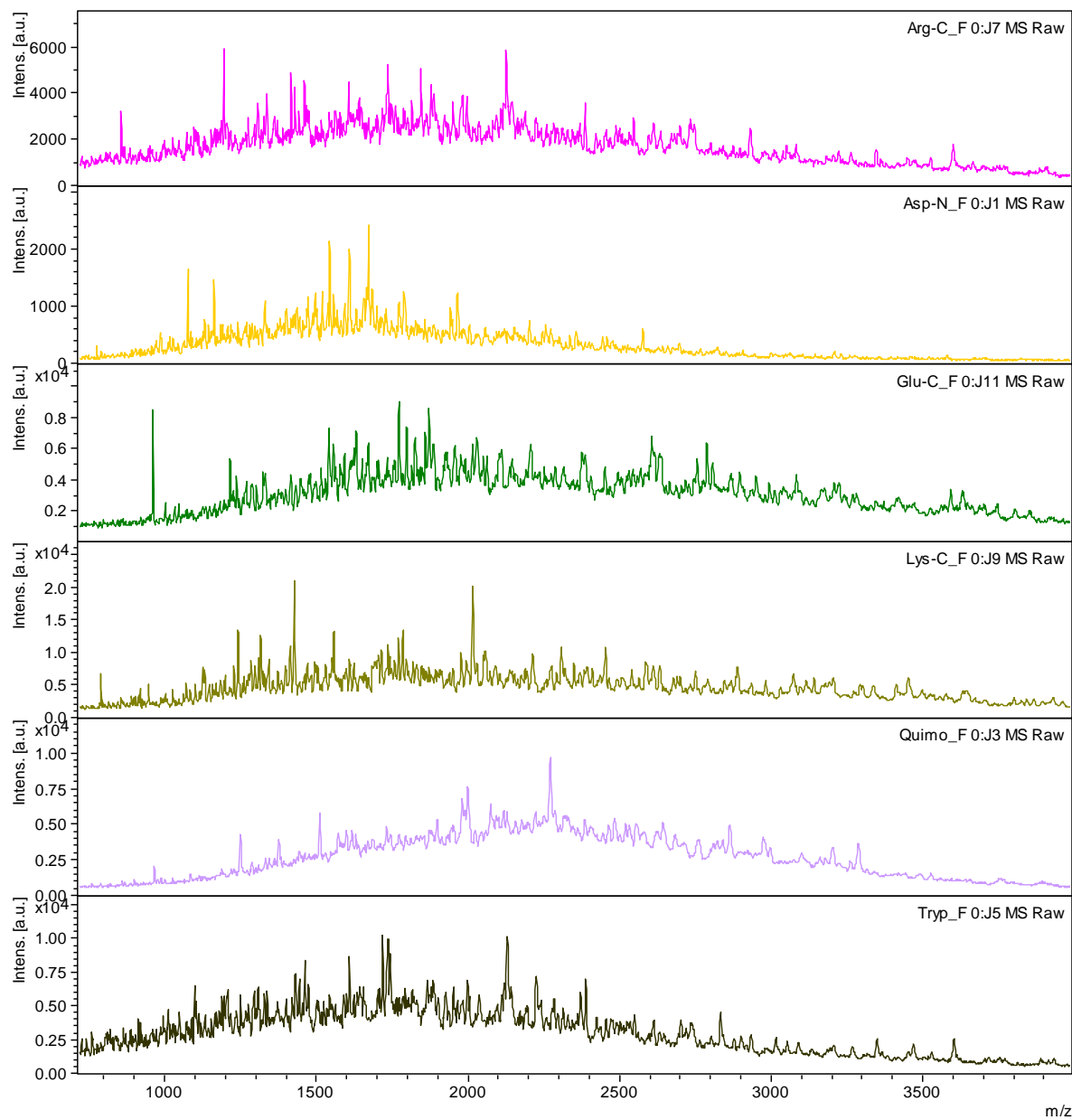


Figura 35. Análise por MALDI-TOF dos peptídeos obtidos por digestão em FASP utilizando as enzimas Arg-C, Asp-N, Glu-C, Lys-C, Quimiotripsina e Tripsina.

A análise por LC-MS/MS das 42 frações geradas resultou na identificação de 1564 proteínas e 14215 peptídeos. Mais especificamente, foram identificadas 340 proteínas e 1705 peptídeos com modificações pós-traducionais. Isso sugere que o uso dessas informações de PTMs para anotação do banco de dados e subsequente uso desse por abordagem *top-down* pode resultar em um aumento considerável do número e acurácia das identificações de proteoformas. Ao analisar as enzimas utilizadas, pode-se perceber que diferentes contribuições para o número de peptídeos modificados identificados. Mais especificamente, os peptídeos resultantes das digestões por Lys-C, tripsina, e Glu-C resultaram na identificação de mais peptídeos com modificações, enquanto a quimiotripsina foi a que identificou o menor número e as enzimas Arg-C e Asp-N alcançaram números intermediários (figura 36 A). Ademais, foram identificados peptídeos com modificações biológicas comuns e possíveis artefatos (figura 36 B).

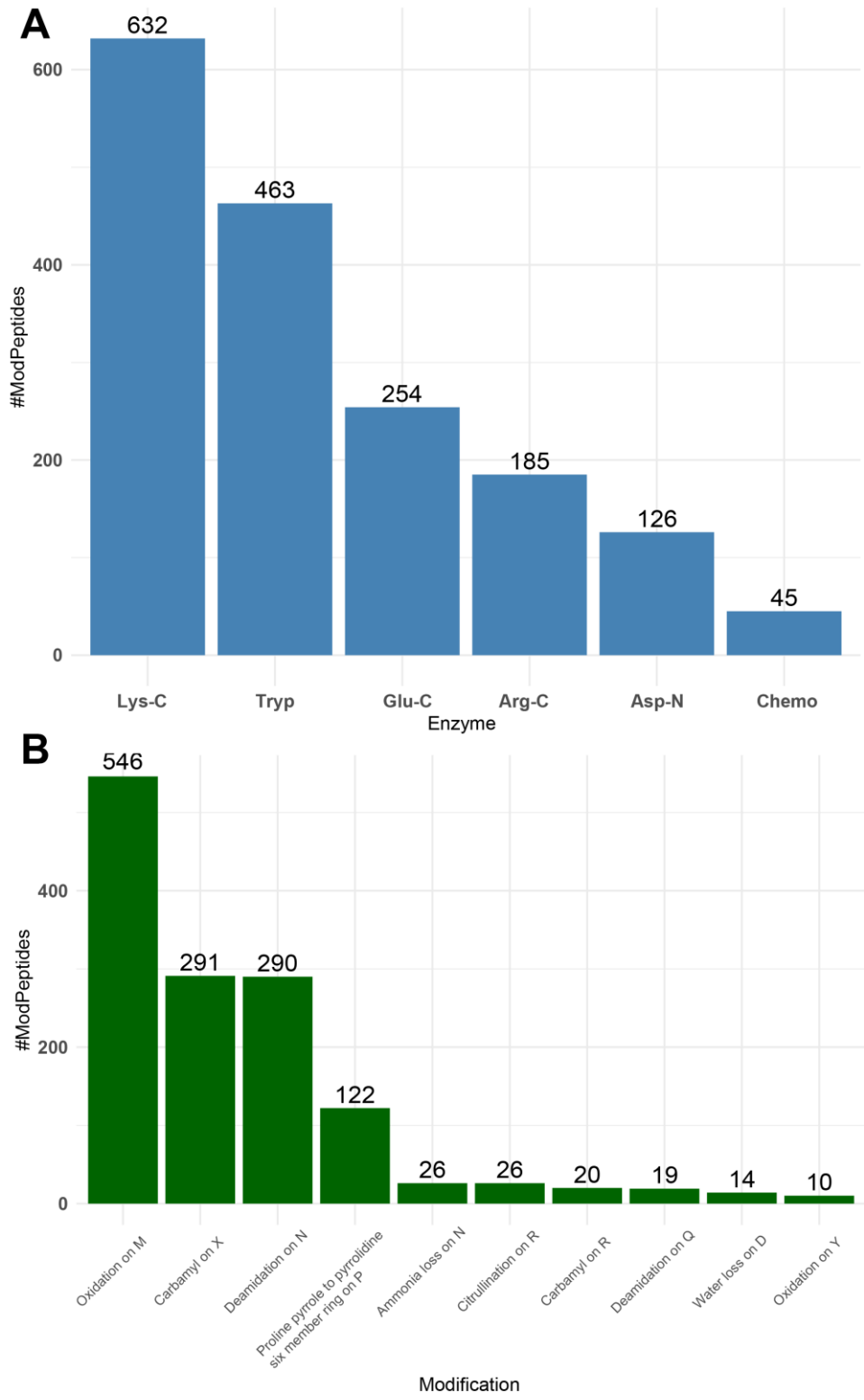


Figura 36. Análise de peptídeos modificados identificados por abordagem multi-enzimática. A) Número de peptídeos modificados identificados pelas diferentes enzimas utilizadas. B) 10 modificações mais identificadas nos peptídeos modificados.

Integração de proteômica *bottom-up* e *top-down* por meio da expansão das proteoformas do banco de dados

As análises de LC-MS/MS dos experimentos de abordagem multi-enzimática e anotação das PTMs identificadas nos peptídeos em banco de dados para subsequente uso na identificação de proteoformas por dados de proteômica *top-down* resultou na identificação de 3549 PrSMs, 243 proteoformas e 160 proteínas (figura 37). Esse resultado demonstra que o uso de abordagem *bottom-up* multi-enzimática para descoberta e anotação de banco de dados resultou em aumento considerável do número de proteínas, proteoformas e PrSMs identificadas em método de busca restrito a erro. Apesar disso, esse aumento não foi o suficiente para igualar às identificações realizadas pelo programa TopPic, que utiliza uma abordagem tolerante a erro (figura 37). Ainda assim, a otimização dos resultados apresentados para a busca restrita é promissora por apresentar meio mais conveniente de interpretação das modificações quando comparado ao TopPic.

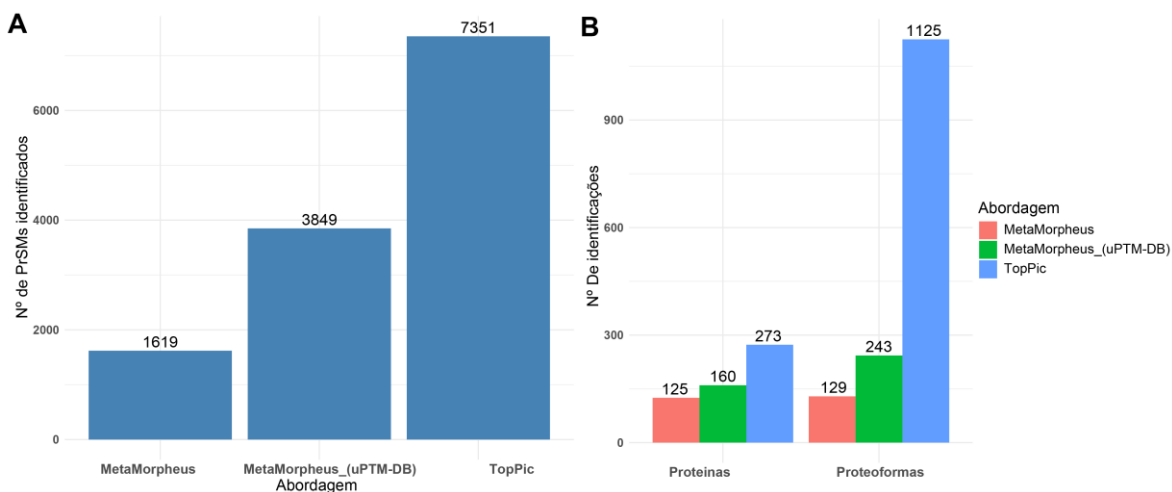


Figura 37. Comparação das abordagens utilizadas para identificação. As abordagens analisadas são a busca tolerante a erro (TopPic), busca restrita com banco de dados do Uniprot (MetaMorpheus) e busca restrita com banco de dados aumentado a partir das PTMs identificadas por abordagem *bottom-up* (MetaMorpheus_(uPTM-DB)). A) Comparação em número de espectros identificados (PrSMs). B) comparação em nível de proteínas e proteoformas identificadas.

4.3.4 Conclusões

A análise proteômica *top-down* de *C. glutamicum* comparando condição controle com indução de glutamato identificou no total 1363 proteoformas, provenientes de 382 proteínas, sendo que 279 dessas proteínas possuem Δm , sugerindo presença de modificação pós-traducional não descrita no banco de dados. Além disso, foram identificadas proteínas e proteoformas exclusivas de cada condição, sendo que as proteoformas apresentaram maior potencial para diferenciar essas condições. Mais especificamente, foram identificadas 526 proteoformas únicas da condição Pen-G, enquanto só foram identificadas 43 proteínas. Já na condição controle, foram identificadas 842 proteoformas únicas e 153 proteínas. Ademais, foram analisadas especificamente as proteínas que em uma primeira análise seriam semelhantes entre as condições, mas que ao observadas em nível de proteoformas apresentaram diferença entre as condições. Essa abordagem possibilitou a detecção de proteoformas características de uma condição (identificadas somente em uma condição) relacionadas a regulação do metabolismo de nitrogênio, síntese de aminoácidos de cadeia ramificada, divisão celular, degradação de proteínas e transcrição.

As proteoformas identificadas por proteômica *top-down* revelaram possíveis mecanismos e PTMs relacionadas à produção de glutamato induzida por penicilina-G em *C. glutamicum*. No entanto, essa identificação e caracterização foi dificultada pela interpretação das variações de massas identificadas nas proteoformas, além do baixo número de PTMs anotadas no proteoma de referência de *C. glutamicum*. Considerando essas limitações, foi feita a abordagem de integração da proteômica *top-down* à *bottom-up* por meio da identificação de PTMs utilizando a estratégia recentemente descrita de G-PTM-D unida ao potencial da abordagem multi-enzimática (Miller et al., 2019; Solntsev et al., 2018). Essa abordagem visa a construção de um banco de dados com anotações de PTMs e identificação das proteoformas por meio de busca com tolerância restrita. A análise utilizando banco de dados anotados com informações de modificações identificadas em *bottom-up* multienzimático resultou no aumento das identificações utilizando a plataforma MetaMorpheus para 3549 PrSMs, 243 proteoformas e 160 proteínas.

Capítulo V. Proteômica de *Trypanosoma cruzi*

Dedicatória

O estudo quantitativo de *T. cruzi* foi iniciado pelo colega de laboratório, e querido amigo, Agnelo Rodrigues de Souza durante seu doutorado sob orientação do prof. Dr. Carlos André Ornelas Ricart. Eu tive a honra de poder acompanhar de perto e auxiliar nas etapas finais desse projeto. Infelizmente, Agnelo faleceu em meados de 2022. Considerando nossa orientação pelo professor Carlos André e os resultados promissores já apresentados pela sua pesquisa, decidimos em conjunto finalizar os seus esforços para concluir seu trabalho e publicá-lo em forma de homenagem ao nosso colega, aluno e amigo. Sendo assim, esse capítulo de minha tese é inteiramente dedicado ao meu falecido amigo Agnelo Rodrigues de Souza e sua família, que iniciou o projeto de análise de amastigotas de *T. cruzi* e me instigou a estudar e trabalhar com esse organismo, até o desenvolvimento da análise proteômica integrando abordagem *bottom-up* e *top-down* para caracterização das proteoformas desse organismo.

5 Resumo

O *Trypanosomas cruzi* é o agente etiológico da doença de Chagas. Essa doença afeta cerca de 7 milhões de pessoas no mundo, ocorrendo predominantemente na América Latina. O protozoário *T. cruzi* possui um complexo ciclo de vida, com estágios e diferentes formas de vida em artrópodes e mamíferos, refletindo um sistema de regulação de expressão complexo. Os genes de *T. cruzi* não possuem regiões promotoras conhecidas e são normalmente transcritos de maneira policistrônica, o que resulta em regulação da expressão predominantemente em nível pós-transcricional. A relação entre a abundância de proteínas e transcritos não é sempre linear, com estimativas de ocorrer correlação de 40% entre esses dois tipos moleculares. Levando em consideração a relevância médica e os mecanismos de regulação da expressão de *T. cruzi*, foram conduzidos dois estudos proteômicos com o objetivo de aprimorar os conhecimentos relacionados a divisão celular desse protozoário: 1) Análise quantitativa de formas de vida amastigota axênicas e intracelulares em condições replicativas e não replicativas (apresentado na forma de manuscrito); 2) Integração de proteômica *bottom-up* multi-enzimática e *top-down* para caracterização de proteoformas da forma de vida epimastigota de *T. cruzi*. Esses estudos resultaram na verificação do enriquecimento de amastigotas intracelulares a partir de células HeLa, utilizando colunas de DEAE-celulose, o que possibilitou detectar diferenças entre parasitos intracelulares e extracelulares. Essa comparação revelou diferenças entre esses dois tipos celulares relacionadas com a tradução, metabolismo energético e de aminoácidos. Além disso, duas proteínas associadas ao cinetoplasto apresentaram indícios de estarem envolvidas no processo de replicação de amastigotas. Alternativamente, a análise proteômica *top-down* de epimastigotas identificou indícios de substituições em algumas proteínas, por exemplo na HSP10. Por outro lado, foi possível observar que melhorias podem ser feitas no processamento de amostras, aquisição de dados e identificação de proteoformas para análises de proteínas intactas de *T. cruzi*.

5.1 Manuscript: Proteomics of axenic and intracellular amastigote lifeforms of *Trypanosoma cruzi* in replicative and non-replicative conditions

Reynaldo Magalhães Melo, Agnelo Rodrigues de Souza, Izabela Dourado Bastos Charneau, Flavia Nader, Jaques Miranda Ferreira de Souza, Marcelo Valle de Sousa, Wagner Fontes, Carlos André Ornelas Ricart

Keywords: *Trypanosoma cruzi*, proteomics, amastigote, replication, intracellular amastigote

5.1.1 Abstract: Proteomics of axenic and intracellular amastigote lifeforms of *Trypanosoma cruzi* in replicative and non-replicative conditions

Trypanosoma cruzi is the etiologic agent of Chagas disease, which affects about 6-7 million people worldwide. *T. cruzi* has a complex life cycle with different life forms inside insect vectors and mammalian hosts. The biological plasticity of the parasite throughout its life cycle is reflected by changes in the proteome. Here, we performed a quantitative proteome analysis of amastigotes, the intracellular replicative form, isolated from infected cells as well as from axenic cultures in replicative and non-replicative conditions. Intracellular amastigotes (AI) were obtained by mechanical disruption of *T. cruzi*-infected Hela cells followed by DEAE-cellulose chromatography. Axenic amastigotes (AD5) were prepared by differentiation of trypomastigotes for 9 h of incubation in DMEM pH 5.0. AD5 cells were then incubated in LIT (AL7) or DMEM (AD7) media with pH 7.2. Manual cell counting showed that AD5 cells was not able to replicate while AL7 and AD7 were replicative. Proteins from all amastigote samples were extracted, digested with trypsin and the resulting peptides subjected to LC-MS/MS followed by protein identification and quantification using bioinformatics tools. Proteomic analyses identified 3134 protein groups and showed remarkable difference between AD5 and

the replicative conditions (AL7 and AD7). Cluster analysis revealed proteins with higher abundance in AD5 which were related to microtubule and dynein complex. Moreover, kinetoplast DNA-associated proteins and trans-sialidases presented increased abundance in replicative conditions, suggesting relevant role of these in amastigote division.

5.1.2 Introduction

The protozoan *Trypanosoma cruzi* is the etiologic agent of Chagas disease, an illness that affects about 6-7 million people worldwide, mostly in Latin America, causing approximately 12,000 death per year (Pérez-Molina & Molina, 2018; WHO, 2022). *T. cruzi* has a complex life cycle with life stages inside insect vectors (epimastigote and metacyclic trypomastigote) and mammalian hosts (bloodstream trypomastigote and amastigote) (W. Souza, 2005). Amastigotes and trypomastigotes are the replicative and infective forms, respectively, of *T. cruzi* inside mammalian hosts (W. Souza, 2005; Zuma, dos Santos Barrias, & de Souza, 2020).

The amastigote life form of *T. cruzi* is derived from trypomastigotes inside mammalian cells. After penetrating host cells, trypomastigotes are restrained in the parasitophorous vacuole (Moretti et al., 2020). Subsequent to escaping the parasitophorous vacuole, the parasite transforms to amastigote, a process known as amastigogenesis, and replicate through binary fission (Pérez-Molina & Molina, 2018). Further, *T. cruzi* amastigogenesis can be conducted *in-vitro* through acidic pH conditions (Tomlinson, Vandekerckhove, Frevert, & Nussenzweig, 1995). Amastigotes obtained through *in-vitro* differentiation using acidic pH and the extracted from cells share morphological characteristics and stage-specific antigen (Ssp-3) (Tomlinson et al., 1995). Some morphological similarities presented between these two types are oval-shaped bodies, longitudinal cracks and short protruding flagellum. Despite these resemblances, intracellular amastigotes present smooth surface, and extracellular acidic pH-induced amastigotes have a finely

grained surface (Navarro, De Lima, Askue, & Contreras, 2003). Further, recent study have assessed differences in the cytometry profile of Ssp-4 expression from axenic amastigotes and intracellular amastigotes, suggesting a more heterogeneous population of axenic obtained amastigotes (Batista, Kessler, Eger, & Soares, 2015).

Intracellular amastigotes of *T. cruzi* can be isolated through Percoll discontinuous gradients (Gamarro, Osuna, Castanys, Pérez-López, & Ruiz-Pérez, 1985), anion exchange chromatography (Marques, Marques, Nakayasu, & Almeida, 2011), and more recently a method using nitrogen decompression and differential centrifugation have been described (Batista et al., 2015).

Beyond the morphological complexity represented by its life cycle, *T. cruzi* has an interesting gene expression regulation machinery. This parasite lacks regulatory promoter in its genes; therefore, great part of its coding genes is constitutively transcribed. Moreover, this transcription is performed polycistronically and the resulting mRNA should be processed to become mature (Herrerros-Cabello et al., 2020), consequently, great part of the gene expression regulation is performed post-transcriptionally in this organism (Clayton, 2019).

The biological plasticity of *T. cruzi* is reflected by its complex life cycle, its capacity of infection of multiple organisms and molecular mechanisms presented. Proteomic analysis has already provided important information regarding molecular mechanisms of this organism. For instance, epimastigote cell cycle was recently studied evidencing differences in translation and oxidative phosphorylation, beyond changes in phosphorylation status of peptides originated from proteins related to pyruvate and carbon metabolism (Santos Júnior et al., 2021). Moreover, the amastigogenesis was investigated through phosphoproteomic analysis, suggesting the influence of cAMP signaling pathway and protein kinases in this process (Queiroz et al., 2014). Other study has focused on *T. cruzi* metacyclogenesis (epimastigote differentiation to metacyclic trypomastigote), demonstrating the relevance of a possible RNA binding protein in early stages of this differentiation, and several regulated proteins comparing early stages of the differentiation with final the stage of this process (de Godoy et al., 2012).

As mentioned above, the power of proteomics to describe *T. cruzi* molecular mechanisms has been demonstrated several times before. However, to our knowledge, there is no proteomic study focused on amastigote biology and the differences between axenic and intracellular cultured cells of this parasite. Therefore, in this study we performed a quantitative proteomic analysis of amastigotes isolated from infected HeLa cells as well as amastigotes from axenic cultures in replicative and non-replicative conditions (Fig. 1), describing significant changes between intracellular and axenic amastigotes, primarily related to energy metabolism and translation machinery. Additionally, significant differences were observed among replicative and non-replicative conditions, indicating the relevance of proteins from the multi-genic family of trans-sialidases and kinetoplast DNA-associated proteins in the replication *T. cruzi* amastigotes.

5.1.3 Methods

Growth and purification of intracellular amastigotes

HeLa cells were cultured in DMEM (Invitrogen) supplemented with 10% fetal bovine serum and streptomycin/ampicillin (100 µg/mL) at 37 °C with 5% CO₂ atmosphere. Infection of HeLa cells was performed in biological triplicate through the addition of *T. cruzi* (strain Y) trypomastigote forms (40:1 HeLa cell) for 12 h. Next, the medium was removed to remove extracellular parasites. HeLa cell layer was then washed with PBS (0.01 M phosphate buffer, 0.0027 M KCl, 0.137 NaCl, pH 7.4). Fresh DMEM was added, and cells were incubated for 60 h at 37 °C with 5% CO₂ atmosphere. After incubation, cells were washed with PBS and detached from bottle using Verseno (trypsin 0.05 %, PBS, EDTA 2mM). Detached cells were centrifuged at 400 x g for 10 min, washed with PBS, and suspended in 2 mL of the same buffer. Approximately 100 µL of 0.1 mm glass beads were added to the suspension of infected cells, and host cells were lysed through vortex for 90 s. After host cells disruption, 13 mL of PBS was added to the cell suspension, and this was submitted to DEAE-cellulose chromatography. The chromatography column was assembled using a syringe obstructed with glass wool, containing 10 mL of DEAE-

cellulose resin, with the slurry pH adjusted to 7.2. The cell suspension was passed through the column, which retained the host cell debris, and the flow-through of the column was centrifuged and washed two times with PBS. An aliquot of the PBS suspension with the flow-through was assessed for the presence of amastigote through counting in a hemocytometer. The resulting intracellular amastigotes (AI) was kept in -80 °C for further analysis.

Fluorescence microscopy

Slides were briefly washed with 100% methanol. Approximately 1.25×10^5 cells were put in a slide and incubated with poly-lysine for 30 min. Next, cells were permeabilized using Triton 0.1% in PBS and incubated for 10 min, followed by three washes with PBS. Then, 100 μ L of enhancer (Invitrogen) were added to the slides and incubated for 30 min. For antibody treatment, slides were washed with PBS, and blocked by PBS/milk 5% for 30 min. Then, cells were incubated with anti-tubulin (commercial informations) for 2 h, and subjected to three PBS washes for 3 min each. Following, cells were incubated with anti-TcSCC1 for 2 h in the dilution 1:100 in PBS/milk 1%, washed three times with PBS, and incubated with the secondary antibody (goat anti-rabbit IgG) conjugated with Alexa Fluor 488 (Invitrogen) in the dilution 1:400, for 2 h. At final, slides were washed three more times with PBS, followed by H₂O Milli-Q, and assembled with ProLong® Gold antifade reagent with DAPI (Invitrogen). Assembled slides were observed in by confocal microscopy using the Leica TCS SP5 (Leica Microsystems, Wetzlar, Germany).

Axenic amastigotes cultures

Axenic cultures were also performed with *T. cruzi* Y strain in biological triplicate. Briefly, these were obtained by the differentiation of trypomastigotes, right after its outbreak (96 -120 h after infection). Trypomastigotes were incubated for 9 h in DMEM, without FBS (pH5.0) at 37 °C under atmosphere of 5 % CO₂ (Tomlinson et al., 1995). After incubation, the cell culture was centrifuged at 2500 x g for 10 min, and the resulting pellet washed twice with PBS, and stored at -80 °C for future analysis (AD5). Replicative axenic amastigotes were obtained after the same

amastigogenesis process, followed by incubation in DMEM pH 7.2 (AD7), supplemented with 10 % FBS (Gibco), or LIT pH 7.2 (AL7), supplemented with 10 % FBS (Gibco), 2 mg/mL hemin, 100 µg/ µL ampicillin, 30 µg/ µL streptomycin for 18 h, at 37 °C, under 5 % CO₂ atmosphere (CAMARGO, 1964). Next, resulting amastigotes were centrifuged at 2500 x g for 10 min, washed twice by PBS, and stored at -80 °C for further analysis. In addition, the amastigotes present in AD7 and AL7 were counted in a hemocytometer, after 18 h, 42 h, and 72 h of incubation.

Protein extraction and sample preparation for mass spectrometry analysis

Amastigote cells were disrupted using 100 µL of 0.1 % RapiGest SF Surfactant (Waters) diluted in 0.04 M TEAB, supplemented with protease and phosphatase inhibitor (Sigma-Aldrich), under ultra-sonication tip. After cell lysis, extracted proteins were quantified through Qubit (Invitrogen), and digestion procedure was performed according to RapiGest SF Surfactant (Waters) instructions. Briefly, dithiothreitol (DTT) was added to protein samples to final concentration of 5 mM, followed by incubation at 60 °C for 30 min. Then samples were cooled to room temperature and submitted to alkylation, by addition of iodoacetamide (IAA) to final concentration of 15 mM and incubation in the dark for 30 min. In the end, trypsin was added (1:100 protein) and incubated at 37 °C for 18 h. The resulting peptides were acidified with TFA to final concentration of 0.5 %, and incubated at 37 °C for 45 min. Next, the solution was centrifuged at 13.000 rpm for 10 min, and the supernatant was submitted to desalinization. Salts removal was performed using a slurry of POROS R2:OLIGO R3 (1:1), retained in a stage-tip by an SDB disc. The desalinized peptides were quantified, and 1 µg of each sample was submitted to LC-MS/MS.

LC-MS/MS

The LC-MS/MS analysis was performed using a nano-UHPLC Dionex Ultimate 3000 coupled to Orbitrap Elite™ mass spectrometer (Thermo Scientific).

Briefly, the chromatography was carried out with homemade packed columns: trap-columns (Reprosil-Pur 120 Å pore C18, 5 µm, 100 µm × 3 cm, Dr. Maich, Ammerbuch, Germany), coupled to the analytical column (75 µm × 35 cm) packed with the same particles. The mobile phases used were solvent A (2% acetonitrile, 0.1 % formic acid), and solvent B (0.1% formic acid in acetonitrile) with flow rate of 0.25 µL/min, under the gradient: 0-10 min 5 % B, 10-35 min 9 % B, 35-130 min 25 %, 130-180 45 % B, 180-181 min 85 % B, 181-191 min 85 % B, 191-192 min 2 % B, 192-210 min 2 % B.

The mass spectrum acquisition was performed in data-dependent mode (DDA), fragmenting the 15 most intense peptides by Higher Collisional Energy Dissociation (HCD), with normalized collision energy of 35, default charge state of 2, activation time of 0.1, and isolation width of 2 *m/z*. The resolution for MS1 and MS2 were 120,000 FWHM and 15,000 FWHM, respectively.

Three biological replicates from four conditions (AD5, AD7, AL7, AI) were analyzed with three technical runs, generating 36 raw files.

Protein identification and quantification

Protein identification and quantification were performed in Metamorphus (Solntsev et al., 2018) with UniProt *T. cruzi* CL Brener reference proteome database. For AI samples a parallel identification was performed using a concatenated database of *T. cruzi* and *Homo sapiens* reference proteome. The first step was the calibration of spectra with default parameters, resulting in calibrated MS1 and MS2, that were searched for common biological PTMs, using the Global Post-Translational Modification Discovery (G-PTM-D) strategy, which generated databases with annotation of possible PTMs, that was used in the identification and quantification of proteins. The identification was performed with the following parameters: MS1 and MS2 tolerances were defined by the calibration of each run, maximum miss cleavage = 2, max mods per peptide = 2, minimum peptide length = 7, require at least two peptides to identify a protein, treat modified peptides as different peptides, initiator methionine as variable, oxidation of methionine as variable modification,

carbamidomethyl of cysteine as fixed modification. The quantification was carried out using the flashLFQ algorithm (Millikin, Solntsev, Shortreed, & Smith, 2018) with match between runs activated and median normalization performed and 5 ppm peak finding tolerance. Peptides and proteins identification were filtered by Metamorphus score (MM Score > 5), PEP-Q.value (< 0.01), and at least one unique peptide.

Statistical analysis and biological annotations

Quality control and statistical analysis were performed mainly in pmartR (Stratton et al., 2019), using Rstudio (RStudio Team, 2021), and R language (R Core Team, 2022). Briefly, after filtering, all quantified proteins were loaded in Rstudio and abundances data were \log_2 transformed. After transformation, proteins were filtered out by the number of missing values (number of N.A. < 4). After filtering, the conditions and replicates quality control were evaluated by probabilistic principal component analysis (PPCA), and correlation heatmap. Next, the statistical difference between conditions was performed by ANOVA, using bonferroni multiple test correction, with q-value threshold of 0.05. The regulated proteins without missing values were submitted to two cluster analysis, a variance sensitive clustering (VSClust) (Schwämmle & Jensen, 2018), and a hierarchical clustering of samples using the heatmaply R package (Galili, O'Callaghan, Sidi, & Sievert, 2018). Moreover, after Uniprot accession conversion to GeneID, using the uniprotKB database (Bateman et al., 2021), protein abundances were submitted to enrichment of KEGG pathways through the R package gage (Luo et al., 2009). The different conditions were compared as groups, considering enriched pathways those with q-value < 0.1.

Annotation of proteins was done using shinyGO (Ge et al., 2020), tritryp (Aslett et al., 2010), and uniprot blastp tool against swissprot database (Bateman et al., 2021).

5.1.4 Results

This study was performed through the proteomic analysis of *T. cruzi* amastigotes obtained through different methods, comparing axenic conditions (replicative and non-replicative) with intracellular obtained amastigotes (Fig. 1).

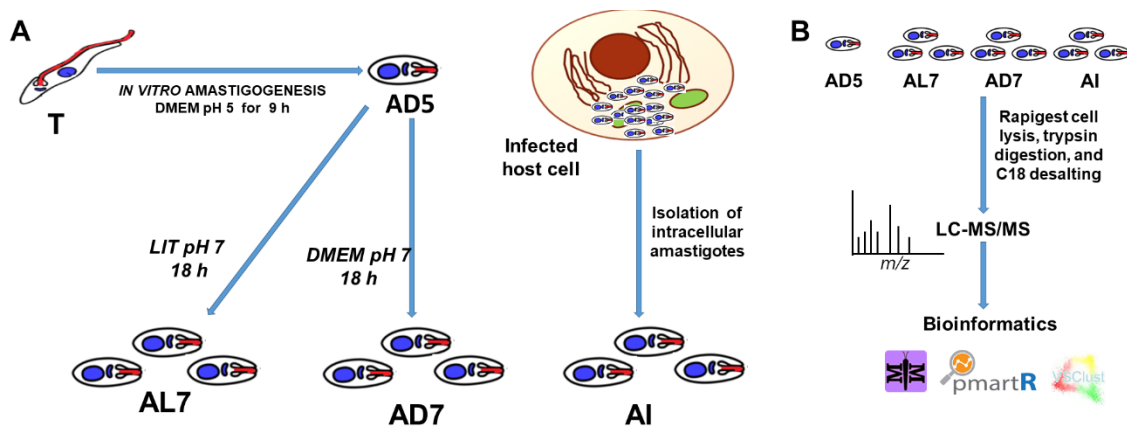


Figure 1. Experimental design for the analysis of intracellular and axenic amastigotes. A) Growth conditions for obtaining not replicative axenic amastigotes (AD5), replicative axenic amastigotes (AL7, AD7), and intracellular amastigotes (AI). B) Sample processing for LC-MS/MS and proteomic analysis.

Analysis of intracellular amastigote purification by DEAE-cellulose chromatography

Figure 2A displays a fluorescence microscopy image of a HeLa cell containing a high number of intracellular amastigotes life forms 72 h post-infection. This time was optimized to obtain a larger number of amastigote cell forms inside HeLa cells, with minimal disruption of these (data not shown). Thus, we chose this time point for the purification of intracellular amastigotes using mild cell lysis followed by DEAE-cellulose chromatography in order to separate the parasites from HeLa cells debris.

The efficiency of DEAE-cellulose for removal of HeLa cell proteins from *T. cruzi*-infected (72 h.p.i) and non-infected samples (HeLa control) was assessed by SDS-PAGE (Fig. 2B). The signal of HeLa control protein profile (Fig. 2B lane 2)

diminished drastically when this sample was submitted to DEAE-cellulose chromatography (Fig. 2B lane 1) indicating that HeLa debris was removed from intracellular amastigote (AI) fraction. On the other hand, Fig. 2B lane 3 shows a protein profile after DEAE chromatography of infected HeLa cells that are due mostly to amastigote proteins. Finally, Fig. 2B lane 4 displays proteins from both HeLa and amastigotes.

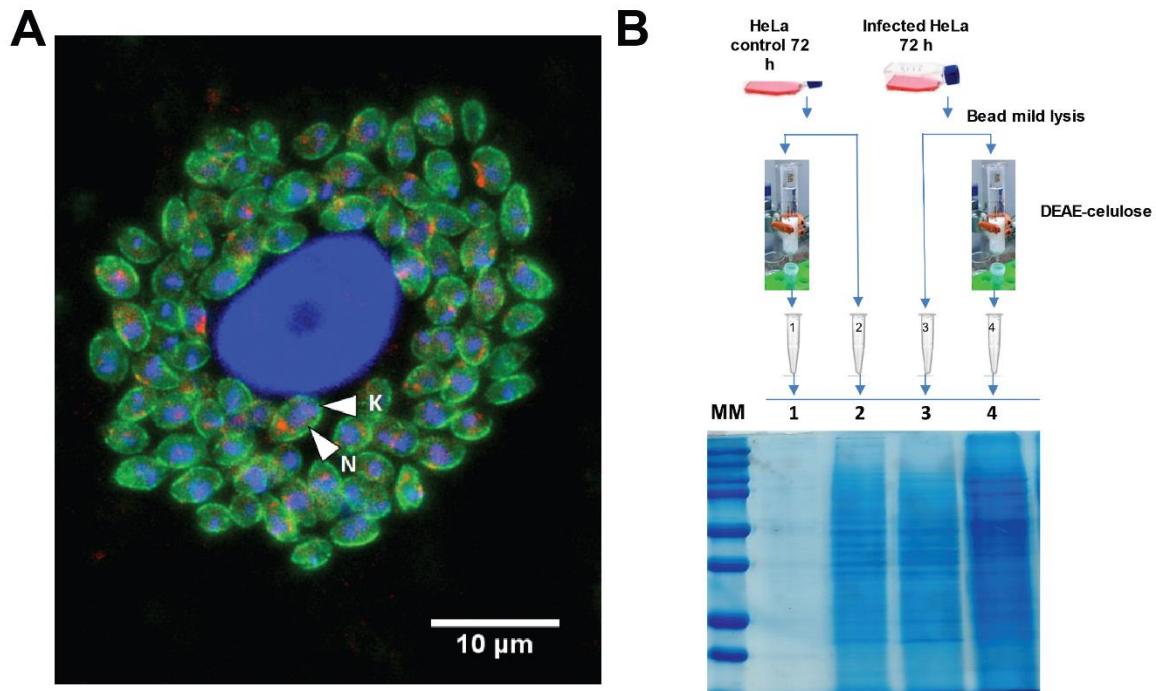


Figure 2. Enrichment of intracellular amastigote proteins A) Fluorescence microscopy analysis of HeLa cells infected with *T. cruzi* (72 h.p.i.) showing the kinetoplast (K) and nucleus (N) of amastigote cells. B) SDS-PAGE for the analysis of DEAE-cellulose purification of intracellular amastigotes present in HeLa cells.

Proteomic analysis supported the efficiency of the *intracellular amastigotes* enrichment procedure. Overall, 2068 protein groups with at least one unique peptide from *T. cruzi* and *H. sapiens* were identified (supplementary table S1). Overall, 245 protein groups belonged to *H. sapiens* and 1823 to *T. cruzi*. Therefore, approximately 90% of the protein groups identified belongs to the parasite (Fig.3 A).

The human proteins identified were enriched to the biological processes terms TCA cycle, aerobic respiration, mitochondrial transmembrane transport (Fig. 3 B). In agreement, the enriched cellular component terms were mitochondrial nucleoid, mitochondrial proton-transporting ATP synthase complex, and nucleoid (Fig.3 C). Overall, the enriched terms suggest the presence of HeLa mitochondrial proteins in the intracellular amastigote enrichment. On the other hand, the enrichment analysis of *T. cruzi* proteins did not show significant enriched terms of biological process and cellular component, suggesting that the purification did not have undesired distinction to specific amastigote components. A recent study of *Leishmania mexicana* have analyzed intracellular amastigotes extracted from infected macrophages and only around 30% of total peptides were from the parasite (Pacakova, Harant, Volf, & Lestinova, 2022). This supports the enrichment of intracellular amastigotes performed in this study, which achieved 90 % *T. cruzi* proteins. Considering the importance of host cells in the biology of parasites, this method can be used in new discoveries regarding *T. cruzi* biological processes in mammalian hosts.

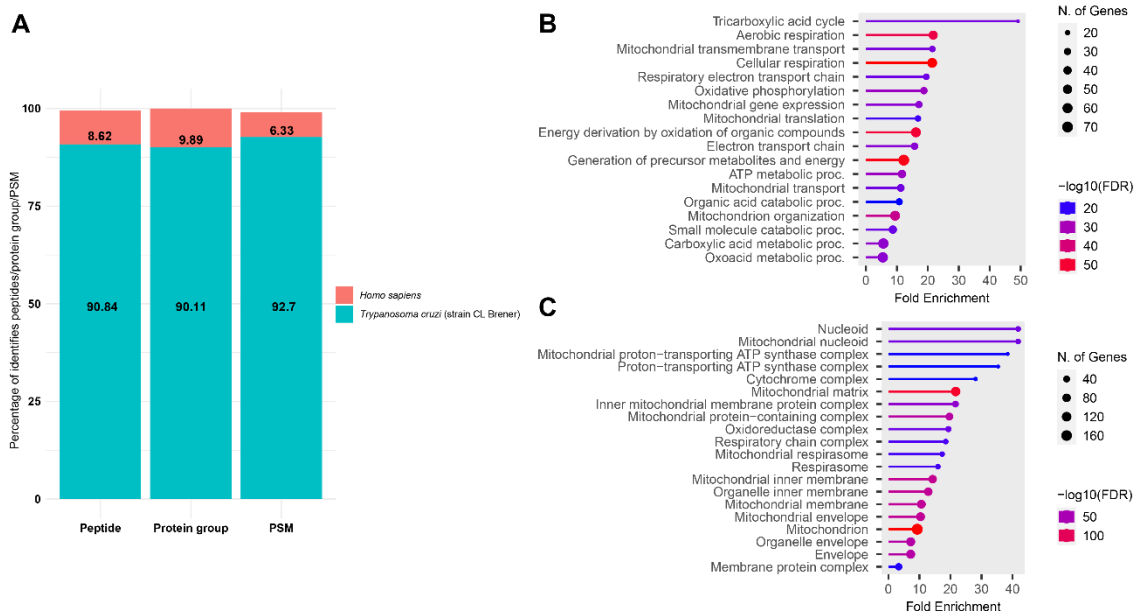


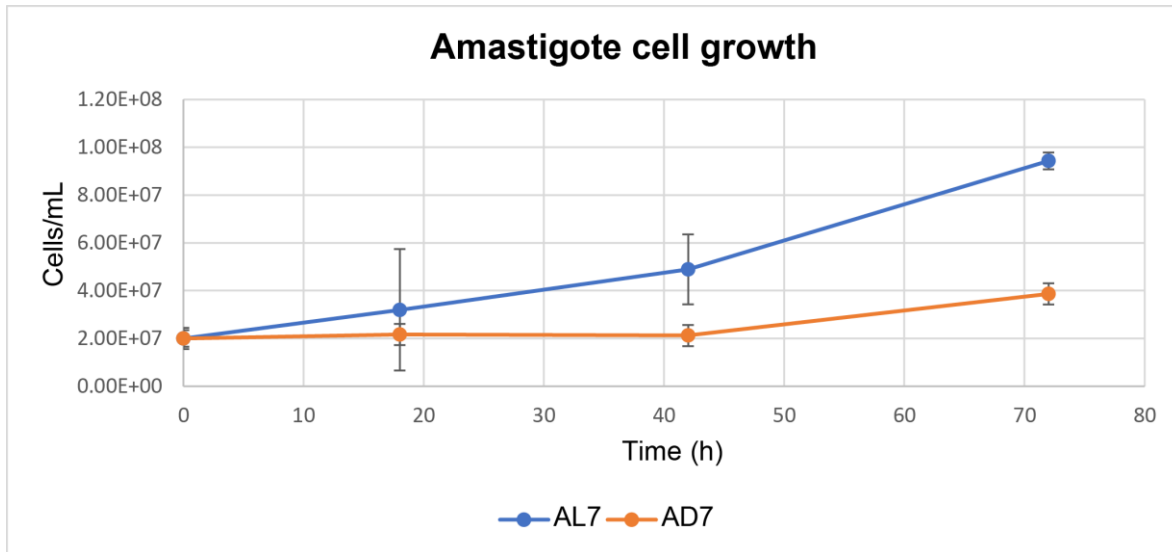
Figure 3. Analysis of intracellular amastigotes purification through identified proteins. A) Percentage *T. cruzi* or *H. sapiens* identified PSMs, peptides and protein groups. The number of peptides spectrum matches (PSM), peptides and non-redundant protein groups were counted for each species after the removal of decoy

and contaminant identifications. B) Gene ontology biological process enrichment analysis of *H. sapiens* identified proteins. C) Gene ontology cellular component enrichment analysis of *H. sapiens* identified proteins. Enrichment analysis performed by ShinyGO (v0.77).

Other methods have been proposed for the purification of *T. cruzi* amastigotes (Carvalho & Souza, 1983; Gamarro et al., 1985). A more recent study has demonstrated the efficient isolation of intracellular amastigotes from VERO cells using nitrogen cavitation (Batista et al., 2015), with viable endocytosis. However, this method requires specialized instruments, such as cell pressures for cell lysis (Batista et al., 2015).

Quantitative proteomic analysis of *T. cruzi* amastigote forms in replicative and non-replicative conditions

The replication characteristic of axenic amastigotes was observed through cell counting in a hemocytometer during 72 h of incubation at 37 °C under atmosphere of 5 % CO₂, demonstrating replication of amastigotes in DEMEM pH 7, and LIT pH 7 (supplementary figure 1). AD5 cells did not replicate (data not shown). On the other hand, AL7 and AD7 were replicative, being the growth of AD7 observed only after 42 h.



Supplementary figure S 1. Number of amastigotes counted in LIT pH 7 and DMEM pH7 after 72 h of incubation. Mean values of biological triplicate was used and error bars represent the standard deviation of these replicates.

The proteomic analysis of *T. cruzi* amastigotes in different conditions resulted in the identification of 3134 (Supplementary table S2) unique protein groups below 1% FDR, and after quality filtering for quantitative analysis 3043 protein groups remained (Supplementary table S3). Probabilistic principal component analysis (PPCA) of protein groups abundances revealed the close grouping of AD5 biological replicates and its separation from the replicative conditions (AD7, AL7, AI) (Fig. 4 A). Moreover, biological replicates from AL7 and AD7 showed some proximity in the PPCA plot, suggesting some degree of similarity, or less difference when compared to other conditions (Fig. 4 A).

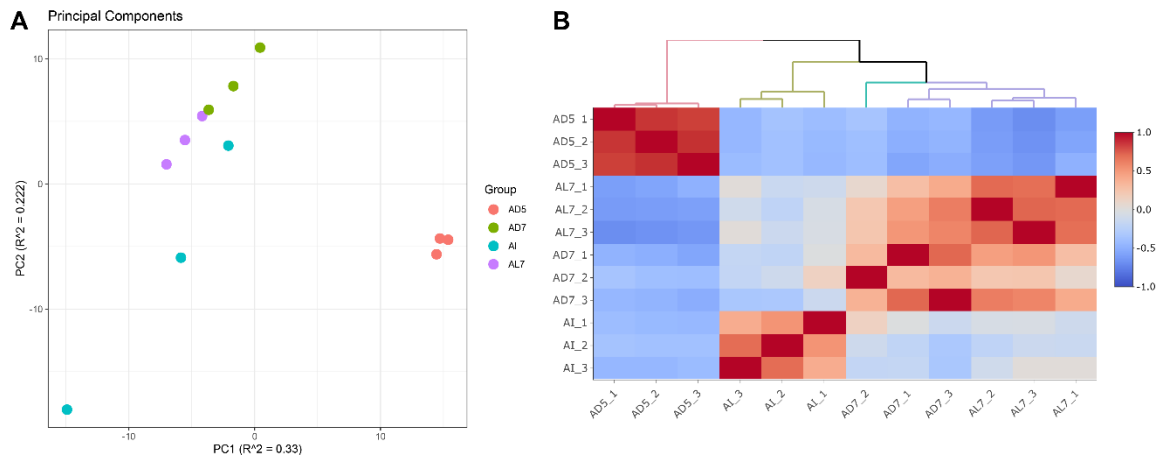


Figure 4. Quality control and correlation analysis of amastigote samples. A) Probabilistic principal component analysis (PPCA) obtained through dimensions' reduction, summarizing the information's of protein groups abundances and number of missing values. B) Heatmap of correlations between samples and hierarchical cluster of samples Euclidean distances.

The ANOVA test followed by bonferroni multiple-tests correction showed 1106 regulated protein groups among all conditions (q-value < 0.05) (supplementary table S4). As suggested by the PPCA plot, the clustering and correlation heatmap of the samples demonstrated remarkable difference between AD5 and replicative conditions, clustering all non-replicative samples distant from every replicative sample in the hierarchic tree (Fig. 4 B). Moreover, the greater correlation of axenic replicative conditions (AD7 and AL7) is evidenced by this analysis, clustering these two conditions in the same group, with the exception of AD7_2 sample (Fig. 4 B). Despite closer to the axenic replicative conditions, intracellular amastigotes (AI) clustered in a separate group (Fig. 4 B), suggesting differences between axenic and intracellular amastigotes.

In agreement with preceding analysis, the AD5 group presented the highest number of regulated protein groups between comparisons, and the comparisons between AI and axenic replicative conditions demonstrated a considerable number of regulated protein groups (Fig. 5 A). Additionally, a low number of protein groups was identified as regulated between AL7 and AD7 (Fig. 5 A). The comparisons between AD5 and replicative conditions also presented a large number of regulated

protein groups with great \log_2FC ($\log_2FC > 2$ or $\log_2FC < -2$) (Fig. 5 B). These great fold changes suggest that the cells are going through drastic changes, when comparing non-replicative conditions to replicative conditions. In AI comparison with axenic replicative conditions (AD7 and AL7), there were also some protein groups with great \log_2FC (Fig. 3 B), suggesting that there are significant differences between axenic replicative conditions and intracellular amastigotes.

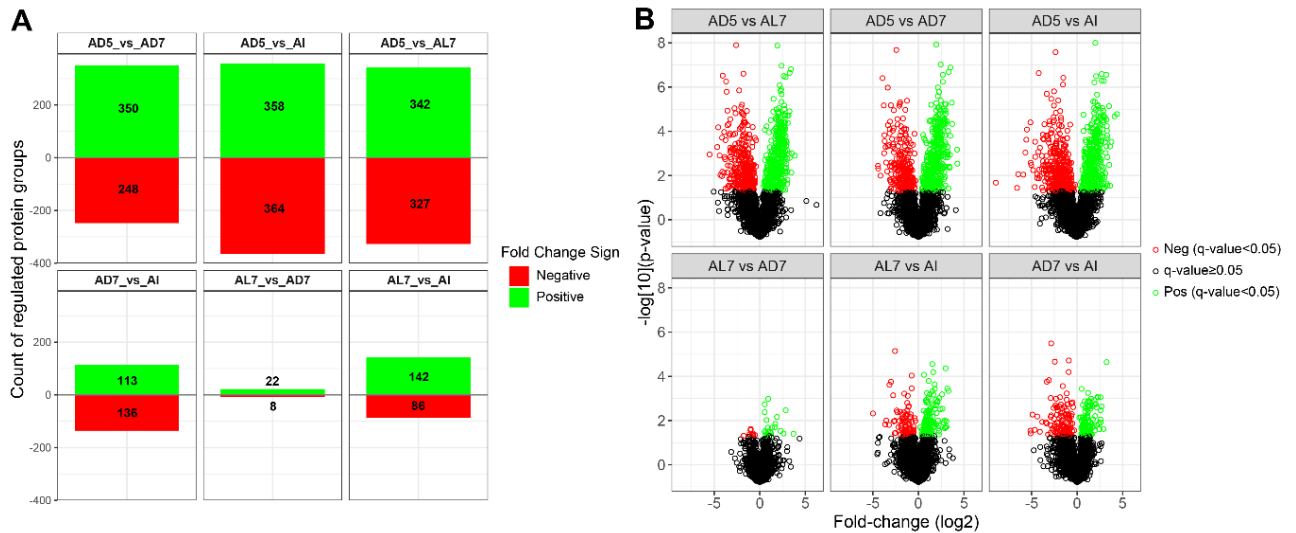


Figure 5. Analysis of regulated proteins. Quantified protein groups were submitted to ANOVA test followed by bonferroni correction, and those with q-value < 0.05 were considered as regulated. A. Number of regulated protein groups in each comparison and its fold change sign, where the first group of the title is the reference. B. Volcano plot showing the $-\log_{10}p$ -value and the fold change, the first group of the titles are also the reference in this case.

The functional enrichment analysis of protein profiles revealed several enriched pathways, mainly in AI comparisons (Fig.6). Considering the great difference between AD5 (non-replicative) and axenic replicative conditions (AD7, AL7), we expected a large number of enriched pathways in these comparisons. However, only AD5_Vs_AL7 comparison resulted in the identification of enriched pathways (Fig. 6). These were related to oxidative phosphorylation, valine, leucine and isoleucine degradation, glyoxylate and dicarboxylate metabolism, propanoate metabolism, t-RNA biosynthesis, ribosome, protein export, vesicular transport, protein processing in endoplasmic reticulum and endocytosis. Overall, the terms

related with energy metabolism, amino acid degradation, endocytosis, and vesicular transport were up-regulated in AD5 condition (positive stat). Conversely, pathways related to protein synthesis, and protein processing were down-regulated in AD5 condition (negative stat). Therefore, the amastigotes in non-replicative condition (AD5) probably have an intense energy metabolism, in the expense of protein biosynthesis and degradation of amino acids. Further, the replicative amastigotes (AL7) have a more active protein synthesis machinery, empowering cell division. Other study from our group revealed that the oxidative phosphorylation, propanoate metabolism, and ribosome pathways were regulated throughout *T. cruzi* epimastigote cell cycle (Santos Júnior et al., 2021). In that study, we demonstrated that the oxidative phosphorylation is up-regulated in early stages of the cell cycle (G1 phase), then it goes through a decrease in abundance during S phase and increases abundance again during mitosis. Conversely, the ribosome and propanoate metabolism start increasing abundances from G1 phase to S phase, then decrease during mitosis. Moreover, it is interesting to notice the pathways of protein processing in endoplasmic reticulum and protein export enhanced in replicative conditions, suggesting possible role in protein secretion during amastigote replication.

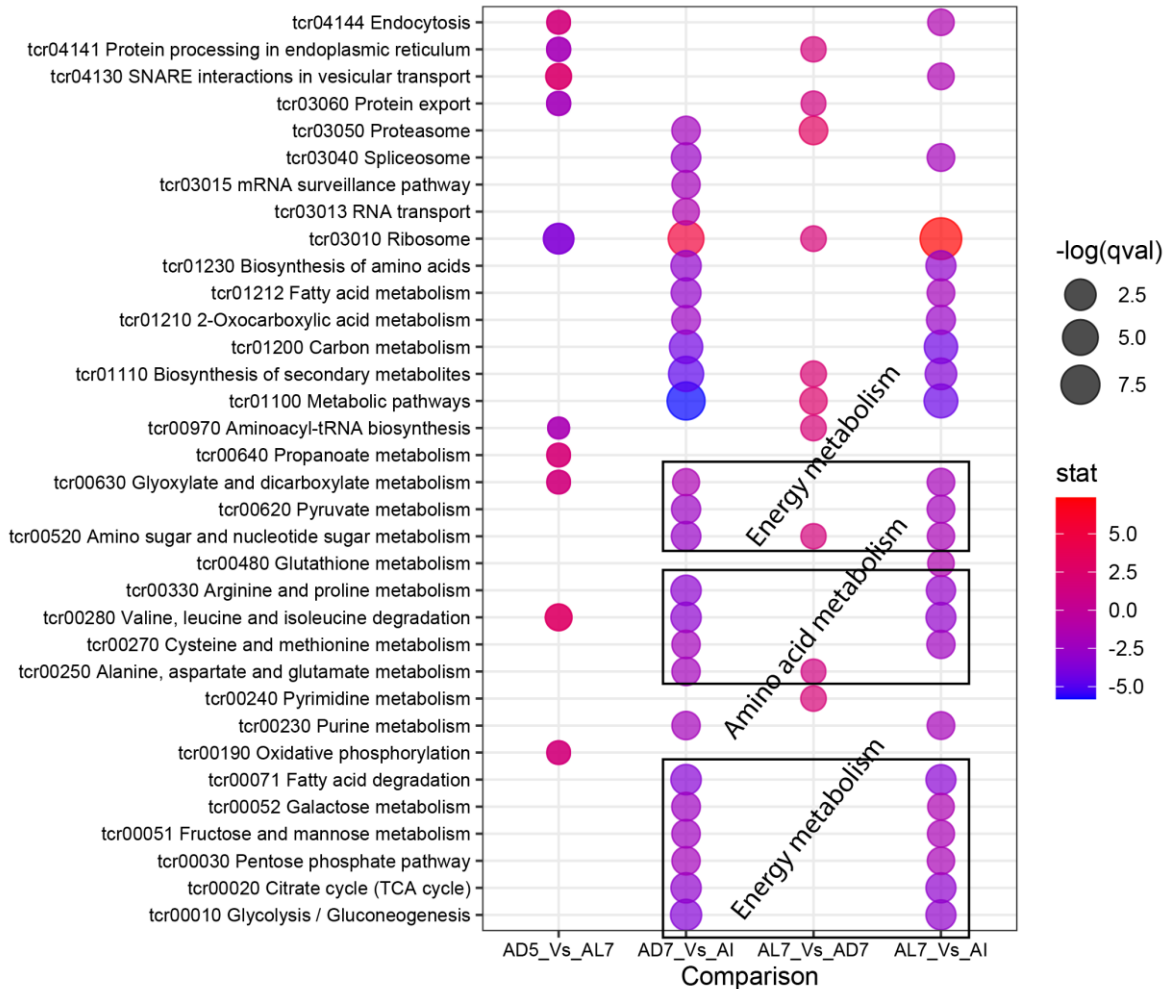


Figure 6. Bubble plot of gage pathway enrichment analysis. All quantified proteins and its abundance were submitted to enrichment analysis using gage R package after conversion of uniprot Accession code to GeneID. The bubble plot was created using the ggplot2 R package with stats and q-value of enriched pathways (q-value < 0.1), and the stat is value representative of the pathway disturb between conditions (a positive value represents that the pathway is up-regulated in the reference condition, that in this case are the first named in the comparison, and a negative value that it is down-regulated in the reference condition).

In addition to the unexpected low number of enriched pathways between replicative and non-replicative conditions, the comparison AL7_Vs_AD7, which demonstrated low number of regulated protein groups, presented some enriched pathways. These enriched pathways were related to protein processing in endoplasmic reticulum, protein export, proteasome, ribosome, biosynthesis of secondary metabolites, tRNA synthesis, amino sugar and nucleotide sugar

metabolism, alanine, aspartate and glutamate metabolism, and pyrimidine metabolism (Fig. 6).

Interestingly, the AI comparison with axenic replicative conditions demonstrated the largest number of enriched pathways (Fig. 6). These pathways were mainly related to energy metabolism, amino acid metabolism, biosynthesis of secondary metabolites, proteasome and ribosome. Only ribosome pathway was up-regulated in axenic replicative conditions (AD7 and AL7), while all others are up-regulated in intracellular amastigotes. Therefore, we hypothesize that some biological processes of *T. cruzi* may be very different in intracellular amastigotes, when compared to axenic replicative amastigotes. Other study of intracellular amastigotes has pointed out that the cell population of axenic amastigotes are more heterogeneous than amastigotes harvested from inside Vero cells. This heterogeneity was suggested to be caused by the presence of cells at different stages of amastigogenesis in axenic culture (Batista et al., 2015). A recent proteomic analysis of *L. mexicana* amastigote forms obtained from different methods (mouse lesion derived, grown in axenic culture, and intracellular amastigotes from macrophage infection) revealed proteins related to virulence factors, glycolysis, lipid metabolism, oxidative phosphorylation, oxidative stress, vesicular transport, transmembrane transport, and amino acid metabolism with significant abundance differences between these amastigote cells (Pacakova et al., 2022).

Clustering analysis of protein groups abundances revealed five clusters with different profiles (Fig. 7). The cluster with more protein groups (cluster 1) suggested the presence of several protein groups with greater abundance in the non-replicative condition (AD5) in comparison to replicative conditions. On the other hand, a cluster that suggest large number of protein groups with greater abundance in the replicative conditions was also formed (cluster 2). However, a modest decrease in the profile could be detected in the AD5 condition. Interestingly, the cluster 3 presented protein groups with greater abundance in the AI conditions, potentially revealing proteins important to intracellular amastigotes replication (Fig. 7). In contrast, cluster 5 demonstrates pattern of lower abundance in the intracellular condition (AI) in

comparison to axenic conditions (AD5, AD7 and AL7). In addition, the cluster 4 showed a pattern of protein groups with greater abundance in axenic replicative conditions (AL7 and AD7). Overrepresentation analysis of proteins belonging to each cluster gave a glimpse in some processes that could be relevant to amastigote replication, intracellular amastigote differences from axenic culture growth amastigotes. Several terms related to structural components or movement were overrepresented in cluster 1 proteins (supplementary figure S2), suggesting that these processes may be relevant to amastigote replication. Another interpretation to this result is that the amastigotes in AD5 conditions are resulting from amastigogenesis and some flagellum protein may remain in these cells. However, this does not exclude the possibility that these proteins may have to be excluded before or during amastigote replication. Moreover, the cluster 3 proteins were overrepresented to terms related to translation or ribosomes (supplementary figure S2), suggesting that the translation machinery may be more active in intracellular amastigotes.

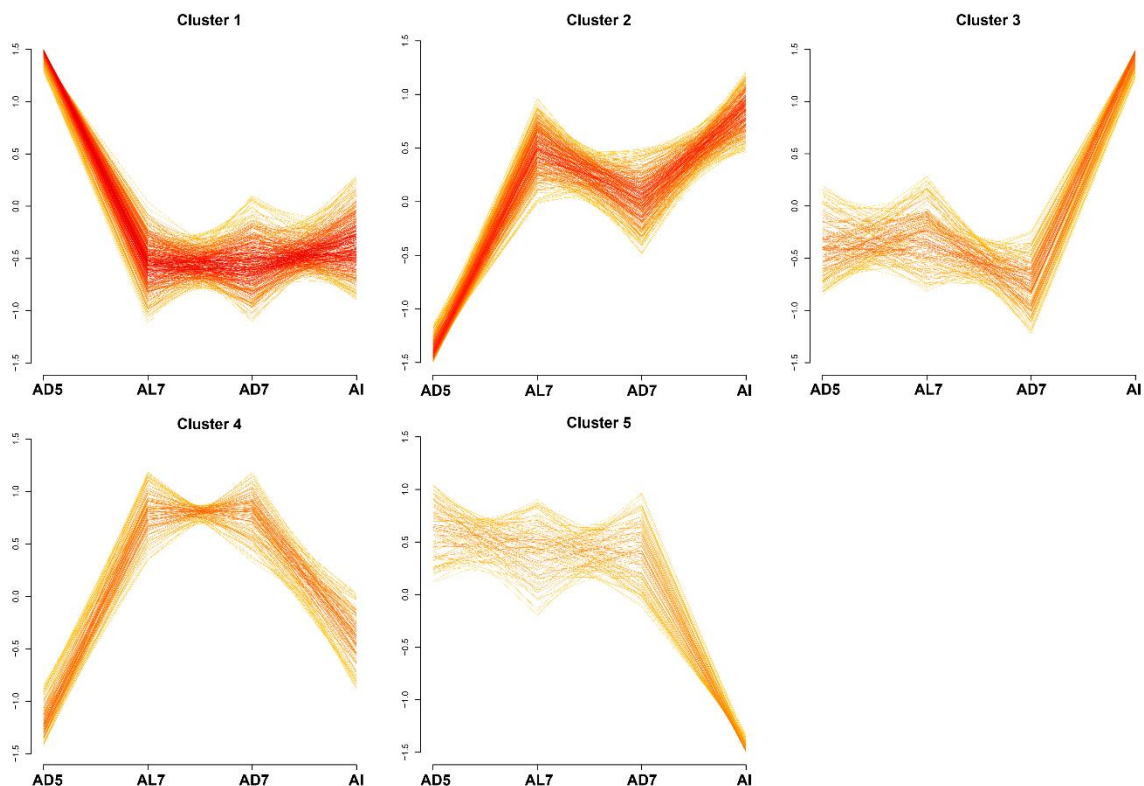
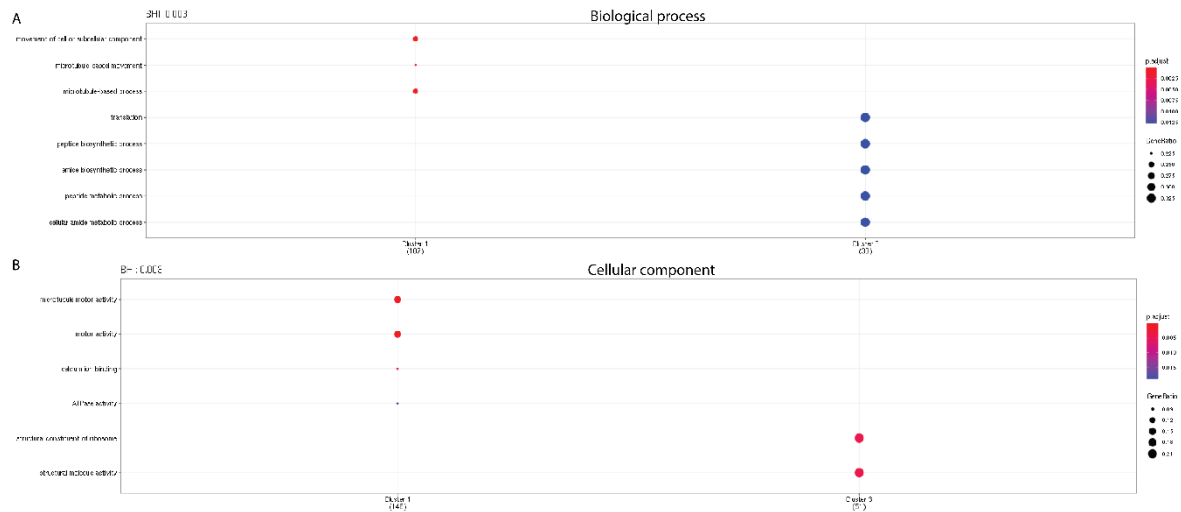


Figure 7. Global variance sensitive clustering profile. Clustering was performed with the abundances of protein groups without missing values. The optimum number of clusters were set as 5, as defined by the method of minimum centroid distance.



Supplementary figure S 2. Overrepresentation analysis of clustered proteins analyzed by VSclust using cellular component and biological process.

Trans-sialidases in amastigote replication

Among the proteins with great FC between AD5 and at least one replicative condition, there were 22 putative trans-sialidases and seven putative paraflagellar rod protein. Interestingly, 12 putative trans-sialidases presented greater abundance in AD5 condition, while 10 were more abundant in replicative conditions. Moreover, beyond those with great FC ($\log_2FC > 2$ or $\log_2FC < -2$), eleven other putative trans-sialidases were regulated in replicative conditions when compared to AD5 (supplementary table S5). Trans-sialidases of *T. cruzi* are a diverse group of proteins derived from a multi-gene family, that can be classified in eight different families (Freitas et al., 2011). These proteins are usually present in the surface of *T. cruzi* cell. This parasite is not capable of producing sialic acid and uses the trans-sialidases to transfer sialic-acid conjugates from the host cell to the protozoan membrane mucins (Herrerros-Cabello et al., 2020).

We have used the tritrypDB (Aslett et al., 2010) to classify the putative trans-sialidases into groups, the Group II had the greatest number of regulated proteins (Fig. 8 A). Beyond Group II, there were identified trans-sialidases from Group I, Group III, Group V, Group VI, Group VIII, and three of them was classified as fragments. Moreover, five trans-sialidases could not be classified (Fig. 8). At first glance, the abundance pattern of the regulated trans-sialidases is divergent with no trend. However, observing only the regulated proteins that clustered together by VSclust analysis, a trend seems to be revealed, mainly in group II trans-sialidases (Fig. 8 B). This tendency indicates an enhanced abundance of group II trans-sialidases in axenic replicative conditions (AD7 and AL7), and a decrease in AD5 and AI. Moreover, another trend seems to be related to a few proteins of group II, group I and group V trans-sialidases, with greater abundance in non-replicative condition (Fig. 8 B). TcS group II proteins are described to be involved in host cell attachment and invasion processes (Freitas et al., 2011), TcS group I have the enzymatic activity of trans-sialidases, and group TcS group V is related to antigenic variation (Herrerros-Cabello et al., 2020).

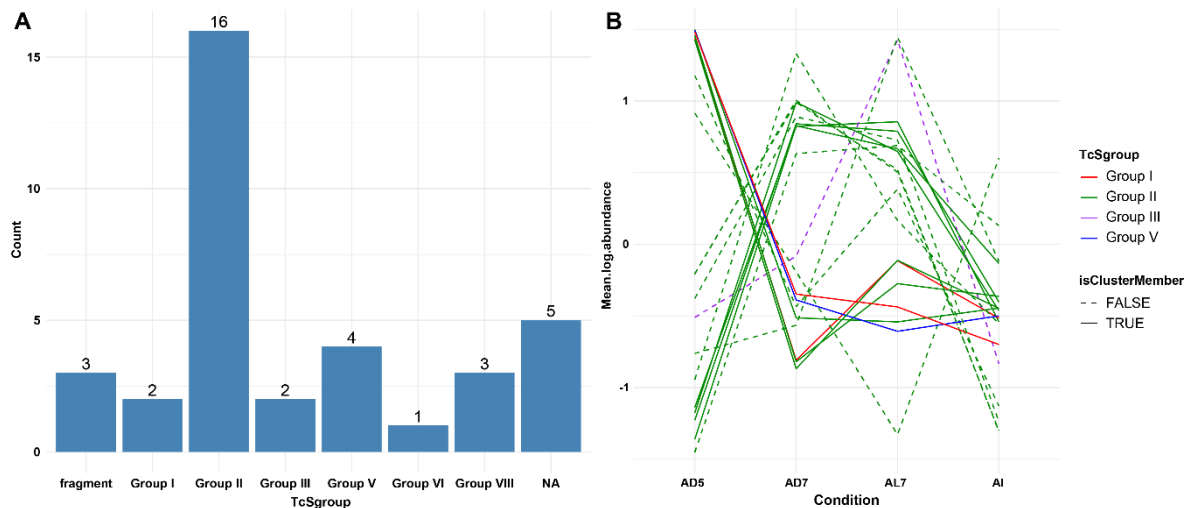


Figure 8. Number and classification in subgroups of regulated trans-sialidases. A. Count of regulated proteins identified in each TcSgroup. B. Abundance profile of regulated trans-sialidases according to TcSgroup and cluster member.

A recent study suggests that active trans-sialidases (sialidases that present enzymatic activity of sialidases/neuramidase, therefore members of TcS group I) are essential for the differentiation of amastigotes to trypomastigotes, and host cell egress. However it was suggested that these proteins have no effect in the host cell infection or replication (Burle-Caldas et al., 2022). On the other hand, several studies have related the function of trans-sialidases in the host cell infection (Herrerros-Cabello et al., 2020). Considering the suggested role of TcS II group trans-sialidases in the host cell infection, it was expected that these proteins would not be very abundant in replicative forms of amastigote, since its replication occurs after the infection of host cell by trypomastigotes. Moreover, if it should appear regulated, one would expect it to be up-regulated in AD5 condition, as residual proteins from trypomastigotes after amastigogenesis. However, TcS II group presented at least four proteins with enhanced abundance in axenic replicative conditions (Fig. 8B), suggesting that a subgroup of TcS II proteins have a role in amastigote replication. Besides this, TcS II group also presented at least three regulated proteins with increased abundance in non-replicative condition AD5, in comparison to replicative conditions. This group followed a similar pattern to group I and group V TcS, suggesting that their abundances are reduced during *T. cruzi* replication, therefore, they may be more important during other cell processes, such as the suggested cell host infection.

Leishmanolysin-like peptidases and kDNA associated proteins may be involved in amastigote replication

Two Leishmanolysin-like peptidase presented high FC with greater abundance in axenic replicative conditions. Leishmanolysin peptidases are also known as major surface proteases (MSP) or GP63 proteases. These proteins represents 1 % of the total protein in promastigotes of *L. major* (Yao, 2010). Moreover, they are implicated in several processes of these parasites, for example: degradation of cytosolic proteins, and survival of intracellular amastigotes. Homologues of leishmanolysin peptidases were found in all trypanosomatids,

including *T. cruzi*. However, the function in different groups are suggested to be different of those performed in *Leishmania* spp. (Yao, 2010). In *T. brucei*, two MSP promote antigenic variation through the cleavage of variant surface glycoproteins, and have been explored as possible drug targets using molecular docking approach (Marufu & Coetzer, 2022). In *T. cruzi*, the leishmanolysin-like peptidases, commonly referred as GP63 proteases, were classified in four clades, based on the sequence analysis of GP63 proteases of trypanosomatids. These clades were named TcGP63-a, TcGP63-b, TcGP63-c, and TcGP63-d (Ma et al., 2011). The Leishmanolysin-like peptidase regulated in axenic replicative conditions (Q4DLF6) was classified as TcGP63-b in this study (called by Trycr293). TcGP63-b peptidases are thought to be extensively expressed in *T. cruzi*, and was implicated by a former study in parasite infection (Cuevas, Cazzulo, & Sánchez, 2003; Ma et al., 2011). On the other hand, a former study suggested the localization of GP63 isoform(s), identified by anti-TcGP63, near the kinetoplast DNA and flagella pocket of metacyclic trypomastigotes. Moreover, this study also described that in epimastigote lifeforms, proteins identified by the anti-TcGP63 was present predominantly in the cell body and flagellum of the parasite (Kulkarni, Olson, Engman, & McGwire, 2009).

Despite the functions in cell host interaction described for *Leishmania* spp., *T. brucei*, and even of the *T. cruzi* GP63 proteases, the two Leishmanolysin-like peptidases regulated in this study may also be involved replication of amastigote cells of *T. cruzi*, considering that they presented higher abundance only in axenic replicative conditions (Q4DLF6, Log₂FC AD5 Vs AD7 = -3.55; Q4DVI5, Log₂FC AD5 Vs AD7 = -3.81).

Despite the function of cell replication, extracellular amastigotes have the ability to infect host cells (Ferri & Edreira, 2021). The presence of regulated transglucosylases and GP63 with greater abundance in replicative conditions may also be a consequence to arrangement of cell surface proteins that is essential to cell-host interactions. Moreover, the main difference between the conditions is the pH of the medium (pH 5 in non-replicative condition Vs pH 7 in replicative conditions), the pH that extracellular amastigotes should be exposed is also a pH 7. On the other hand,

extracellular amastigotes infection of host cells is suggested to be promoted by δ -Amastin (Ferri & Edreira, 2021). However, the two proteins identified as putative amastins in this study were not regulated (supplementary table S3). This observation suggests that the differences between conditions of this study did not promote a phenotype leading to a cell with more cell-host interactions, but this hypothesis should not be discarded.

Two putative kinetoplast DNA-associated protein also presented great FC with greater abundance in axenic replicative conditions (Q4CY62, FC = -3.61; Q4CY61, FC = -2.99). Kinetoplast DNA (kDNA) is a concatenated network of minicircles and maxicircles located in kinetoplast, only present in trypanosomes. This network is greatly condensed with the aid of associated proteins (C. Xu & Ray, 1993) and its fission is coupled to cell abscission (Zhou, Hu, & Li, 2014). The uniqueness characteristic of kDNA and its proteins makes them attractive targets for drug development. Both regulated kDNA associated proteins possess the High mobility group (HMG)-box domain according to InterPro annotation in Uniprot. HMG-box domain containing proteins are known to be involved in the compaction, protection, regulation, and segregation of mitochondrial DNA (Vozáriková et al., 2020). Recently, a *T. cruzi* HMGB protein (TcHMGB, high mobility group B) was demonstrated to bind to DNA and impair intracellular amastigote and epimastigote replication (Tavernelli et al., 2019). Moreover, it was demonstrated in *T. brucei* that a HMG-box containing protein, tbKAP-6 (kinetoplast-associated protein 6), is essential in kDNA replication and maintenance (Vozáriková et al., 2020; Wang, Pappas-Brown, Englund, & Jensen, 2014). On the other hand, in *T. cruzi* the kinetoplast-associated protein 3 (tcKAP 3) was suggested to be involved in the kDNA replication, however its gene knockout had no effect in proliferation, differentiation or infectivity of *T. cruzi* (F. S. P. de Souza et al., 2010). It was suggested that others kinetoplast-associated proteins may be compensating its function. According to tritryp, one of the proteins with greater abundance in axenic replicative conditions is the TcKAP4 protein (Q4CY61). This protein was identified in a chromatin enriched fraction previously, suggested to be caused by mitochondrial DNA extraction during chromatin enrichment (Leandro de Jesus et al., 2017). In

another study, this same protein was identified during metacyclogenesis of *T. cruzi*, quantified in all conditions (de Godoy et al., 2012), suggesting the expression of this protein in metacyclic trypomastigote and epimastigote forms.

Proteins with unknown functions

There were 449 uncharacterized proteins among regulated protein groups (approx. 41%). All proteins predicted in the *T. cruzi* (CL Brener) reference proteome presented only 1795 uncharacterized proteins (approx. 9 %). The sequences of regulated uncharacterized proteins were submitted to blastp against the swissprot database, resulting in the identification of similarity (e-value < 0.01) of 131 protein groups submitted to the search (supplementary table S6). From those, several protein groups up-regulated in AD5 condition were related to cilium GO terms. Moreover, four protein groups with greater abundance in AD5 were related to cell division GO terms and had similarity with Leucine-rich repeat-containing protein. Interestingly, two of the similar proteins up-regulated in the AD5 condition were CARD-15-like protein, related to intracellular signal transduction. On the other hand, protein groups up-regulated in replicative conditions presented similarity with two plectin proteins, involved in actin cytoskeleton reorganization. Other two protein groups up-regulated in replicative conditions had similarity to tubulin polymerization-protein family member 2 and Microtubule-actin cross-linking factor 1. In addition, the uncharacterized protein that presented the greatest FC between AD5 and all replicative conditions ($-5.5 < \log_2FC < -3.8$) presented similarity to a protein related to cilium assembly and centriolar satellite.

5.1.5 Concluding remarks

In this study, we presented the isolation of *T. cruzi* intracellular amastigotes inside HeLa cells through a simple procedure using DEAE-cellulose columns and the assessment of these isolation through different methods, supporting the efficient enrichment of amastigotes with less than 10 % of human proteins contamination.

Moreover, the quantitative proteomic analysis of amastigotes obtained through different methods, in replicative and non-replicative conditions, revealed differences in processes energy and amino acid metabolism between intracellular and axenic amastigotes.

The proteomic comparisons of non-replicative and replicative amastigotes highlighted several proteins with possible relevance in the replication *T. cruzi*, such as some members of the trans-sialidases family from groups I, II, III, and V. Further, this analysis also suggested the role of Leishmanolysin-like peptidases and kDNA associated proteins that should be thoroughly investigated and validated through complementary techniques.

5.1.6 References

- Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., ... Wang, H. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Research*, 38(suppl_1), D457–D462. <https://doi.org/10.1093/nar/gkp851>
- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., ... Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- Batista, C. M., Kessler, R. L., Eger, I., & Soares, M. J. (2015). Trypanosoma cruzi Intracellular Amastigotes Isolated by Nitrogen Decompression Are Capable of Endocytosis and Cargo Storage in Reserosomes. *PLOS ONE*, 10(6), e0130165. <https://doi.org/10.1371/journal.pone.0130165>
- Burle-Caldas, G. de A., dos Santos, N. S. A., de Castro, J. T., Mugge, F. L. B., Grazielle-Silva, V., Oliveira, A. E. R., ... Teixeira, S. M. R. (2022). Disruption of Active Trans-Sialidase Genes Impairs Egress from Mammalian Host Cells and Generates Highly Attenuated Trypanosoma cruzi Parasites. *MBio*, 13(1). <https://doi.org/10.1128/MBIO.03478-21>

- CAMARGO, E. P. (1964). GROWTH AND DIFFERENTIATION IN TRYPANOSOMA CRUZI. I. ORIGIN OF. *Revista Do Instituto de Medicina Tropical de São Paulo*, 12, 93–100. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14177814>
- Carvalho, T. U., & Souza, W. (1983). Separation of amastigotes and trypomastigotes of *Trypanosoma cruzi* from cultured cells. *Zeitschrift Für Parasitenkunde Parasitology Researchr Parasitenkunde Parasitology Research*, 69(5), 571–575. <https://doi.org/10.1007/BF00926668>
- Clayton, C. (2019). Regulation of gene expression in trypanosomatids: Living with polycistronic transcription. *Open Biology*, 9(6). <https://doi.org/10.1098/rsob.190072>
- Cuevas, I. C., Cazzulo, J. J., & Sánchez, D. O. (2003). gp63 Homologues in *Trypanosoma cruzi*: Surface Antigens with Metalloprotease Activity and a Possible Role in Host Cell Infection. *Infection and Immunity*, 71(10), 5739–5749. <https://doi.org/10.1128/IAI.71.10.5739-5749.2003>
- de Godoy, L. M. F., Marchini, F. K., Pavoni, D. P., Rampazzo, R. de C. P., Probst, C. M., Goldenberg, S., & Krieger, M. A. (2012). Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis. *PROTEOMICS*, 12(17), 2694–2703. <https://doi.org/10.1002/pmic.201200078>
- Ferri, G., & Edreira, M. M. (2021). All Roads Lead to Cytosol: *Trypanosoma cruzi* Multi-Strategic Approach to Invasion. *Frontiers in Cellular and Infection Microbiology*, 11(March), 1–17. <https://doi.org/10.3389/fcimb.2021.634793>
- Freitas, L. M., dos Santos, S. L., Rodrigues-Luiz, G. F., Mendes, T. A. O., Rodrigues, T. S., Gazzinelli, R. T., ... Bartholomeu, D. C. (2011). Genomic Analyses, Gene Expression and Antigenic Profile of the Trans-Sialidase Superfamily of *Trypanosoma cruzi* Reveal an Undetected Level of Complexity. *PLoS ONE*, 6(10), e25914. <https://doi.org/10.1371/journal.pone.0025914>
- Galili, T., O’Callaghan, A., Sidi, J., & Sievert, C. (2018). Heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, 34(9), 1600–1602. <https://doi.org/10.1093/bioinformatics/btx657>

- Gamarro, F., Osuna, A., Castanys, S., Pérez-López, M. I., & Ruiz-Pérez, L. M. (1985). Isolation and purification of amastigotes of *Trypanosoma cruzi* from cultured Vero cells. *Zeitschrift Für Parasitenkunde Parasitology Research*, 71(1), 15–17. <https://doi.org/10.1007/BF00932914>
- Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36(8), 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>
- Herreros-Cabello, A., Callejas-Hernández, F., Gironès, N., & Fresno, M. (2020). *Trypanosoma cruzi* genome: Organization, multi-gene families, transcription, and biological implications. *Genes*, 11(10), 1–26. <https://doi.org/10.3390/genes11101196>
- Kulkarni, M. M., Olson, C. L., Engman, D. M., & McGwire, B. S. (2009). *Trypanosoma cruzi* GP63 Proteins Undergo Stage-Specific Differential Posttranslational Modification and Are Important for Host Cell Infection. *Infection and Immunity*, 77(5), 2193–2200. <https://doi.org/10.1128/IAI.01542-08>
- Leandro de Jesus, T. C., Calderano, S. G., Vitorino, F. N. de L., Llanos, R. P., Lopes, M. D. C., de Araújo, C. B., ... Chagas da Cunha, J. P. (2017). Quantitative Proteomic Analysis of Replicative and Nonreplicative Forms Reveals Important Insights into Chromatin Biology of *Trypanosoma cruzi*. *Molecular & Cellular Proteomics*, 16(1), 23–38. <https://doi.org/10.1074/mcp.M116.061200>
- Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., & Woolf, P. J. (2009). GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*, 10, 1–17. <https://doi.org/10.1186/1471-2105-10-161>
- Ma, L., Chen, K., Meng, Q., Liu, Q., Tang, P., Hu, S., & Yu, J. (2011). An evolutionary analysis of trypanosomatid GP63 proteases. *Parasitology Research*, 109(4), 1075–1084. <https://doi.org/10.1007/s00436-011-2348-x>
- Marques, A., Marques, A., Nakayasu, E., & Almeida, I. (2011). Purification of extracellular and intracellular amastigotes of *Trypanosoma cruzi* from mammalian host-infected cells. *Protocol Exchange*, 1–7.

<https://doi.org/10.1038/protex.2011.265>

- Marufu, L., & Coetzer, T. H. T. (2022). Homology modelling of *Trypanosoma brucei* major surface proteases and molecular docking of variant surface glycoproteins and inhibitor ligands for drug design. *Journal of Molecular Graphics and Modelling*, 111(August 2021), 108104. <https://doi.org/10.1016/j.jmgm.2021.108104>
- Millikin, R. J., Solntsev, S. K., Shortreed, M. R., & Smith, L. M. (2018). Ultrafast Peptide Label-Free Quantification with FlashLFQ. *Journal of Proteome Research*, 17(1), 386–391. <https://doi.org/10.1021/acs.jproteome.7b00608>
- Moretti, N. S., Mortara, R. A., & Schenkman, S. (2020). *Trypanosoma cruzi*. *Trends in Parasitology*, 36(4), 404–405. <https://doi.org/10.1016/j.pt.2019.10.002>
- Navarro, M. C., De Lima, A. R., Askue, J., & Contreras, V. T. (2003). Morphological comparison of axenic amastigogenesis of trypomastigotes and metacyclic forms of *Trypanosoma cruzi*. *Memórias Do Instituto Oswaldo Cruz*, 98(1), 83–91. <https://doi.org/10.1590/S0074-02762003000100012>
- Pacakova, L., Harant, K., Volf, P., & Lestinova, T. (2022). Three types of *Leishmania mexicana* amastigotes: Proteome comparison by quantitative proteomic analysis. *Frontiers in Cellular and Infection Microbiology*, 12(November), 1–18. <https://doi.org/10.3389/fcimb.2022.1022448>
- Pérez-Molina, J. A., & Molina, I. (2018). Chagas disease. *The Lancet*, 391(10115), 82–94. [https://doi.org/10.1016/S0140-6736\(17\)31612-4](https://doi.org/10.1016/S0140-6736(17)31612-4)
- Queiroz, R. M. L., Charneau, S., Mandacaru, S. C., Schwämmle, V., Lima, B. D., Roepstorff, P., & Ricart, C. A. O. (2014). Quantitative proteomic and phosphoproteomic analysis of *Trypanosoma cruzi* amastigogenesis. *Molecular and Cellular Proteomics*, 13(12), 3457–3472. <https://doi.org/10.1074/mcp.M114.040329>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Retrieved from <https://www.r-project.org/>.

- RStudio Team. (2021). *RStudio: integrated development for R*. Bonton, MA.
- Santos Júnior, A. de C. M. dos, Melo, R. M., Ferreira, B. V. G., Pontes, A. H., Lima, C. M. R. de, Fontes, W., ... Ricart, C. A. O. (2021). Quantitative proteomics and phosphoproteomics of *Trypanosoma cruzi* epimastigote cell cycle. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1869(5), 140619. <https://doi.org/10.1016/j.bbapap.2021.140619>
- Schwämmle, V., & Jensen, O. N. (2018). VSClust: Feature-based variance-sensitive clustering of omics data. *Bioinformatics*, 34(17), 2965–2972. <https://doi.org/10.1093/bioinformatics/bty224>
- Solntsev, S. K., Shortreed, M. R., Frey, B. L., & Smith, L. M. (2018). Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *Journal of Proteome Research*, 17(5), 1844–1851. <https://doi.org/10.1021/acs.jproteome.7b00873>
- Souza, F. S. P. de, Rampazzo, R. de C. P., Manhaes, L., Soares, M. J., Cavalcanti, D. P., Krieger, M. A., ... Fragoso, S. P. (2010). Knockout of the gene encoding the kinetoplast-associated protein 3 (KAP3) in *Trypanosoma cruzi*: Effect on kinetoplast organization, cell proliferation and differentiation. *Molecular and Biochemical Parasitology*, 172(2), 90–98. <https://doi.org/10.1016/j.molbiopara.2010.03.014>
- Souza, W. (2005). Basic Cell Biology of *Trypanosoma cruzi*. *Current Pharmaceutical Design*, 8(4), 269–285. <https://doi.org/10.2174/1381612023396276>
- Stratton, K. G., Webb-Robertson, B. J. M., McCue, L. A., Stanfill, B., Claborne, D., Godinez, I., ... Bramer, L. M. (2019). PmartR: Quality Control and Statistics for Mass Spectrometry-Based Biological Data. *Journal of Proteome Research*, 18(3), 1418–1425. <https://doi.org/10.1021/acs.jproteome.8b00760>
- Tavernelli, L. E., Motta, M. C. M., Gonçalves, C. S., da Silva, M. S., Elias, M. C., Alonso, V. L., ... Cribb, P. (2019). Overexpression of *Trypanosoma cruzi* High Mobility Group B protein (TcHMGB) alters the nuclear structure, impairs cytokinesis and reduces the parasite infectivity. *Scientific Reports*, 9(1), 192.

<https://doi.org/10.1038/s41598-018-36718-0>

- Tomlinson, S., Vandekerckhove, F., Frevert, U., & Nussenzweig, V. (1995). The induction of *Trypanosoma cruzi* trypomastigote to amastigote transformation by low pH. *Parasitology*, *110*(5), 547–554. <https://doi.org/10.1017/S0031182000065264>
- Vozáriková, V., Kunová, N., Bauer, J. A., Frankovský, J., Kotrasová, V., Procházková, K., ... Tomáška, L. (2020). Mitochondrial HMG-Box Containing Proteins: From Biochemical Properties to the Roles in Human Diseases. *Biomolecules*, *10*(8), 1193. <https://doi.org/10.3390/biom10081193>
- Wang, J., Pappas-Brown, V., Englund, P. T., & Jensen, R. E. (2014). TbKAP6, a Mitochondrial HMG Box-Containing Protein in *Trypanosoma brucei*, Is the First Trypanosomatid Kinetoplast-Associated Protein Essential for Kinetoplast DNA Replication and Maintenance. *Eukaryotic Cell*, *13*(7), 919–932. <https://doi.org/10.1128/EC.00260-13>
- WHO. (2022). Fact sheet: Chagas disease (also known as American trypanosomiasis). *Chagas Disease Fact Sheets*. Retrieved from [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis))
- Xu, C., & Ray, D. S. (1993). Isolation of proteins associated with kinetoplast DNA networks in vivo. *Proceedings of the National Academy of Sciences*, *90*(5), 1786–1789. <https://doi.org/10.1073/pnas.90.5.1786>
- Yao, C. (2010). Major Surface Protease of Trypanosomatids: One Size Fits All? *Infection and Immunity*, *78*(1), 22–31. <https://doi.org/10.1128/IAI.00776-09>
- Zhou, Q., Hu, H., & Li, Z. (2014). New Insights into the Molecular Mechanisms of Mitosis and Cytokinesis in Trypanosomes. In *International Review of Cell and Molecular Biology* (1st ed., Vol. 308, pp. 127–166). <https://doi.org/10.1016/B978-0-12-800097-7.00004-X>
- Zuma, A. A., dos Santos Barrias, E., & de Souza, W. (2020). Basic Biology of

Trypanosoma cruzi. *Current Pharmaceutical Design*, 27(14), 1671–1732.
<https://doi.org/10.2174/1381612826999201203213527>

5.2 Integração de proteômica *bottom-up* e *top-down* para caracterização de proteoformas da forma de vida epimastigota de *Trypanosoma cruzi*

5.2.1 Objetivos

A análise proteômica da forma de vida epimastigota de *T. cruzi* integrando abordagens *bottom-up* e *top-down* foi realizada para identificar proteoformas desse protozoário em nível de análise intacta dessas por espectrometria de massas e complementar essas informações com as modificações pós-traducionais (PTMs) identificadas por abordagem *bottom-up* descritiva. A abordagem *top-down* foi baseada no fracionamento de proteínas e a *bottom-up* utilizou estratégia multi-enzimática acoplada a fracionamento para identificação de PTMs utilizando o método *enhanced Global Post Translational Modification-Discovery* (G-PTM-D)(Miller et al., 2019; Solntsev et al., 2018) (figura 38).

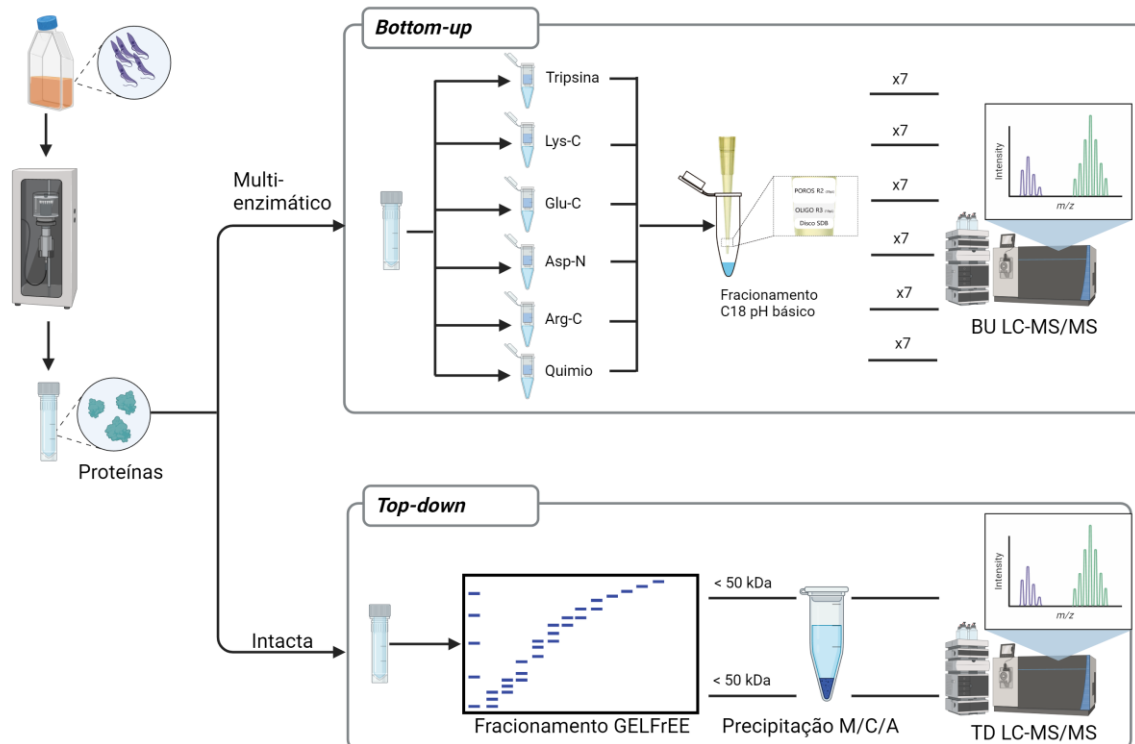


Figura 38. Fluxograma para preparo de amostra da forma de vida epimastigota de *T. cruzi* para análises proteômicas *bottom-up* e *top-down*.

5.2.2 Métodos

Células e extração de proteínas

Aproximadamente 8×10^8 células epimastigotas de *T. cruzi* (CL Brener) foram rompidas usando tampão de lise SDS (100 mM Tris-HCl, 50 mM DTT, 4% SDS, pH = 7,6), suplementado com inibidor de fosfatase (1x PhosStop, Roche) e inibidor de protease (1x Ultra Complete EDTA-free, Roche). Após ressuspender as células, elas foram incubadas a 95°C e 500 rpm por 10 min, seguido de ultrassom em sonda com 3 ciclos de 10 s a 40% e 10 s desligado. A seguir, foram centrifugadas a 16.000 g por 15 min (22 °C) e o sobrenadante foi coletado e armazenado a -80 °C para análises proteômicas. A quantidade de proteínas total do lisado de epimastigotas foi estimada usando uma solução diluída 200 vezes por método de fluorescência, utilizando o kit Qubit™ (Invitrogen) de acordo com as instruções do fabricante.

Preparo de amostra e análise *top-down* e *bottom-up* por LC-MS/MS

Cerca de 1 mg de proteínas foi reduzido e alquilado. Em seguida, essas proteínas foram submetidas a abordagens proteômica *top-down* ou *bottom-up* multi-enzimática. Ambas as abordagens foram realizadas da mesma forma como descrito no capítulo III (Metodologia Comum). Brevemente, cerca de 400 µg de proteínas foram submetidas a fracionamento por GELFrEE (Tran & Doucette, 2009), gerando 17 frações. O perfil de massa das frações foi determinado por SDS-PAGE 12% e as frações com massa menor que 50 kDa foram submetidas a precipitação metanol/clorofórmio/água (3:1:2) (Wessel & Flügge, 1984). As proteínas resultantes da precipitação foram solubilizadas em solução de injeção e direcionadas para a análise proteômica *top-down* por LC-MS/MS. Para a abordagem proteômica *bottom-up*, cerca de 100 µg de proteínas foram submetidas a digestão por seis enzimas separadamente (tripsina, quimiotripsina, Lys-C, Glu-C, Arg-C, Asp-N). Então, os peptídeos resultantes de cada digestão foram dessalinizados em *stagetip* e fracionados por meio de fase reversa com pH básico, utilizando microcolunas com disco de SDB, e resinas OLIGO R3 e POROS R2. Essa abordagem gerou 7 frações para cada digestão, resultando em 42 frações. Após o fracionamento, os peptídeos resultantes foram secados, solubilizados, e submetidos a LC-MS/MS.

Análise de dados *bottom-up*, *top-down* e integração das abordagens

A análise de dados, assim como o preparo de amostras, foi realizada como descrito no capítulo III (Metodologia Comum). Sucintamente, os espectros de peptídeos foram identificadas pelo programa MetaMorpheus, utilizando abordagem de inferência de proteínas a partir de múltiplas digestões (Miller et al., 2019) e identificação de PTMs por meio de estratégia *Global-Post Translational Modification-Discovery* (G-PTM-D) contra o proteoma de referência de *T. cruzi* CL Brener (UP000002296) e banco de contaminantes comuns (Solntsev et al., 2018). Ademais, foram buscadas as modificações “biológicas comuns”, “adutos” e “metais” na estratégia G-PTM-D. Para análise de dados proteômicos *top-down* foram utilizados dois *softwares*, inicialmente foi usado o programa TopPic para a

identificação das proteoformas de maneira tolerante a erro (Kou et al., 2016). Em seguida foi utilizado o programa MetaMorpheus para a identificação das proteoformas de maneira restrita a erros (Yuling Dai et al., 2022). Ambas identificações foram realizadas inicialmente contra o proteoma de referência de *T. cruzi* CL Brener (UP000002296), no entanto, a abordagem pelo MetaMorpheus passou por mais uma rodada de identificação dos espectros contra o proteoma de referência de *T. cruzi* complementado com as modificações identificadas em abordagem *bottom-up* por G-PTM-D.

5.2.3 Resultados

Análise global das proteoformas detectadas

As proteínas obtidas a partir de células de epimastigotas de *T. cruzi* passaram por fracionamento em GELFrEE. Como esperado, esse fracionamento foi eficiente em separar as proteínas em frações com diferentes padrões de massas moleculares. As frações 0, 1, 2, 3, 4, 5, 7, 9, 11, 13, 15 e 20 foram submetidas a LC-MS/MS, considerando a presença de proteínas abaixo de 50 kDa (figura 37).

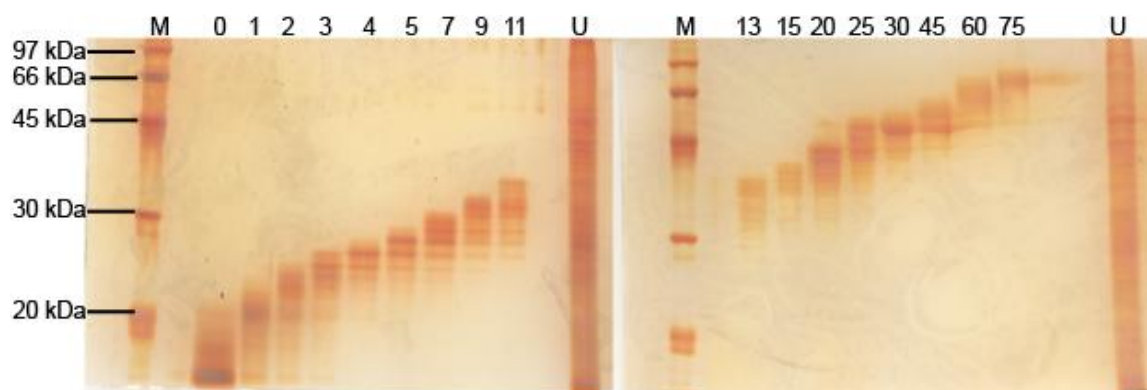


Figura 39. SDS-PAGE das frações de GELFrEE do lisado das células epimastigotas de *T. cruzi*. As frações são representadas por números que indicam o tempo em minutos após a saída do azul de bromofenol durante o fracionamento por GELFrEE.

A análise dessas frações possibilitou a detecção de eventos com predominância de massa molecular de até 30 kDa (figura 40). Foram detectados mais eventos de maior massa nas frações finais (figura 40 B). Isso sugere que proteoformas maiores do que 25 kDa estão muito mais presentes nas frações de maior massa (figura 40 B). Apesar disso, praticamente não foram detectados eventos com massas maiores do que 30 kDa (figura 40 B).

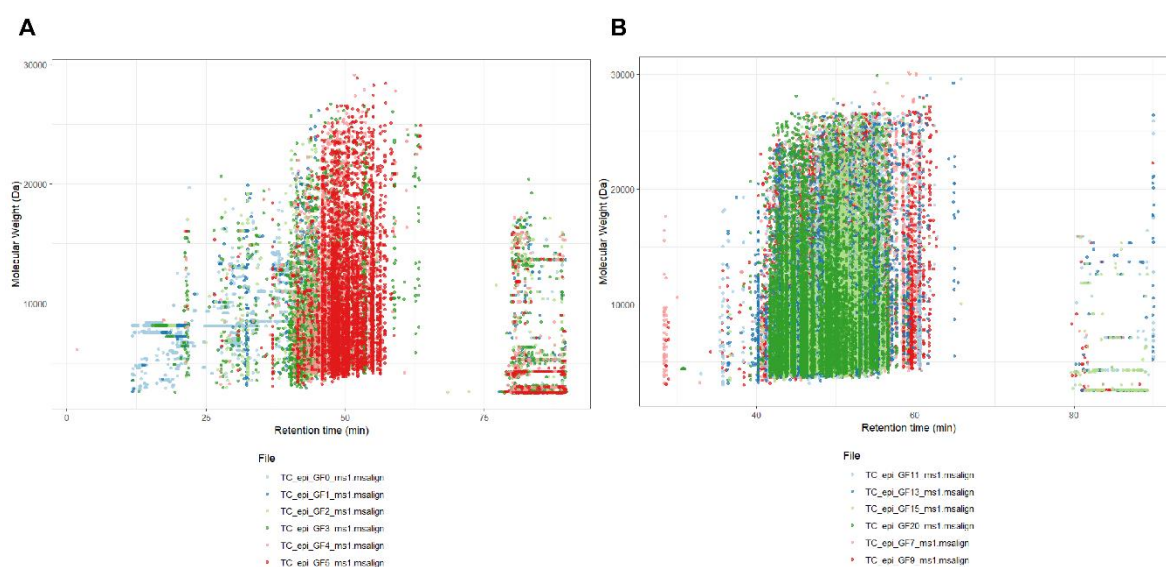


Figura 40. Análise dos eventos detectados por LC-MS e suas massas moleculares de acordo com o tempo de retenção. A) Frações de GELFrEE 0, 1, 2, 3, 4 e 5. B) Frações de GELFrEE 7, 9, 11, 13, 15 e 20. Análise realizada com *threshold* de 20% para intensidade dos eventos detectados usando o programa VisioProt-MS (Locard-Paulet et al., 2019).

Os eventos de MS1 detectados e deconvoluídos por TopFD (Basharat et al., 2023) foram submetidos a análise de diferença de massas (Δm - *mass shift*) por MSTopDiff (Kaulich et al., 2022). Essa análise possibilita a interpretação visual das modificações muito frequentes nos precursores, que podem ser indicativas de PTMs, ou artefatos do processamento da amostra (Kaulich et al., 2022). Foi analisada uma fração com proteoformas de baixa massa (GF0) e uma fração de alta massa (GF11) (figura 41). Na fração de baixa massa foram observados padrões de Δm de 1 Da, 16 Da, 28 Da, 57 Da, entre outros (figura 41 A). A variação de massa correspondente a 1 Da provavelmente está relacionada com a dificuldade em

determinar a massa monoisotópicas de envelopes isotópicos de moléculas grandes (Yunxiang Dai et al., 2017; Gallagher et al., 2020). Por outro lado, a Δm de 16 Da sugere presença de proteoformas oxidadas e não oxidadas da mesma proteína. A presença frequente de proteoformas formiladas (UnimodAC: 122, $\Delta m = 27.994915$) poderia explicar a $\Delta m = 28,04$ Da. A variação de massa de 57 Da foi observada nas frações de baixa massa e de alta massa (figura 41 A, B), essa pode estar relacionada com modificações por carbamidometilação causadas pela alquilação de cisteínas, processo causado pelo preparo de amostras pela incubação das proteínas em iodoacetoamida (IAA). A intensidade do pico desse Δm na fração GF0 sugere que ela ocorra com intensidade semelhante ao seu íon de massa menor (figura 41 A). Além disso, foi possível observar nas frações GF0 e GF11 $\Delta m = 74,02$ Da, e diferenças maiores que sugerem a presença de múltiplas modificações iguais (148,04 Da (2x), 222,06 Da (3x); e 296,08 Da (4x)) (figura 41). Essa variação de massa já foi relacionada a modificação causada por glicerol à resíduos de glutamato e aspartato (Xing et al., 2008). Ademais, esse padrão de modificações também foi observado na análise proteômica *top-down* de *C. glutamicum* apresentado no capítulo II, sessão “Integração de proteômica bottom-up multi-enzimática e top-down para caracterização de proteoformas da bactéria *C. glutamicum*”. A presença dessas diferenças de massas frequentes na análise de *C. glutamicum* e *T. cruzi* sugere que essas modificações podem estar sendo causadas pelo preparo da amostra. Essa possibilidade deve ser melhor investigada para otimizar os métodos de preparo de amostra e/ou identificação das proteoformas.

Por meio da identificação de modificações causadas por artefatos de processamento de amostras e adutos, podem ser realizadas otimizações de protocolos (Kaulich et al., 2022), além de possibilitar a identificação de proteoformas com essas modificações indesejadas (Yunxiang Dai et al., 2019, 2017).

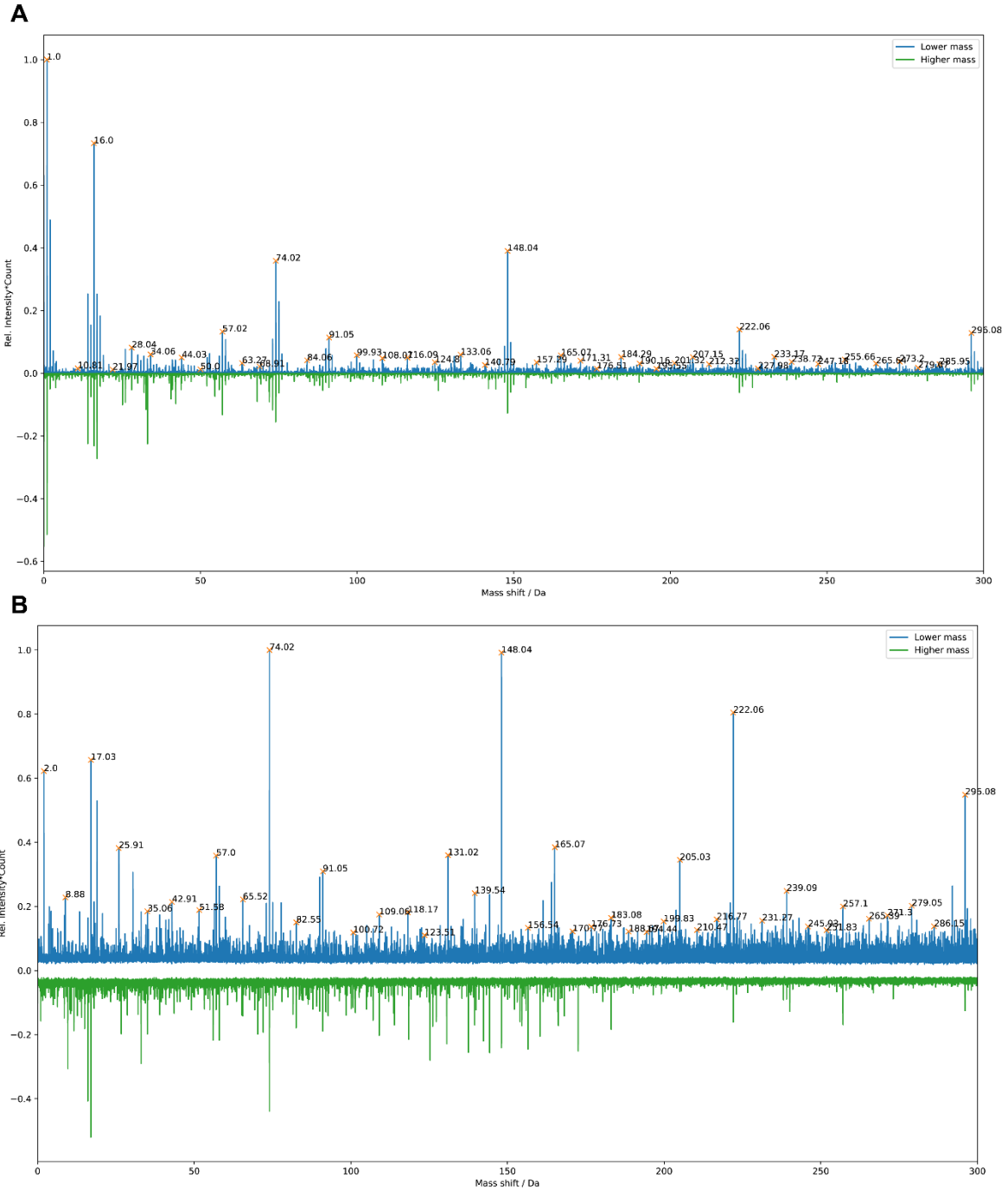


Figura 41. Análise de diferença de massas (Δm) de eventos de MS1 detectados. A) Δm calculadas na fração GF0. B) Δm calculadas na fração GF11. Ambos histogramas foram construídos com $binwidth = 0,01$. Cálculos e construção de histograma realizados pelo programa MStopDiff (Kaulich et al., 2022).

Proteoformas identificadas por TopPic Suite

A análise proteômica *top-down* das frações por TopPic, utilizando estratégia de busca tolerante a erro, identificou 658 PrSMs, 244 proteoformas e 163 proteínas (Tabela suplementar 10). O número de PrSMs e proteínas identificadas em cada fração apresentou um decréscimo entre as frações com proteínas de menor massa para as frações com proteínas de maior massa (figura 42 A). A identificação de proteoformas com alta massa apresenta um desafio para a proteômica *top-down*, principalmente proteoformas com massa maior que 30 kDa (Fornelli & Toby, 2022). Essa limitação pode ser causada pela diluição de sinal entre isótopos e envelopes de cargas, além de decaimento de proteínas por colisão (Fornelli & Toby, 2022). Apesar dessas limitações a identificação de proteoformas com massa maior que 30 kDa foi observada, mais especificamente nas frações 5, 7, e 11 (figura 42 B). Assim como esperado a densidade de massas das proteoformas identificadas até a fração 5 apresentou claro aumento na região de alta massa de acordo com o aumento da fração. No entanto, pequena redução foi observada nas frações 7 e 9, voltando a aumentar a densidade na região de alta massa na fração 11 (figura 42 B). Além disso, as frações 13, 15 e 20 apresentaram densidade de massa das proteoformas até abaixo das proteoformas da fração 0, sugerindo grande dificuldade para identificação de suas proteoformas de alta massa (figura 42 B).

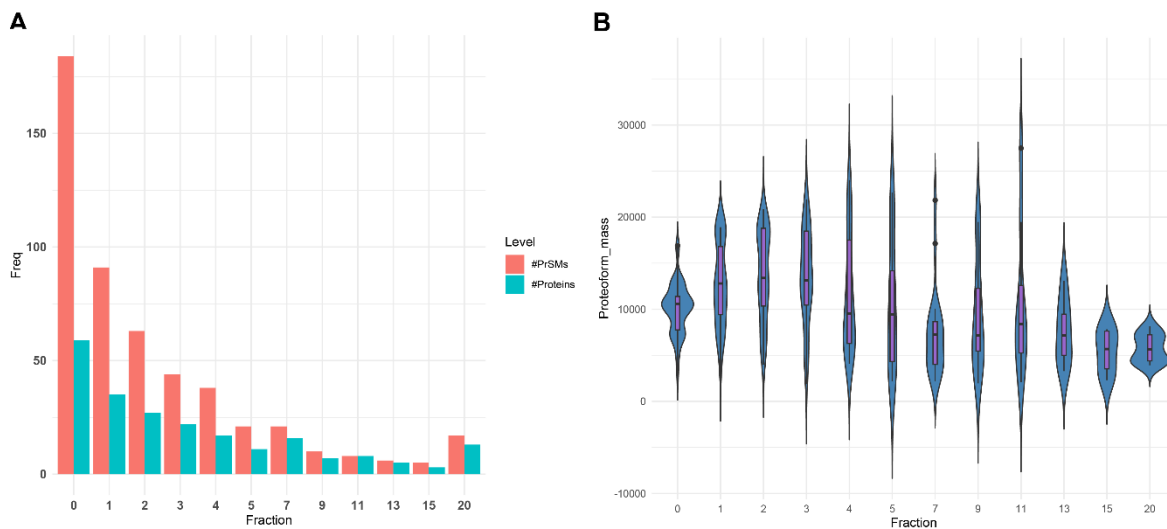


Figura 42. Análise dos PrSMs, proteoformas e proteínas identificadas em cada fração. A) Número de PrSMs e proteínas identificadas em cada fração. B) *Violin plot* representando a distribuição de massa das proteoformas presentes em cada fração.

Foram identificados 319 PrSMs e 104 proteínas contendo variação de massa (Δm), sugerindo presença de PTMs (Tabela suplementar 11). As proteínas identificadas com Δm estão relacionadas a família GroES chaperonina, proteínas ribossomais, histonas e enovelamento de histonas (figura 43), sugerindo presença de PTMs em proteínas relacionadas a esses processos.

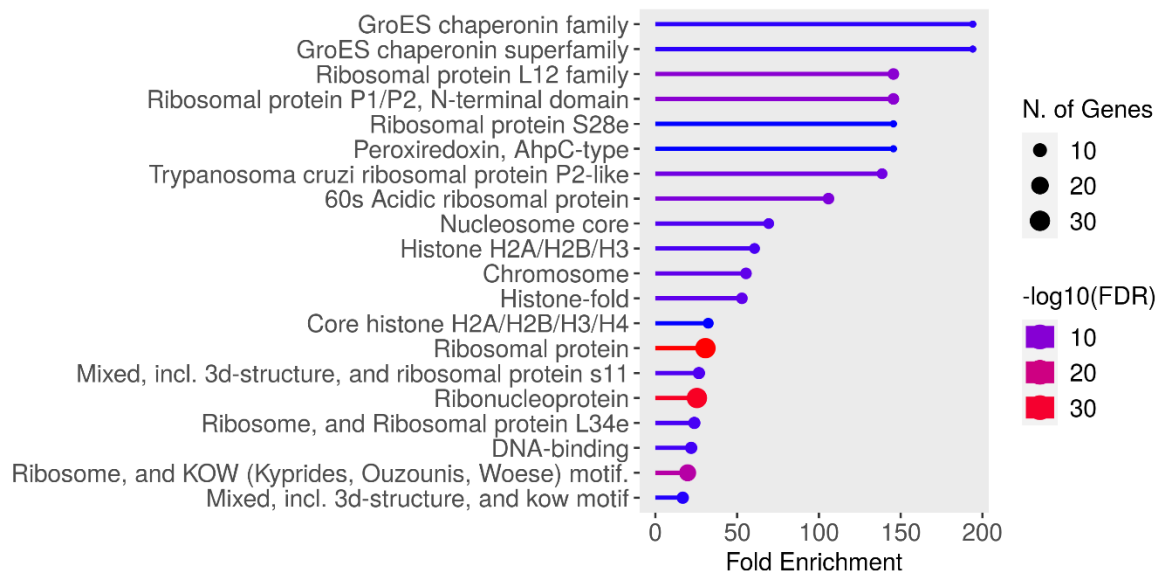


Figura 43. Análise de super-representação das proteínas identificadas com variação de massa (Δm). Foram considerados termos super-representados se q-valor < 0,05 e mínimo de duas proteínas no termo. Análise realizada com ShinyGO 0.77 (Ge et al., 2020).

Variações de massas (Δm) mais frequentes

As variações de massas mais frequentes foram 27 Da, - 57 Da, 26 Da, -29 Da, -103 Da, -14 Da, -23 Da, -28 Da, e -85 Da. Além das proteínas com Δm , foram identificadas 43 proteínas (213 PrSMs) com N-terminal acetilado (figura 44 A). As modificações por 27 Da, 26 Da, -29 Da, -28 Da, -23 Da, -14 Da e -85 Da podem ser causadas por substituições (figura 44 B). Algumas dessas possibilidades estão discutidas adiante. Ademais, a modificação por -57 Da pode ser indicativo de proteínas com alquilação incompleta. Congruente com essa hipótese, todas as proteoformas identificadas com $\Delta m = -57$ apresentaram resíduos de cisteína (Tabela suplementar 12). Apesar disso, apenas duas proteínas foram identificadas com esse possível artefato, sugerindo que a alquilação das proteínas ocorreu de maneira eficiente.

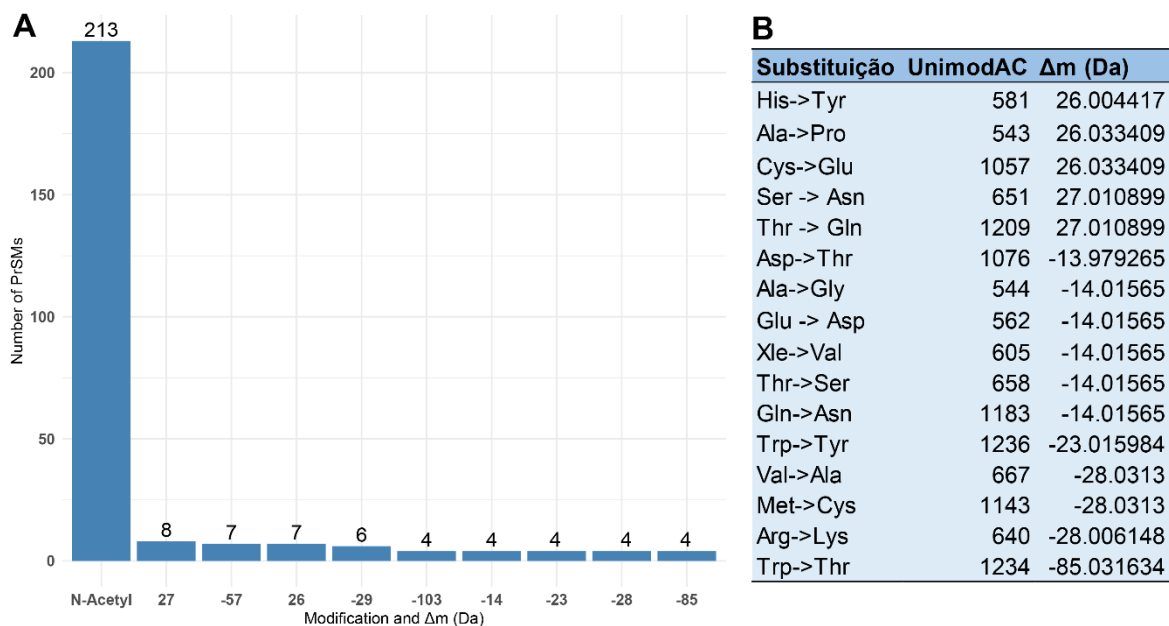


Figura 44. Análise dos PrSMs com Δm mais frequentes ou N-terminal acetilado. A) O número de PrSMs para as variações de massas foi calculado considerando o valor de Δm arredondado, sem casas decimais. B) Possíveis causas para as variações de massas decorrentes de substituições de resíduos de aminoácido. Obtido em unimod.org.

As proteínas identificadas com o N-terminal acetilado foram clusterizadas em quatro grupos, relacionados com proteínas ribossomais, peroxiredoxina, catabolismo de tirosina e nucleossomo (figura 45), sugerindo que esses processos estão suscetíveis a essa modificação. A acetilação do N-terminal de proteínas é uma modificação predominantemente co-traducional, ocorrendo em cerca de 80 % das proteínas humanas. Essa modificação é orquestrada por diferentes NATs (*N-terminal Acetyl Transferases*), e já foi relacionada a alterações no envelhecimento de proteínas, tempo de vida e capacidade de ligação (Ree, Varland, & Arnesen, 2018).

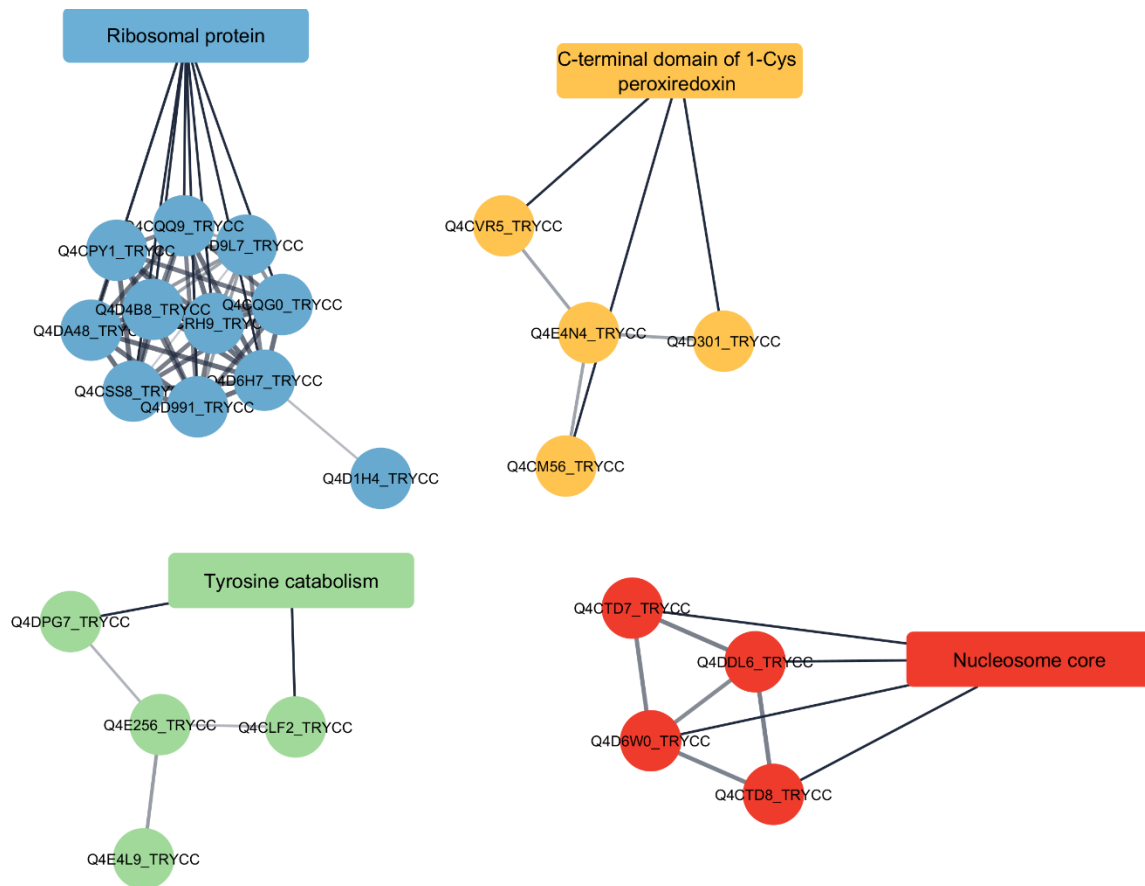


Figura 45. Análise de função das proteoformas com N-terminal acetilado. Códigos das proteínas foram submetidos a análise de interação e clusterização (inflation = 10) em cytoscape stringAPP (Nadezhda T. Doncheva, Morris, Gorodkin, & Jensen, 2019).

Possíveis substituições de resíduos de aminoácidos

Foram identificadas quatro proteínas com a variação de massa de 27 Da, entre elas, três são descritas como proteínas *heat shock* (HSP) putativas de 10 kDa. Apesar de possuírem códigos diferentes, todas apresentaram a variação de massa aproximadamente no mesmo local, e clivagem de aminoácidos próximos ao N-terminal. No entanto, a proteoforma da proteína Q4D000 foi identificada com clivagem dos 89 resíduos próximos ao N-terminal (figura 46 B), enquanto as proteoformas de Q4DFB0 e Q4DFA8 estavam sem os 9 primeiros resíduos (figura

46 A). Em análise de alinhamento global da sequência de resíduos das proteoformas identificadas foi observado que as sequências de duas são idênticas, enquanto uma tem a diferença de um resíduo de aminoácido, ácido aspártico para asparagina (D57->N57) (figura 46), essa diferença ocasiona em diferença de apenas 0,98401 Da. Sendo assim, a sequência das proteoformas de Q4DFB0 e Q4DFA8 são virtualmente idênticas, logo os PrSMs de suas identificações são compartilhados. Além disso, a diferença de massa dos precursores dessa proteína para a proteoforma de Q4DFA8 é muito pequena. Considerando a Δm de aproximadamente 27 Da identificada nessas proteoformas, é possível que tenha ocorrido dois tipos de substituição nessas sequências: Ser->Asn (UnimodAC: 651; $\Delta m = 27.010899$) ou Thr->Gln (UnimodAC: 1209; $\Delta m = 27.010899$). Levando em consideração as regiões identificadas com essa variação de massa, é provável que um dos resíduos de serina (S77/S78, usando Q4DFB0 como referência) sofra essa substituição. HSPs são proteínas com papel fundamental no enovelamento, degradação, secreção, e localização espacial de proteínas (Folgueira & Requena, 2007). Essas proteínas costumam estar relacionadas a mudanças de temperatura em parasitos com ciclo de vida digenético, sendo fortemente induzidas durante mudanças de temperaturas, por exemplo quando o parasito transita de seu vetor inseto para seu hospedeiro mamífero (Folgueira & Requena, 2007). O genoma de *T. cruzi* apresenta três cópias do gene *HSP10* com sequência de nucleotídeos idênticas, a não ser por uma substituição silenciosa (154G->A) (Fernandes et al., 2005). A região dessa mutação silenciosa é bem distante da região codificante da diferença entre as proteoformas identificadas aqui. Sendo assim, foram apresentadas evidências em nível de proteínas da expressão e tradução de proteoformas de HSP10 com D57, e N57, além de uma provável substituição de serina por asparagina (S77/S88-> N77/N78). Ademais, todas proteoformas foram identificadas com clivagem dos 9 resíduos do N-terminal, sugerindo processamento em nível de mRNA ou pós-traducional.

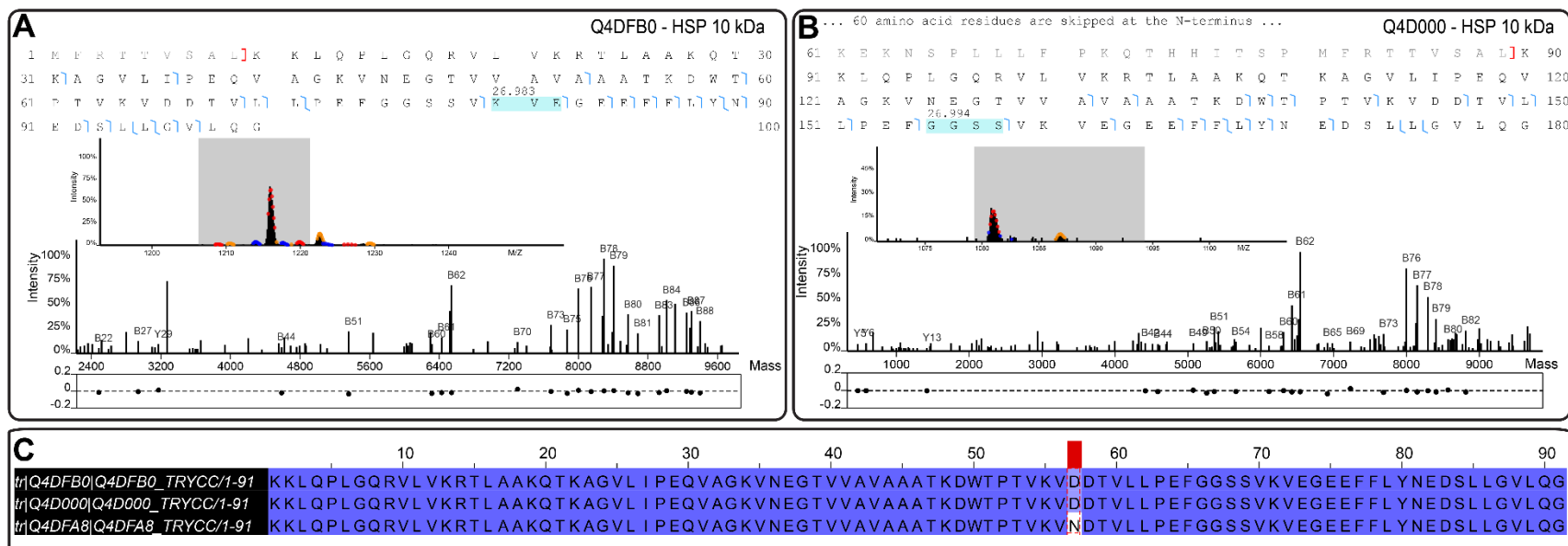


Figura 46. Análise de espectros e sequências das proteoformas identificadas com Δm de 27 Da. Em A e B são apresentadas análises dos PrSMs identificados. No topo dessas imagens está a sequência identificada, com traços em azul representando os fragmentos identificados, destaque em azul na região de identificação para a variação de massa, e colchetes em vermelho representando onde ocorreu uma possível clivagem da proteína. No meio das imagens, é mostrado o MS1 e a janela de exclusão utilizada no precursor identificado. Na parte inferior está o espectro de massas e gráfico de erros em ppm da massa dos fragmentos identificados. A) Análise de PrSM de HSP 10 kDa (UniprotAC: Q4DFB0). B) Análise de PrSM de HSP 10 kDa (UniprotAC: Q4D000). C) Alinhamento global entre as três proteínas HSP 10 kDa (UniprotACs: Q4DFB0; Q4D000; Q4DFA8) identificadas com Δm de 27 Da. A região onde ocorre 100 % de identidade está preenchida em azul, e onde não ocorre está destacada em vermelho. Alinhamento realizado pelo programa Clustal omega (Sievers & Higgins, 2018) e imagem gerada por Jalview (Waterhouse, Procter, Martin, Clamp, & Barton, 2009).

Proteínas identificadas com maior número de proteoformas

Muitas proteínas foram identificadas por mais de uma proteoformas, mais especificamente, 43 proteínas foram identificadas com 2 ou mais proteoformas (Tabela suplementar 13). Dessas, nove foram identificadas por pelo menos 4 proteoformas (tabela 6). Entre essas proteínas, a quantidade de proteoformas identificadas para triparedoxina peroxidase putativa (TxP, Q4CM56) e peptidil-prolil cis-trans isomerase (PPIase, Q4E4L9) chama atenção.

Tabela 6. Proteínas identificadas com mais do que 3 proteoformas

Protein accession	Protein description	#Proteoforms
Q4CM56	Tryparedoxin peroxidase, putative	21
Q4E4L9	Peptidyl-prolyl cis-trans isomerase	17
Q4CLU9	40S ribosomal protein S8	5
Q4CWE9	Histone H3, putative	5
Q4D991	60S acidic ribosomal protein, putative	5
Q4CU61	60S ribosomal protein L5, putative	4
Q4DJB6	14-3-3 protein, putative	4
Q4E0P2	60S ribosomal protein L28, putative	4
Q4E4N4	Eukaryotic translation initiation factor 5A	4

A TxP foi identificada por proteoformas com grande diferença de massas, entre 2,16 kDa e 13,38 kDa, mas nenhuma com a massa esperada de sua proteoforma canônica (22,39 kDa). Apesar de ter seus fragmentos identificados em MS2, os precursores das proteoformas identificadas para a TxP apresentaram baixa relação sinal/ruído (S/N – *Signal to Noise*), o que pode ter dificultado a deconvolução dos espectros, e conseqüentemente a determinação da massa do precursor. Tendo em vista essa limitação, as diferentes proteoformas apresentadas para essa proteína podem ser artefato de deconvolução devido ao seu baixo S/N. Por outro lado, a PPIase foi identificada por proteoforma com massa próxima da derivada de sua sequência canônica, com diferença ocasionada pela clivagem da metionina e acetilação da serina no N-terminal (figura 47 A). No entanto, diversas proteoformas de PPIase também apresentaram massa muito menor do que a da proteoforma

canônica, ao observar os seus precursores pode-se notar novamente um padrão de proteoformas com baixa razão S/N (figura 47 B). Esses eventos de baixo S/N podem dificultar a identificação de proteoformas e podem ser causados pela diluição de sinal em proteoformas de alta massa (Fornelli & Toby, 2022). Para avaliar a pertinência dessa hipótese, foram analisados os tempos de retenção (RT) das proteoformas de PPlase e TxP, demonstrando poucas variações de RT oscilando em torno de 30 segundos, mesmo com as diferentes massas das proteoformas de TxP (figura 48 A) e PPlase (figura 48 B). Padrão semelhante que sugerem a presença desses artefatos foram observados nas proteoformas da 40S proteína ribossomal S8 (Q4CLU9), histona H3 putativa (Q4CWE9), proteínas ribossomal L5 putativa (Q4CU61), proteína 14-3-3 putativa (Q4DJB6), proteína ribossomal L28 60S (Q4E0P2), fator de iniciação da tradução 5^a (Q4E4N4) e 3-oxo-5alfa-esteróide 4-desidrogenase putativa (Q4CM68).

A dificuldade de identificação de proteoformas de alta massa podem ser atenuadas por métodos como o aumento no número de *uscan*s em MS1, ou redução da resolução (Fornelli & Toby, 2022). Métodos mais modernos como análise de proteoformas nativas ou *próton-transfer reaction* (PTR) também podem auxiliar nessas identificações (Donnelly et al., 2019b; Fornelli & Toby, 2022; Kline et al., 2021). Ademais, recentemente foi descrita a complementariedade entre os algoritmos de deconvolução de proteoformas (McIlwain et al., 2020; Tabb et al., 2023), sendo assim, abordagens com outros algoritmos podem resultar em identificações complementares/acuradas.

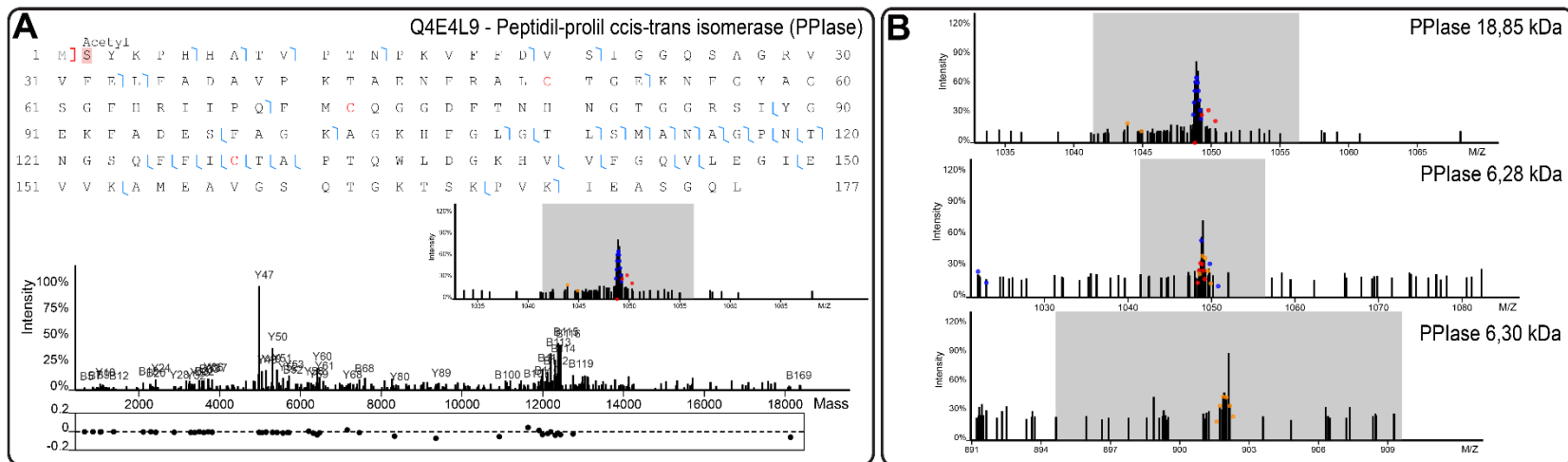


Figura 47. Análise dos PrSMs identificadores da proteína peptidil-prolil cis-trans isomerase (PPIase, Q4E4L9). A) Análise de PrSM da proteoforma identificada como mais próxima da canônica. No superior está a sequência da proteína com os fragmentos identificados em MS2 representados por traços azuis, possível perda de resíduos apresentada por um colchete em vermelho e local de modificação por acetilação do N-terminal apresentada por um sombreamento em vermelho. No meio está o precursor dos fragmentos identificados dessa proteoforma. Na parte inferior está apresentado o espectro de massas dos fragmentos identificados juntos a sua barra de erros (ppm). B) Precursores identificadores de diferentes proteoformas de PPIase.

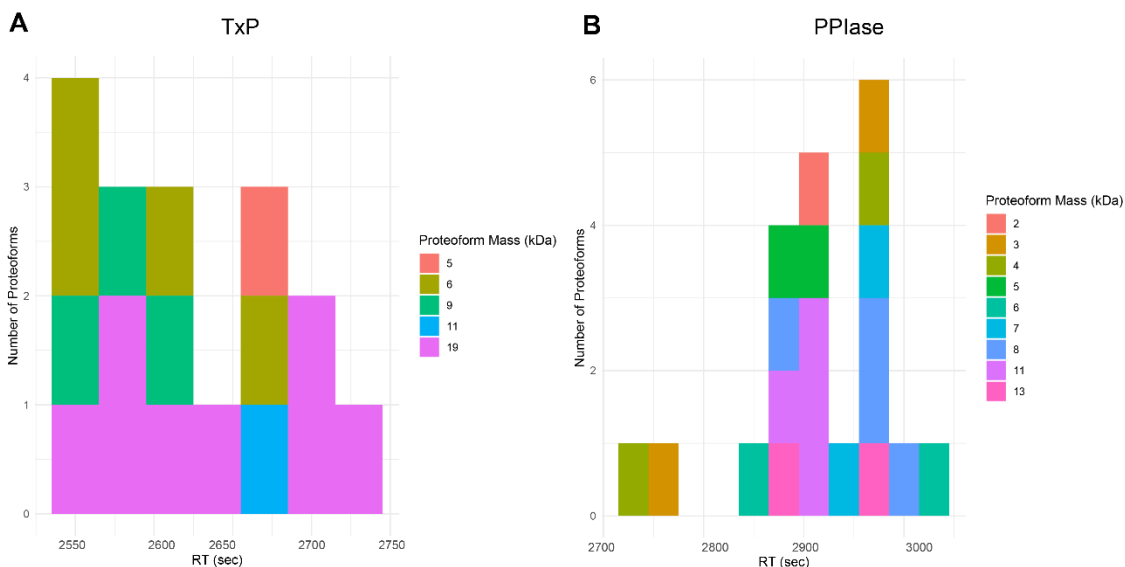


Figura 48. Análise do tempo de retenção (RT) das proteoformas de TxP e PPlase de acordo com suas massas.

Por outro lado, a proteína ácida ribossomal 60S putativa foi identificada com proteoformas que sugerem PTM com Δm de 16 Da, 222 Da e 58 Da, todas identificadas com excisão da metionina e N-terminal acetilado. A modificação por 16 Da provavelmente é causada por oxidação. Já o Δm de 58 Da pode ser atribuído às seguintes substituições: Ala->Glu (UnimodAC: 545; $\Delta m = 58,005479$ Da), Gly- >Asp (UnimodAC: 576; $\Delta m = 58,005479$), e Gln->Trp (UnimodAC: 1187; $\Delta m = 58,020735$). Por outro lado, o Δm de 222 Da não tem nenhuma modificação anotada no Unimod, sugerindo que mais de uma PTM pode estar causando essa variação.

Proteoformas identificadas por MetaMorpheus

Considerando a complementariedade entre os programas de identificação e deconvolução, foi utilizado o programa MetaMorpheus para deconvolução e identificação de proteoformas de *T. cruzi*. O programa MetaMorpheus utiliza uma abordagem restrita a erro em suas buscas, objetivando a identificação de proteoformas anotadas nos bancos de dados. Utilizando essa abordagem, foram identificados 21 PrSMs, 9 proteoformas e 9 proteínas, a maioria relacionada com

ribossomo ou processo de tradução (Tabela suplementar 14). *T. cruzi* possui número extremamente baixo de PTMs anotadas em seu proteoma de referência (UP000002296), para fins de comparação, o proteoma do organismo modelo *Saccharomyces cerevisiae* (UP000002311) foi usado como referência (figura 49). Apesar de ambos serem eucariotos unicelulares, a diferença em todos os níveis de PTMs anotadas no Uniprot é perceptível. Considerando o baixo número de PTMs anotadas no proteoma de referência de *T. cruzi* CL Brener (UP000002296), a redução no número de identificações pelo software MetaMorpheus em comparação ao TopPic suite é esperada, principalmente, devido a sua característica de identificação por busca restrita. Apesar disso, a redução no número de identificações foi muita drástica, diferente do observado para *C. glutamicum*, apresentado no capítulo II sessão “Integração de proteômica *bottom-up* multi-enzimática e *top-down* para caracterização de proteoformas da bactéria *C. glutamicum*”, que demonstrou a redução de 7351PrSms, 1363 proteoformas e 382 proteínas identificadas pelo programa TopPic Suite para 1619 PrSms, 129 proteoformas e 125 proteínas identificadas pelo programa MetaMorpheus. Essa diferença e o número de proteoformas identificadas com baixa relação S/N sugerem que a abordagem utilizada para identificação de proteoformas de *T. cruzi* ainda pode ser otimizada. Modificações no protocolo de aquisição já sugeridas como redução da resolução e aumento no número de *uscsans* podem auxiliar. Além disso, otimizações no preparo de amostra, como a utilização de métodos de solubilização por ácido fórmico gelado (Doucette et al., 2014) para ressuspender proteínas hidrofóbicas após precipitação podem melhorar a recuperação, e consequentemente o S/N de algumas proteínas.



Figura 49. Análise de PTMs anotadas em bancos de dados de *T. cruzi* e *S. cerevisiae*. Informações obtidas do UniprotKB a partir dos proteomas de referência de *T. cruzi* CL Brener (UP000002296) e *S. cerevisiae* (UP000002311) (Bateman et al., 2021).

Integração de abordagem *bottom-up* multi-enzimática com *top-down*

A digestão das proteínas de *T. cruzi* utilizando seis enzimas (Arg-C, Asp-N, Glu-C, Lys-C, quimiotripsina e tripsina) e o fracionamento dos peptídeos resultantes de cada digestão por fase reversa em pH básico já foram realizados. As análises LC-MS/MS estão em andamento, apresentando grande potencial para descoberta de PTMs e anotação no banco de dados de *T. cruzi* por meio da abordagem G-PTM-D (Solntsev et al., 2018). O uso de banco de dados anotados com PTMs para a identificação de proteoformas de *T. cruzi* pode resultar em aumento significativo

da cobertura de PrSMs, proteoformas e proteínas identificadas, assim como apresentado de maneira preliminar para *C. glutamicum* no capítulo II, sessão “Integração de proteômica bottom-up multi-enzimática e top-down para caracterização de proteoformas da bactéria *C. glutamicum*”. Para *C. glutamicum*, com dados preliminares, utilizando apenas os dados das enzimas tripsina, Lys-C e Glu-C, foi possível obter aumento de 157 %, 72 % e 37 % nas identificações de PrSMs, proteoformas e proteínas, respectivamente.

5.2.4 Conclusões

A análise proteômica *top-down* de *T. cruzi* foi capaz de identificar proteoformas relacionadas a processos de tradução, e resposta a estresses ambientais. Especificamente, foram identificadas proteoformas da HSP10 e proteínas ribossomais que sugerem diferentes tipos de substituição de resíduos de aminoácidos. Além disso, foram detectadas diversas proteoformas com N-terminal acetilado relacionadas com proteínas ribossomais, peroxiredoxina, catabolismo de tirosina e nucleossomo. Entretanto, a análise apontou indícios de que melhorias podem ser implementadas tanto no processamento das amostras quanto na análise dos dados. Por exemplo, a diferença no número de identificações entre os programas TopPic Suite e MetaMorpheus sugere que diversas proteoformas com PTMs não anotadas no banco de dados de *T. cruzi* não estão sendo identificadas. A identificação de PTMs e anotação do banco de dados, como realizada por abordagem G-PTM-D, pode resultar em aumento do número de identificações. Por outro lado, otimizações dos métodos de aquisição e preparo de amostra tem potencial para aperfeiçoar ainda mais a identificação de proteoformas presentes em *T. cruzi*.

Capítulo VI. Considerações finais e perspectivas

Os estudos proteômicos apresentados nessa tese contribuíram de forma significativa para a compreensão do metabolismo da bactéria *C. glutamicum* e para o apontamento de diversas proteínas com relevância no processo de replicação em *T. cruzi*. Além disso também foram apresentados dados que suportam a integração de proteômica *bottom-up* e *top-down*, principalmente quando utilizando abordagem de identificação restritas a erro.

Sobre *C. glutamicum*, foram identificadas pela primeira vez proteoformas de OdhI e mepB que sugerem potenciais mecanismos de regulação a serem explorados nessas proteínas. Ademais, duas proteínas pouco caracterizadas, uma proteína de micomembrana e outra secretada, apresentaram possível influência no mecanismo inicial que desencadeia a produção de glutamato por penicilina-G. Diversos aspectos desse mecanismo já foram elucidados, como a cadeia de sinalização proposta envolvendo GlnX, GlnH, PknG e OdhI (Sundermeyer et al., 2022). Além desse, a influência das modificações de OdhI por fosforilação (Boulaya et al., 2010) e succinilação (Komine-Abe et al., 2017) já foram descritas na produção de glutamato. No entanto, o mecanismo exato por trás do início desse processo continua obscuro e deve ser alvo de estudos futuros (Sundermeyer et al., 2022). Considerando possível relevância das proteínas periplasmática e extracelular reguladas no início da indução para produção de glutamato, essas podem participar de alguma maneira do processo que desencadeia o aumento da produção de glutamato. Estudos futuros devem ser conduzidos para verificar a função dessas proteínas.

Em relação aos estudos proteômicos de *T. cruzi*, foi apresentado um método para isolamento/enriquecimento de amastigotas intracelulares e avaliação desse em nível molecular, apresentando enriquecimento de aproximadamente 90 % das proteínas do parasito em comparação proteínas da célula hospedeira. No mesmo estudo, os amastigotas intracelulares purificados foram comparados com amastigotas axênicos obtidos a partir de amastigogênese em pH ácido (Tomlinson et al., 1995), revelando diferenças significativas entre essas células. Para finalizar, foi discutida a relevância de proteínas relacionadas ao kDNA, *Leishmanolysin-like*

peptidases e subfamílias de trans-sialidases na replicação de amastigotas. Por sua vez, a análise proteômica *top-down* da forma de vida epimastigota de *T. cruzi* revelou algumas proteoformas possivelmente decorrentes de substituições e apontou possíveis meios para otimização dessa análise.

Por fim, a integração das abordagens proteômicas *bottom-up* e *top-down* em *C. glutamicum* foi capaz de aumentar significativamente o número de PrSMs, proteoformas e proteínas identificadas. Mais especificamente, a abordagem *bottom-up* para identificação de PTMs em peptídeos, por meio digestão multi-enzimática acoplada a *enhanced* G-PTM-D (Miller et al., 2019), seguida pela criação de bancos de dados expandidos foi o que resultou nesse aumento. A utilização desses dados para criação de um banco de dados reduzido, e subsequente busca de proteoformas por meio da análise de massas intactas com Proteoform Suite (Cesnik et al., 2018; Yunxiang Dai et al., 2019) pode resultar em otimização da caracterização de proteoformas. De maneira mais geral, a abordagem para expansão de bancos utilizada parece muito promissora para identificação de proteoformas em organismos pouco caracterizados, com poucas PTMs anotadas em seus bancos de dados, quando utilizando algoritmos de identificação não tolerantes a erro.

Referências

- Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198–207. <https://doi.org/10.1038/nature01511>
- Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620), 347–355. <https://doi.org/10.1038/nature19949>
- Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., ... Wang, H. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Research*, 38(suppl_1), D457–D462. <https://doi.org/10.1093/nar/gkp851>
- Barthe, P., Roumestand, C., Canova, M. J., Kremer, L., Hurard, C., Molle, V., & Cohen-Gonsaud, M. (2009). Dynamic and Structural Characterization of a Bacterial FHA Protein Reveals a New Autoinhibition Mechanism. *Structure*, 17(4), 568–578. <https://doi.org/10.1016/j.str.2009.02.012>
- Basharat, A. R., Zang, Y., Sun, L., & Liu, X. (2023). TopFD: A Proteoform Feature Detection Tool for Top–Down Proteomics. *Analytical Chemistry*, 95(21), 8189–8196. <https://doi.org/10.1021/acs.analchem.2c05244>
- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., ... Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- Batista, C. M., Kessler, R. L., Eger, I., & Soares, M. J. (2015). Trypanosoma cruzi Intracellular Amastigotes Isolated by Nitrogen Decompression Are Capable of Endocytosis and Cargo Storage in Reservosomes. *PLOS ONE*, 10(6), e0130165. <https://doi.org/10.1371/journal.pone.0130165>
- Becker, J., & Wittmann, C. (2012). Bio-based production of chemicals, materials and fuels – *Corynebacterium glutamicum* as versatile cell factory. *Current Opinion in Biotechnology*, 23(4), 631–640. <https://doi.org/10.1016/j.copbio.2011.11.012>
- Beckers, G., Burkovski, A., & Nolden, L. (2001). Glutamate synthase of *Corynebacterium glutamicum* is not essential for glutamate synthesis and is regulated by the nitrogen status. *Microbiology*, 147(11), 2961–2970. <https://doi.org/10.1099/00221287-147-11-2961>
- Boulahya, K. A., Guedon, E., Delaunay, S., Schultz, C., Boudrant, J., Bott, M., & Goergen, J. L. (2010). OdhI dephosphorylation kinetics during different glutamate production processes involving *Corynebacterium glutamicum*. *Applied Microbiology and Biotechnology*, 87(5), 1867–1874. <https://doi.org/10.1007/s00253-010-2599-y>
- Brown, C. W., Sridhara, V., Boutz, D. R., Person, M. D., Marcotte, E. M., Barrick, J. E., & Wilke, C. O. (2017). Large-scale analysis of post-translational modifications in *E.coli* under glucose-limiting conditions. *BMC Genomics*, 18(1),

1–21. <https://doi.org/10.1186/s12864-017-3676-8>

- Brown, K. A., Chen, B., Guardado-Alvarez, T. M., Lin, Z., Hwang, L., Ayaz-Guner, S., ... Ge, Y. (2019). A photocleavable surfactant for top-down proteomics. *Nature Methods*, *16*(5), 417–420. <https://doi.org/10.1038/s41592-019-0391-1>
- Burkovski, A. (2003). Ammonium assimilation and nitrogen control in *Corynebacterium glutamicum* and its relatives: An example for new regulatory mechanisms in actinomycetes. *FEMS Microbiology Reviews*, *27*(5), 617–628. [https://doi.org/10.1016/S0168-6445\(03\)00067-6](https://doi.org/10.1016/S0168-6445(03)00067-6)
- Burle-Caldas, G. de A., dos Santos, N. S. A., de Castro, J. T., Mugge, F. L. B., Grazielle-Silva, V., Oliveira, A. E. R., ... Teixeira, S. M. R. (2022). Disruption of Active Trans-Sialidase Genes Impairs Egress from Mammalian Host Cells and Generates Highly Attenuated *Trypanosoma cruzi* Parasites. *MBio*, *13*(1). <https://doi.org/10.1128/MBIO.03478-21>
- Cai, W., Tucholski, T., Chen, B., Alpert, A. J., McIlwain, S., Kohmoto, T., ... Ge, Y. (2017). Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. *Analytical Chemistry*, *89*(10), 5467–5475. <https://doi.org/10.1021/acs.analchem.7b00380>
- CAMARGO, E. P. (1964). GROWTH AND DIFFERENTIATION IN TRYPANOSOMA CRUZI. I. ORIGIN OF. *Revista Do Instituto de Medicina Tropical de São Paulo*, *12*, 93–100. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14177814>
- Carbonara, K., Andonovski, M., & Coorsen, J. R. (2021). Proteomes are of proteoforms: Embracing the complexity. *Proteomes*, *9*(3). <https://doi.org/10.3390/proteomes9030038>
- Carvalho, T. U., & Souza, W. (1983). Separation of amastigotes and trypomastigotes of *Trypanosoma cruzi* from cultured cells. *Zeitschrift Für Parasitenkunde Parasitology Researchr Parasitenkunde Parasitology Research*, *69*(5), 571–575. <https://doi.org/10.1007/BF00926668>
- Catherman, A. D., Skinner, O. S., & Kelleher, N. L. (2014). Top Down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications*, *445*(4), 683–693. <https://doi.org/10.1016/j.bbrc.2014.02.041>
- Cesnik, A. J., Shortreed, M. R., Schaffer, L. V., Knoener, R. A., Frey, B. L., Scalf, M., ... Smith, L. M. (2018). Proteoform Suite: Software for Constructing, Quantifying, and Visualizing Proteoform Families. *Journal of Proteome Research*, *17*(1), 568–578. <https://doi.org/10.1021/acs.jproteome.7b00685>
- Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., ... Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, *30*(10), 918–920. <https://doi.org/10.1038/nbt.2377>
- Chamot-Rooke, J., Mikaty, G., Malosse, C., Soyer, M., Dumont, A., Gault, J., ... Duménil, G. (2011). Posttranslational Modification of Pili upon Cell Contact

- Triggers N. meningitidis Dissemination. *Science*, 331(6018), 778–782. <https://doi.org/10.1126/science.1200729>
- Chávez, S., Eastman, G., Smircich, P., Becco, L. L., Oliveira-Rizzo, C., Fort, R., ... Duhagon, M. A. (2017). Transcriptome-wide analysis of the *Trypanosoma cruzi* proliferative cycle identifies the periodically expressed mRNAs and their multiple levels of control. *PLoS ONE*, 12(11), e0188441. <https://doi.org/10.1371/journal.pone.0188441>
- Clayton, C. (2019). Regulation of gene expression in trypanosomatids: Living with polycistronic transcription. *Open Biology*, 9(6). <https://doi.org/10.1098/rsob.190072>
- Cuevas, I. C., Cazzulo, J. J., & Sánchez, D. O. (2003). gp63 Homologues in *Trypanosoma cruzi*: Surface Antigens with Metalloprotease Activity and a Possible Role in Host Cell Infection. *Infection and Immunity*, 71(10), 5739–5749. <https://doi.org/10.1128/IAI.71.10.5739-5749.2003>
- Dai, Yuling, Millikin, R. J., Rolfs, Z., Shortreed, M. R., & Smith, L. M. (2022). A Hybrid Spectral Library and Protein Sequence Database Search Strategy for Bottom-Up and Top-Down Proteomic Data Analysis. *Journal of Proteome Research*, 21(11), 2609–2618. <https://doi.org/10.1021/acs.jproteome.2c00305>
- Dai, Yunxiang, Buxton, K. E., Schaffer, L. V., Miller, R. M., Millikin, R. J., Scalf, M., ... Smith, L. M. (2019). Constructing Human Proteoform Families Using Intact-Mass and Top-Down Proteomics with a Multi-Protease Global Post-Translational Modification Discovery Database. *Journal of Proteome Research*, 18(10), 3671–3680. <https://doi.org/10.1021/acs.jproteome.9b00339>
- Dai, Yunxiang, Shortreed, M. R., Scalf, M., Frey, B. L., Cesnik, A. J., Solntsev, S., ... Smith, L. M. (2017). Elucidating *Escherichia coli* Proteoform Families Using Intact-Mass Proteomics and a Global PTM Discovery Database. *Journal of Proteome Research*, 16(11), 4156–4165. <https://doi.org/10.1021/acs.jproteome.7b00516>
- de Godoy, L. M. F., Marchini, F. K., Pavoni, D. P., Rampazzo, R. de C. P., Probst, C. M., Goldenberg, S., & Krieger, M. A. (2012). Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis. *PROTEOMICS*, 12(17), 2694–2703. <https://doi.org/10.1002/pmic.201200078>
- Doncheva, N.T., Morris, J. H., Gorodkin, J., & Jensen, L. J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res*, 18(2), 1623–1632. <https://doi.org/10.1021/acs.jproteome.8b00702>
- Doncheva, Nadezhda T., Morris, J. H., Gorodkin, J., & Jensen, L. J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of Proteome Research*, 18(2), 623–632. <https://doi.org/10.1021/acs.jproteome.8b00702>
- Donnelly, D. P., Rawlins, C. M., DeHart, C. J., Fornelli, L., Schachner, L. F., Lin, Z., ... Agar, J. N. (2019a). Best practices and benchmarks for intact protein analysis

- for top-down mass spectrometry. *Nature Methods*, 16(7), 587–594. <https://doi.org/10.1038/s41592-019-0457-0>
- Donnelly, D. P., Rawlins, C. M., DeHart, C. J., Fornelli, L., Schachner, L. F., Lin, Z., ... Agar, J. N. (2019b). Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nature Methods*, 16(7), 587–594. <https://doi.org/10.1038/s41592-019-0457-0>
- Doucette, A. A., Vieira, D. B., Orton, D. J., & Wall, M. J. (2014). Resolubilization of Precipitated Intact Membrane Proteins with Cold Formic Acid for Analysis by Mass Spectrometry. *Journal of Proteome Research*, 13(12), 6001–6012. <https://doi.org/10.1021/pr500864a>
- Esquilin-Lebron, K., Dubrac, S., Barras, F., & Boyd, J. M. (2021). Bacterial Approaches for Assembling Iron-Sulfur Proteins. *MBio*, 12(6), 1–17. <https://doi.org/10.1128/mBio.02425-21>
- Fernandes, M., Silva, R., Rössle, S. C., Bisch, P. M., Rondinelli, E., & Ürményi, T. P. (2005). Gene characterization and predicted protein structure of the mitochondrial chaperonin HSP10 of *Trypanosoma cruzi*. *Gene*, 349, 135–142. <https://doi.org/10.1016/j.gene.2004.11.047>
- Ferri, G., & Edreira, M. M. (2021). All Roads Lead to Cytosol: *Trypanosoma cruzi* Multi-Strategic Approach to Invasion. *Frontiers in Cellular and Infection Microbiology*, 11(March), 1–17. <https://doi.org/10.3389/fcimb.2021.634793>
- Folgueira, C., & Requena, J. M. (2007). A postgenomic view of the heat shock proteins in kinetoplastids. *FEMS Microbiology Reviews*, 31(4), 359–377. <https://doi.org/10.1111/j.1574-6976.2007.00069.x>
- Fornelli, L., Durbin, K. R., Fellers, R. T., Early, B. P., Greer, J. B., LeDuc, R. D., ... Kelleher, N. L. (2017). Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *Journal of Proteome Research*, 16(2), 609–618. <https://doi.org/10.1021/acs.jproteome.6b00698>
- Fornelli, L., & Toby, T. K. (2022). Characterization of large intact protein ions by mass spectrometry: What directions should we follow? *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1870(4), 140758. <https://doi.org/10.1016/j.bbapap.2022.140758>
- Freitas, L. M., dos Santos, S. L., Rodrigues-Luiz, G. F., Mendes, T. A. O., Rodrigues, T. S., Gazzinelli, R. T., ... Bartholomeu, D. C. (2011). Genomic Analyses, Gene Expression and Antigenic Profile of the Trans-Sialidase Superfamily of *Trypanosoma cruzi* Reveal an Undetected Level of Complexity. *PLoS ONE*, 6(10), e25914. <https://doi.org/10.1371/journal.pone.0025914>
- Galili, T., O'Callaghan, A., Sidi, J., & Sievert, C. (2018). Heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, 34(9), 1600–1602. <https://doi.org/10.1093/bioinformatics/btx657>

- Gallagher, K. J., Palasser, M., Hughes, S., Mackay, C. L., Kilgour, D. P. A., & Clarke, D. J. (2020). Isotope Depletion Mass Spectrometry (ID-MS) for Accurate Mass Determination and Improved Top-Down Sequence Coverage of Intact Proteins. *Journal of the American Society for Mass Spectrometry*, 31(3), 700–710. <https://doi.org/10.1021/jasms.9b00119>
- Gamarro, F., Osuna, A., Castanys, S., Pérez-López, M. I., & Ruiz-Pérez, L. M. (1985). Isolation and purification of amastigotes of *Trypanosoma cruzi* from cultured Vero cells. *Zeitschrift Für Parasitenkunde Parasitology Research*, 71(1), 15–17. <https://doi.org/10.1007/BF00932914>
- Gao, Y., & Yates, J. R. (2019). Protein Analysis by Shotgun Proteomics. In *Mass Spectrometry-Based Chemical Proteomics* (pp. 1–38). <https://doi.org/10.1002/9781118970195.ch1>
- Gargano, A. F. G., Roca, L. S., Fellers, R. T., Bocxe, M., Domínguez-Vega, E., & Somsen, G. W. (2018). Capillary HILIC-MS: A New Tool for Sensitive Top-Down Proteomics. *Analytical Chemistry*, 90(11), 6601–6609. <https://doi.org/10.1021/acs.analchem.8b00382>
- Gault, J., Malosse, C., Machata, S., Millien, C., Podglajen, I., Ploy, M.-C., ... Chamot-Rooke, J. (2014). Complete posttranslational modification mapping of pathogenic *Neisseria meningitidis* pilins requires top-down mass spectrometry. *PROTEOMICS*, 14(10), 1141–1151. <https://doi.org/10.1002/pmic.201300394>
- Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36(8), 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>
- Georgi, T., Rittmann, D., & Wendisch, V. F. (2005). Lysine and glutamate production by *Corynebacterium glutamicum* on glucose, fructose and sucrose: Roles of malic enzyme and fructose-1,6-bisphosphatase. *Metabolic Engineering*, 7(4), 291–301. <https://doi.org/10.1016/j.ymben.2005.05.001>
- Giglione, C., Fieulaine, S., & Meinel, T. (2009). Cotranslational processing mechanisms: towards a dynamic 3D model. *Trends in Biochemical Sciences*, 34(8), 417–426. <https://doi.org/10.1016/j.tibs.2009.04.003>
- Gola, S., Munder, T., Casonato, S., Manganelli, R., & Vicente, M. (2015). The essential role of <scp>SepF</scp> in mycobacterial division. *Molecular Microbiology*, 97(3), 560–576. <https://doi.org/10.1111/mmi.13050>
- Grangeasse, C., Stülke, J., & Mijakovic, I. (2015). Regulatory potential of post-translational modifications in bacteria. *Frontiers in Microbiology*, 6(MAY), 2014–2015. <https://doi.org/10.3389/fmicb.2015.00500>
- Hallgren, J., Tsirigos, K. D., Damgaard Pedersen, M., Juan, J., Armenteros, A., Marcatili, P., ... Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *BioRxiv*, 2022.04.08.487609. <https://doi.org/10.1101/2022.04.08.487609>

- Han, D., Jin, J., Woo, J., Min, H., & Kim, Y. (2014). Proteomic analysis of mouse astrocytes and their secretome by a combination of FASP and StageTip-based, high pH, reversed-phase fractionation. *PROTEOMICS*, 14(13–14), 1604–1609. <https://doi.org/10.1002/pmic.201300495>
- Han, D., Moon, S., Kim, Y., Kim, J., Jin, J., & Kim, Y. (2013). In-depth proteomic analysis of mouse microglia using a combination of FASP and StageTip-based, high pH, reversed-phase fractionation. *Proteomics*, 13(20), 2984–2988. <https://doi.org/10.1002/pmic.201300091>
- Hashimoto, S. (2016). Discovery and History of Amino Acid Fermentation. In *Bioreactor Systems for Tissue Engineering II* (Vol. 123, pp. 15–34). https://doi.org/10.1007/10_2016_24
- Herreros-Cabello, A., Callejas-Hernández, F., Gironès, N., & Fresno, M. (2020). Trypanosoma cruzi genome: Organization, multi-gene families, transcription, and biological implications. *Genes*, 11(10), 1–26. <https://doi.org/10.3390/genes11101196>
- Hirasawa, T., Saito, M., Yoshikawa, K., Furusawa, C., & Shmizu, H. (2018). Integrated Analysis of the Transcriptome and Metabolome of Corynebacterium glutamicum during Penicillin-Induced Glutamic Acid Production. *Biotechnology Journal*, 13(5), 1700612. <https://doi.org/10.1002/biot.201700612>
- Hirasawa, T., & Shimizu, H. (2016). Recent advances in amino acid production by microbial cells. *Current Opinion in Biotechnology*, 42, 133–146. <https://doi.org/10.1016/j.copbio.2016.04.017>
- Houssin, C., de Sousa d'Auria, C., Constantinesco, F., Dietrich, C., Labarre, C., & Bayan, N. (2020). Architecture and Biogenesis of the Cell Envelope of Corynebacterium glutamicum. In M. Inui & K. Toyoda (Eds.), *Corynebacterium glutamicum: Biology and Biotechnology* (pp. 25–60). https://doi.org/10.1007/978-3-030-39267-3_2
- Huergo, L. F., Chandra, G., & Merrick, M. (2013). P II signal transduction proteins: nitrogen regulation and beyond. *FEMS Microbiology Reviews*, 37(2), 251–283. <https://doi.org/10.1111/j.1574-6976.2012.00351.x>
- Huguet, R., Mullen, C., Srzentić, K., Greer, J. B., Fellers, R. T., Zabrouskov, V., ... Fornelli, L. (2019). Proton Transfer Charge Reduction Enables High-Throughput Top-Down Analysis of Large Proteoforms. *Analytical Chemistry*, 91(24), 15732–15739. <https://doi.org/10.1021/acs.analchem.9b03925>
- Hüser, A. T., Chassagnole, C., Lindley, N. D., Merkamm, M., Guyonvarch, A., Elisáková, V., ... Pu, A. (2005). Rational Design of a Corynebacterium glutamicum Pantothenate Production Strain and Its Characterization by Metabolic Flux Analysis and Genome-Wide Transcriptional Profiling Rational Design of a Corynebacterium glutamicum Pantothenate Production Strain and. *Applied and Environmental Microbiology*, 71(6), 3255–3268. <https://doi.org/10.1128/AEM.71.6.3255>

- Ikedo, M., & Takeno, S. (2020). Recent Advances in Amino Acid Production. In M. Inui & K. Toyoda (Eds.), *Corynebacterium glutamicum: Biology and Biotechnology* (pp. 175–226). https://doi.org/10.1007/978-3-030-39267-3_7
- Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., ... Gaigalat, L. (2003). The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol*, *104*. [https://doi.org/10.1016/S0168-1656\(03\)00154-8](https://doi.org/10.1016/S0168-1656(03)00154-8)
- Kanzaki, T., Isobe, K., Okazaki, H., Motizuki, K., & Fukuda, H. (1967). L-Glutamic Acid Fermentation. *Agricultural and Biological Chemistry*, *31*(11), 1307–1317. <https://doi.org/10.1080/00021369.1967.10858972>
- Kaulich, P. T., Winkels, K., Kaulich, T. B., Treitz, C., Cassidy, L., & Tholey, A. (2022). MStoDiff: A Tool for the Visualization of Mass Shifts in Deconvoluted Top-Down Proteomics Data for the Database-Independent Detection of Protein Modifications. *Journal of Proteome Research*, *21*(1), 20–29. <https://doi.org/10.1021/acs.jproteome.1c00766>
- Keilhauer, C., Eggeling, L., & Sahl, H. (1993). Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon. *Journal of Bacteriology*, *175*(17), 5595–5603. <https://doi.org/10.1128/jb.175.17.5595-5603.1993>
- Kim, H., Dan, K., Shin, H., Lee, J., Wang, J. I., & Han, D. (2019). An efficient method for high-pH peptide fractionation based on C18 StageTips for in-depth proteome profiling. *Analytical Methods*. <https://doi.org/10.1039/c9ay01269a>
- Kim, J., Hirasawa, T., Saito, M., Furusawa, C., & Shimizu, H. (2011). Investigation of phosphorylation status of OdhI protein during penicillin- and Tween 40-triggered glutamate overproduction by *Corynebacterium glutamicum*. *Applied Microbiology and Biotechnology*, *91*(1), 143–151. <https://doi.org/10.1007/s00253-011-3275-6>
- Kim, J., Hirasawa, T., Sato, Y., Nagahisa, K., Furusawa, C., & Shimizu, H. (2009). Effect of *odhA* overexpression and *odhA* antisense RNA expression on Tween-40-triggered glutamate production by *Corynebacterium glutamicum*. *Applied Microbiology and Biotechnology*, *81*(6), 1097–1106. <https://doi.org/10.1007/s00253-008-1743-4>
- Kimura, E. (2002). Triggering mechanism of L-glutamate overproduction by DtsR1 in coryneform bacteria. *Journal of Bioscience and Bioengineering*, *94*(6), 545–551. [https://doi.org/10.1016/S1389-1723\(02\)80193-1](https://doi.org/10.1016/S1389-1723(02)80193-1)
- Kimura, E., Yagoshi, C., Kawahara, Y., Ohsumi, T., Nakamatsu, T., & Tokuda, H. (1999). Glutamate overproduction in *Corynebacterium glutamicum* triggered by a decrease in the level of a complex comprising *dtsR* and a biotin-containing subunit. *Bioscience, Biotechnology and Biochemistry*, *63*(7), 1274–1278. <https://doi.org/10.1271/bbb.63.1274>

- Kinugawa, H., Kondo, N., Komine-Abe, A., Tomita, T., Nishiyama, M., & Kosono, S. (2020). In vitro reconstitution and characterization of pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase hybrid complex from *Corynebacterium glutamicum*. *MicrobiologyOpen*, 9(10), 1–14. <https://doi.org/10.1002/mbo3.1113>
- Kline, J. T., Mullen, C., Durbin, K. R., Oates, R. N., Huguet, R., Syka, J. E. P., & Fornelli, L. (2021). Sequential Ion-Ion Reactions for Enhanced Gas-Phase Sequencing of Large Intact Proteins in a Tribrid Orbitrap Mass Spectrometer. *Journal of the American Society for Mass Spectrometry*, 32(9), 2334–2345. <https://doi.org/10.1021/jasms.1c00062>
- Komine-Abe, A., Kondo, N., Kubo, S., Kawasaki, H., Nishiyama, M., & Kosono, S. (2021). Characterization of lysine acetylation in the peripheral subunit-binding domain of the E2 subunit of the pyruvate dehydrogenase-2-oxoglutarate dehydrogenase hybrid complex from *Corynebacterium glutamicum*. *Bioscience, Biotechnology, and Biochemistry*, 85(4), 874–881. <https://doi.org/10.1093/bbb/zbaa114>
- Komine-Abe, A., Nagano-Shoji, M., Kubo, S., Kawasaki, H., Yoshida, M., Nishiyama, M., & Kosono, S. (2017). Effect of lysine succinylation on the regulation of 2-oxoglutarate dehydrogenase inhibitor, OdhI, involved in glutamate production in *Corynebacterium glutamicum*. *Bioscience, Biotechnology and Biochemistry*, 81(11), 2130–2138. <https://doi.org/10.1080/09168451.2017.1372182>
- Kou, Q., Xun, L., & Liu, X. (2016). TopPIC: A software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics*, 32(22), 3495–3497. <https://doi.org/10.1093/bioinformatics/btw398>
- Krawczyk, S., Raasch, K., Schultz, C., Hoffelder, M., Eggeling, L., & Bott, M. (2010). The FHA domain of OdhI interacts with the carboxyterminal 2-oxoglutarate dehydrogenase domain of OdhA in *Corynebacterium glutamicum*. *FEBS Letters*, 584(8), 1463–1468. <https://doi.org/10.1016/j.febslet.2010.03.028>
- Kulkarni, M. M., Olson, C. L., Engman, D. M., & McGwire, B. S. (2009). Trypanosoma cruzi GP63 Proteins Undergo Stage-Specific Differential Posttranslational Modification and Are Important for Host Cell Infection. *Infection and Immunity*, 77(5), 2193–2200. <https://doi.org/10.1128/IAI.01542-08>
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1), 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259), 680–685. <https://doi.org/10.1038/227680a0>
- Leandro de Jesus, T. C., Calderano, S. G., Vitorino, F. N. de L., Llanos, R. P., Lopes, M. D. C., de Araújo, C. B., ... Chagas da Cunha, J. P. (2017). Quantitative

- Proteomic Analysis of Replicative and Nonreplicative Forms Reveals Important Insights into Chromatin Biology of *Trypanosoma cruzi*. *Molecular & Cellular Proteomics*, 16(1), 23–38. <https://doi.org/10.1074/mcp.M116.061200>
- Lee, J. Y., Na, Y. A., Kim, E., Lee, H. S., & Kim, P. (2016). The actinobacterium *Corynebacterium glutamicum*, an industrial workhorse. *J Microbiol Biotechnol*, 26(5), 807–822. <https://doi.org/10.4014/jmb.1601.01053>
- Lennicke, C., Rahn, J., Heimer, N., Lichtenfels, R., Wessjohann, L. A., & Seliger, B. (2016). Redox proteomics: Methods for the identification and enrichment of redox-modified proteins and their applications. *PROTEOMICS*, 16(2), 197–213. <https://doi.org/10.1002/pmic.201500268>
- Letek, M., Fiuza, M., Ordóñez, E., Villadangos, A. F., Ramos, A., Mateos, L. M., & Gil, J. A. (2008). Cell growth and cell division in the rod-shaped actinomycete *Corynebacterium glutamicum*. *Antonie van Leeuwenhoek*, 94(1), 99–109. <https://doi.org/10.1007/s10482-008-9224-4>
- Leuchtenberger, W., Huthmacher, K., & Drauz, K. (2005). Biotechnological production of amino acids and derivatives: Current status and prospects. *Applied Microbiology and Biotechnology*, 69(1), 1–8. <https://doi.org/10.1007/s00253-005-0155-y>
- Lima, D. B., Dupré, M., Duchateau, M., Gai, Q., Rey, M., Matondo, M., & Chamorro, J. (2020). *ProteoCombiner: integrating bottom-up with top-down proteomics data for improved proteoform*. 4–6.
- Litsanov, B., Brocker, M., & Bott, M. (2012). Toward homosuccinate fermentation: Metabolic engineering of *Corynebacterium glutamicum* for anaerobic production of succinate from glucose and formate. *Applied and Environmental Microbiology*, 78(9), 3325–3337. <https://doi.org/10.1128/AEM.07790-11>
- Liu, Q., Ouyang, S. ping, Kim, J., & Chen, G. Q. (2007). The impact of PHB accumulation on l-glutamate production by recombinant *Corynebacterium glutamicum*. *Journal of Biotechnology*, 132(3), 273–279. <https://doi.org/10.1016/j.jbiotec.2007.03.014>
- Liu, Yadi, Li, Y., & Wang, X. (2016). Acetohydroxyacid synthases: evolution, structure, and function. *Applied Microbiology and Biotechnology*, 100(20), 8633–8649. <https://doi.org/10.1007/s00253-016-7809-9>
- Liu, Yansheng, Beyer, A., & Aebbersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, Vol. 165, pp. 535–550. <https://doi.org/10.1016/j.cell.2016.03.014>
- Locard-Paulet, M., Parra, J., Albigot, R., Mouton-Barbosa, E., Bardi, L., Burlet-Schiltz, O., & Marcoux, J. (2019). VisioProt-MS: Interactive 2D maps from intact protein mass spectrometry. *Bioinformatics*, 35(4), 679–681. <https://doi.org/10.1093/bioinformatics/bty680>
- Lubitz, D., & Wendisch, V. F. (2016). Ciprofloxacin triggered glutamate production

- by *Corynebacterium glutamicum*. *BMC Microbiology*, 16(1), 1–12. <https://doi.org/10.1186/s12866-016-0857-6>
- Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., & Woolf, P. J. (2009). GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*, 10, 1–17. <https://doi.org/10.1186/1471-2105-10-161>
- Ma, L., Chen, K., Meng, Q., Liu, Q., Tang, P., Hu, S., & Yu, J. (2011). An evolutionary analysis of trypanosomatid GP63 proteases. *Parasitology Research*, 109(4), 1075–1084. <https://doi.org/10.1007/s00436-011-2348-x>
- Mahne, M., Tauch, A., Puhler, A., & Kalinowski, J. (2006). The *Corynebacterium glutamicum* gene pmt encoding a glycosyltransferase related to eukaryotic protein- O -mannosyltransferases is essential for glycosylation of the resuscitation promoting factor (Rpf2) and other secreted proteins. *FEMS Microbiology Letters*, 259(2), 226–233. <https://doi.org/10.1111/j.1574-6968.2006.00269.x>
- Marchand, C. H., Salmeron, C., Bou Raad, R., Méniche, X., Chami, M., Masi, M., ... Bayan, N. (2012). Biochemical Disclosure of the Mycolate Outer Membrane of *Corynebacterium glutamicum*. *Journal of Bacteriology*, 194(3), 587–597. <https://doi.org/10.1128/JB.06138-11>
- Marques, A., Marques, A., Nakayasu, E., & Almeida, I. (2011). Purification of extracellular and intracellular amastigotes of *Trypanosoma cruzi* from mammalian host-infected cells. *Protocol Exchange*, 1–7. <https://doi.org/10.1038/protex.2011.265>
- Marufu, L., & Coetzer, T. H. T. (2022). Homology modelling of *Trypanosoma brucei* major surface proteases and molecular docking of variant surface glycoproteins and inhibitor ligands for drug design. *Journal of Molecular Graphics and Modelling*, 111(August 2021), 108104. <https://doi.org/10.1016/j.jmgm.2021.108104>
- Matsumoto, K., Kitagawa, K., Jo, S. J., Song, Y., & Taguchi, S. (2011). Production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) in recombinant *Corynebacterium glutamicum* using propionate as a precursor. *Journal of Biotechnology*, 152(4), 144–146. <https://doi.org/10.1016/j.jbiotec.2010.07.031>
- McCool, E. N., Lubeckyj, R. A., Shen, X., Chen, D., Kou, Q., Liu, X., & Sun, L. (2018). Deep Top-Down Proteomics Using Capillary Zone Electrophoresis-Tandem Mass Spectrometry: Identification of 5700 Proteoforms from the *Escherichia coli* Proteome. *Analytical Chemistry*, 90(9), 5529–5533. <https://doi.org/10.1021/acs.analchem.8b00693>
- McIlwain, S. J., Wu, Z., Wetzel, M., Belongia, D., Jin, Y., Wenger, K., ... Ge, Y. (2020). Enhancing Top-Down Proteomics Data Analysis by Combining Deconvolution Results through a Machine Learning Strategy. *Journal of the American Society for Mass Spectrometry*, 31(5), 1104–1113. <https://doi.org/10.1021/jasms.0c00035>

- Melby, J. A., Roberts, D. S., Larson, E. J., Brown, K. A., Bayne, E. F., Jin, S., & Ge, Y. (2021). Novel Strategies to Address the Challenges in Top-Down Proteomics. *Journal of the American Society for Mass Spectrometry*, 32(6), 1278–1294. <https://doi.org/10.1021/jasms.1c00099>
- Melo, R. M., de Souza, J. M. F., Williams, T. C. R., Fontes, W., de Sousa, M. V., Ricart, C. A. O., & do Vale, L. H. F. (2023). Revealing *Corynebacterium glutamicum* proteoforms through top-down proteomics. *Scientific Reports*, 13(1), 2602. <https://doi.org/10.1038/s41598-023-29857-6>
- Miller, R. M., Millikin, R. J., Hoffmann, C. V., Solntsev, S. K., Sheynkman, G. M., Shortreed, M. R., & Smith, L. M. (2019). Improved Protein Inference from Multiple Protease Bottom-Up Mass Spectrometry Data. *Journal of Proteome Research*, 18(9), 3429–3438. <https://doi.org/10.1021/acs.jproteome.9b00330>
- Millikin, R. J., Solntsev, S. K., Shortreed, M. R., & Smith, L. M. (2018). Ultrafast Peptide Label-Free Quantification with FlashLFQ. *Journal of Proteome Research*, 17(1), 386–391. <https://doi.org/10.1021/acs.jproteome.7b00608>
- Mizuno, Y., Nagano-Shoji, M., Kubo, S., Kawamura, Y., Yoshida, A., Kawasaki, H., ... Kosono, S. (2016). Altered acetylation and succinylation profiles in *Corynebacterium glutamicum* in response to conditions inducing glutamate overproduction. *MicrobiologyOpen*, 5(1), 152–173. <https://doi.org/10.1002/mbo3.320>
- Moretti, N. S., Mortara, R. A., & Schenkman, S. (2020). *Trypanosoma cruzi*. *Trends in Parasitology*, 36(4), 404–405. <https://doi.org/10.1016/j.pt.2019.10.002>
- Nagano-Shoji, M., Hamamoto, Y., Mizuno, Y., Yamada, A., Kikuchi, M., Shirouzu, M., ... Kosono, S. (2017). Characterization of lysine acetylation of a phosphoenolpyruvate carboxylase involved in glutamate overproduction in *Corynebacterium glutamicum*. *Molecular Microbiology*, 104(4), 677–689. <https://doi.org/10.1111/mmi.13658>
- Nakunst, D., Larisch, C., Hüser, A. T., Tauch, A., Pühler, A., & Kalinowski, J. (2007). The extracytoplasmic function-type sigma factor SigM of *Corynebacterium glutamicum* ATCC 13032 is involved in transcription of disulfide stress-related genes. *Journal of Bacteriology*, 189(13), 4696–4707. <https://doi.org/10.1128/JB.00382-07>
- Nara, T., Samejima, H., & Kinoshita, S. (1964). Effect of Penicillin on Amino Acid Fermentation. *Agricultural and Biological Chemistry*, 28(2), 120–124. <https://doi.org/10.1080/00021369.1964.10858215>
- Navarro, M. C., De Lima, A. R., Askue, J., & Contreras, V. T. (2003). Morphological comparison of axenic amastigogenesis of trypomastigotes and metacyclic forms of *Trypanosoma cruzi*. *Memórias Do Instituto Oswaldo Cruz*, 98(1), 83–91. <https://doi.org/10.1590/S0074-02762003000100012>
- Nickerson, J. L., Baghalabadi, V., Rajendran, S. R. C. K., Jakubec, P. J., Said, H., McMillen, T. S., ... Doucette, A. A. (2021). Recent advances in top-down

- proteome sample processing ahead of MS analysis. *Mass Spectrometry Reviews*, mas.21706. <https://doi.org/10.1002/mas.21706>
- Niebisch, A., Kabus, A., Schultz, C., Weil, B., & Bott, M. (2006). Corynebacterial Protein Kinase G Controls 2-Oxoglutarate Dehydrogenase Activity via the Phosphorylation Status of the Odhl Protein. *Journal of Biological Chemistry*, 281(18), 12300–12307. <https://doi.org/10.1074/jbc.M512515200>
- Ntai, I., LeDuc, R. D., Fellers, R. T., Erdmann-Gilmore, P., Davies, S. R., Rumsey, J., ... Kelleher, N. L. (2016). Integrated Bottom-Up and Top-Down Proteomics of Patient-Derived Breast Tumor Xenografts. *Molecular & Cellular Proteomics*, 15(1), 45–56. <https://doi.org/10.1074/mcp.M114.047480>
- Okino, S., Inui, M., & Yukawa, H. (2005). Production of organic acids by *Corynebacterium glutamicum* under oxygen deprivation. *Applied Microbiology and Biotechnology*, 68(4), 475–480. <https://doi.org/10.1007/s00253-005-1900-y>
- Pacakova, L., Harant, K., Volf, P., & Lestinova, T. (2022). Three types of *Leishmania mexicana* amastigotes: Proteome comparison by quantitative proteomic analysis. *Frontiers in Cellular and Infection Microbiology*, 12(November), 1–18. <https://doi.org/10.3389/fcimb.2022.1022448>
- Pandey, A., & Mann, M. (2000, June). Proteomics to study genes and genomes. *Nature*, Vol. 405, pp. 837–846. <https://doi.org/10.1038/35015709>
- Park, S. H., Kim, H. U., Kim, T. Y., Park, J. S., Kim, S.-S., & Lee, S. Y. (2014). Metabolic engineering of *Corynebacterium glutamicum* for L-arginine production. *Nature Communications*, 5(1), 4618. <https://doi.org/10.1038/ncomms5618>
- Pérez-Molina, J. A., & Molina, I. (2018). Chagas disease. *The Lancet*, 391(10115), 82–94. [https://doi.org/10.1016/S0140-6736\(17\)31612-4](https://doi.org/10.1016/S0140-6736(17)31612-4)
- Petras, D., Heiss, P., Süßmuth, R. D., & Calvete, J. J. (2015). Venom proteomics of Indonesian king cobra, *Ophiophagus hannah*: Integrating top-down and bottom-up approaches. *Journal of Proteome Research*, 14(6), 2539–2556. <https://doi.org/10.1021/acs.jproteome.5b00305>
- Piatkov, K., Vu, T., Hwang, C.-S., & Varshavsky, A. (2015). Formyl-methionine as a degradation signal at the N-termini of bacterial proteins. *Microbial Cell*, 2(10), 376–393. <https://doi.org/10.15698/mic2015.10.231>
- Pue, N., & Guddat, L. (2014). Acetohydroxyacid Synthase: A Target for Antimicrobial Drug Discovery. *Current Pharmaceutical Design*, 20(5), 740–753. <https://doi.org/10.2174/13816128113199990009>
- Queiroz, R. M. L., Charneau, S., Mandacaru, S. C., Schwämmle, V., Lima, B. D., Roepstorff, P., & Ricart, C. A. O. (2014). Quantitative proteomic and phosphoproteomic analysis of *Trypanosoma cruzi* amastigogenesis. *Molecular and Cellular Proteomics*, 13(12), 3457–3472.

<https://doi.org/10.1074/mcp.M114.040329>

- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Retrieved from <https://www.r-project.org/>.
- Ranjan, A., Mercier, E., Bhatt, A., & Wintermeyer, W. (2017). Signal recognition particle prevents N-terminal processing of bacterial membrane proteins. *Nature Communications*, 8(1), 15562. <https://doi.org/10.1038/ncomms15562>
- Rappsilber, J., Mann, M., & Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols*, 2(8), 1896–1906. <https://doi.org/10.1038/nprot.2007.261>
- Rassi, A., Rassi, A., & Marcondes de Rezende, J. (2012). American Trypanosomiasis (Chagas Disease). *Infectious Disease Clinics of North America*, 26(2), 275–291. <https://doi.org/10.1016/j.idc.2012.03.002>
- Ray, D., Anand, U., Jha, N. K., Korzeniewska, E., Bontempi, E., Proćków, J., & Dey, A. (2022). The soil bacterium, *Corynebacterium glutamicum*, from biosynthesis of value-added products to bioremediation: A master of many trades. *Environmental Research*, 213(May), 113622. <https://doi.org/10.1016/j.envres.2022.113622>
- Ree, R., Varland, S., & Arnesen, T. (2018). Spotlight on protein N-terminal acetylation. *Experimental & Molecular Medicine*, 50(7), 1–13. <https://doi.org/10.1038/s12276-018-0116-z>
- RStudio Team. (2021). *RStudio: integrated development for R*. Bonton, MA.
- Ruprecht, B., Zecha, J., Zolg, D. P., & Kuster, B. (2017). High pH Reversed-Phase Micro-Columns for Simple, Sensitive, and Efficient Fractionation of Proteome and (TMT labeled) Phosphoproteome Digests. In L. Comai, J. E. Katz, & P. Mallick (Eds.), *Test* (pp. 83–98). https://doi.org/10.1007/978-1-4939-6747-6_8
- Santos Júnior, A. de C. M. dos, Melo, R. M., Ferreira, B. V. G., Pontes, A. H., Lima, C. M. R. de, Fontes, W., ... Ricart, C. A. O. (2021). Quantitative proteomics and phosphoproteomics of *Trypanosoma cruzi* epimastigote cell cycle. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1869(5), 140619. <https://doi.org/10.1016/j.bbapap.2021.140619>
- Savaryn, J. P., Toby, T. K., & Kelleher, N. L. (2016). A researcher's guide to mass spectrometry-based proteomics. *PROTEOMICS*, 16(18), 2435–2443. <https://doi.org/10.1002/pmic.201600113>
- Schaffer, L. V., Millikin, R. J., Miller, R. M., Anderson, L. C., Fellers, R. T., Ge, Y., ... Smith, L. M. (2019). Identification and Quantification of Proteoforms by Mass Spectrometry. *Proteomics*, 19(10), 1–15. <https://doi.org/10.1002/pmic.201800361>
- Schaffer, L. V., Millikin, R. J., Shortreed, M. R., Scalf, M., & Smith, L. M. (2020). Improving Proteoform Identifications in Complex Systems through Integration of

- Bottom-Up and Top-Down Data. *Journal of Proteome Research*, 19(8), 3510–3517. <https://doi.org/10.1021/acs.jproteome.0c00332>
- Schaffer, L. V., Rensvold, J. W., Shortreed, M. R., Cesnik, A. J., Jochem, A., Scalf, M., ... Smith, L. M. (2018). Identification and Quantification of Murine Mitochondrial Proteoforms Using an Integrated Top-Down and Intact-Mass Strategy. *Journal of Proteome Research*, 17(10), 3526–3536. <https://doi.org/10.1021/acs.jproteome.8b00469>
- Schubert, O. T., Röst, H. L., Collins, B. C., Rosenberger, G., & Aebersold, R. (2017). Quantitative proteomics: Challenges and opportunities in basic and applied research. *Nature Protocols*, 12(7), 1289–1294. <https://doi.org/10.1038/nprot.2017.040>
- Schwämmle, V., & Jensen, O. N. (2018). VSClust: Feature-based variance-sensitive clustering of omics data. *Bioinformatics*, 34(17), 2965–2972. <https://doi.org/10.1093/bioinformatics/bty224>
- Sheng, Q., Wu, X.-Y., Xu, X., Tan, X., Li, Z., & Zhang, B. (2021). Production of l-glutamate family amino acids in *Corynebacterium glutamicum*: Physiological mechanism, genetic modulation, and prospects. *Synthetic and Systems Biotechnology*, 6(4), 302–325. <https://doi.org/10.1016/j.synbio.2021.09.005>
- Shimizu, H., Tanaka, H., Nakato, A., Nagahisa, K., Kimura, E., & Shioya, S. (2003). Effects of the changes in enzyme activities on metabolic flux redistribution around the 2-oxoglutarate branch in glutamate production by *Corynebacterium glutamicum*. *Bioprocess and Biosystems Engineering*, 25(5), 291–298. <https://doi.org/10.1007/s00449-002-0307-8>
- Shirai, T., Fujimura, K., Furusawa, C., Nagahisa, K., Shioya, S., & Shimizu, H. (2007). Study on roles of anaplerotic pathways in glutamate overproduction of *Corynebacterium glutamicum* by metabolic flux analysis. *Microbial Cell Factories*, 6(1), 19. <https://doi.org/10.1186/1475-2859-6-19>
- Shuken, S. R. (2023). An Introduction to Mass Spectrometry-Based Proteomics. *Journal of Proteome Research*, 22(7), 2151–2171. <https://doi.org/10.1021/acs.jproteome.2c00838>
- Sievers, F., & Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*, 27(1), 135–145. <https://doi.org/10.1002/pro.3290>
- Sinha, A., & Mann, M. (2020). A beginner's guide to mass spectrometry-based proteomics. *The Biochemist*, 42(5), 64–69. <https://doi.org/10.1042/BIO20200057>
- Smith, L. M., & Kelleher, N. L. (2013). Proteoform: A single term describing protein complexity. *Nature Methods*, 10(3), 186–187. <https://doi.org/10.1038/nmeth.2369>
- Smith, R. D., Anderson, G. A., Lipton, M. S., Pasa-Tolic, L., Shen, Y., Conrads, T.

- P., ... Udseth, H. R. (2002). An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics*, 2(5), 513–523. [https://doi.org/10.1002/1615-9861\(200205\)2:5<513::AID-PROT513>3.0.CO;2-W](https://doi.org/10.1002/1615-9861(200205)2:5<513::AID-PROT513>3.0.CO;2-W)
- Sogues, A., Martinez, M., Gaday, Q., Ben Assaya, M., Graña, M., Voegelé, A., ... Alzari, P. M. (2020). Essential dynamic interdependence of FtsZ and SepF for Z-ring and septum formation in *Corynebacterium glutamicum*. *Nature Communications*, 11(1), 1641. <https://doi.org/10.1038/s41467-020-15490-8>
- Solntsev, S. K., Shortreed, M. R., Frey, B. L., & Smith, L. M. (2018). Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *Journal of Proteome Research*, 17(5), 1844–1851. <https://doi.org/10.1021/acs.jproteome.7b00873>
- Souza, F. S. P. de, Rampazzo, R. de C. P., Manhaes, L., Soares, M. J., Cavalcanti, D. P., Krieger, M. A., ... Fragoso, S. P. (2010). Knockout of the gene encoding the kinetoplast-associated protein 3 (KAP3) in *Trypanosoma cruzi*: Effect on kinetoplast organization, cell proliferation and differentiation. *Molecular and Biochemical Parasitology*, 172(2), 90–98. <https://doi.org/10.1016/j.molbiopara.2010.03.014>
- Souza, W. (2005). Basic Cell Biology of *Trypanosoma cruzi*. *Current Pharmaceutical Design*, 8(4), 269–285. <https://doi.org/10.2174/1381612023396276>
- Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M., & Henrissat, B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of -amylase-related proteins. *Protein Engineering Design and Selection*, 19(12), 555–562. <https://doi.org/10.1093/protein/gzl044>
- Stratton, K. G., Webb-Robertson, B. J. M., McCue, L. A., Stanfill, B., Claborne, D., Godinez, I., ... Bramer, L. M. (2019). PmartR: Quality Control and Statistics for Mass Spectrometry-Based Biological Data. *Journal of Proteome Research*, 18(3), 1418–1425. <https://doi.org/10.1021/acs.jproteome.8b00760>
- Sundermeyer, L., Bosco, G., Gujar, S., Brocker, M., Baumgart, M., Willbold, D., ... Bott, M. (2022). Characteristics of the GlnH and GlnX Signal Transduction Proteins Controlling PknG-Mediated Phosphorylation of OdhI and 2-Oxoglutarate Dehydrogenase Activity in *Corynebacterium glutamicum*. *Microbiology Spectrum*. <https://doi.org/10.1128/spectrum.02677-22>
- Tabb, D. L., Jeong, K., Druart, K., Gant, M. S., Brown, K. A., Nicora, C., ... Chamot-Rooke, J. (2023). Comparing Top-Down Proteoform Identification: Deconvolution, PrSM Overlap, and PTM Detection. *Journal of Proteome Research*. <https://doi.org/10.1021/acs.jproteome.2c00673>
- Takemori, A., Butcher, D. S., Harman, V. M., Brownridge, P., Shima, K., Higo, D., ... Takemori, N. (2020). PEPPI-MS: Polyacrylamide-Gel-Based Prefractionation for Analysis of Intact Proteoforms and Protein Complexes by Mass

- Spectrometry. *Journal of Proteome Research*, 19(9), 3779–3791. <https://doi.org/10.1021/acs.jproteome.0c00303>
- Tauch, A., & Burkovski, A. (2015). Molecular armory or niche factors: Virulence determinants of corynebacterium species. *FEMS Microbiology Letters*, 362(23), 1–6. <https://doi.org/10.1093/femsle/fnv185>
- Tavernelli, L. E., Motta, M. C. M., Gonçalves, C. S., da Silva, M. S., Elias, M. C., Alonso, V. L., ... Cribb, P. (2019). Overexpression of Trypanosoma cruzi High Mobility Group B protein (TcHMGB) alters the nuclear structure, impairs cytokinesis and reduces the parasite infectivity. *Scientific Reports*, 9(1), 192. <https://doi.org/10.1038/s41598-018-36718-0>
- Tomlinson, S., Vandekerckhove, F., Frevert, U., & Nussenzweig, V. (1995). The induction of Trypanosoma cruzi trypomastigote to amastigote transformation by low pH. *Parasitology*, 110(5), 547–554. <https://doi.org/10.1017/S0031182000065264>
- Tran, J. C., & Doucette, A. A. (2008). Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: A proteome prefractionation tool. *Journal of Proteome Research*, 7(4), 1761–1766. <https://doi.org/10.1021/pr700677u>
- Tran, J. C., & Doucette, A. A. (2009). Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Analytical Chemistry*, 81(15), 6201–6209. <https://doi.org/10.1021/ac900729r>
- Tryfona, T., & Bustard, M. T. (2005). Fermentative production of lysine by Corynebacterium glutamicum: Transmembrane transport and metabolic flux analysis. *Process Biochemistry*, 40(2), 499–508. <https://doi.org/10.1016/j.procbio.2004.01.037>
- Vozáriková, V., Kunová, N., Bauer, J. A., Frankovský, J., Kotrasová, V., Procházková, K., ... Tomáška, L. (2020). Mitochondrial HMG-Box Containing Proteins: From Biochemical Properties to the Roles in Human Diseases. *Biomolecules*, 10(8), 1193. <https://doi.org/10.3390/biom10081193>
- Wang, J., Pappas-Brown, V., Englund, P. T., & Jensen, R. E. (2014). TbKAP6, a Mitochondrial HMG Box-Containing Protein in Trypanosoma brucei, Is the First Trypanosomatid Kinetoplast-Associated Protein Essential for Kinetoplast DNA Replication and Maintenance. *Eukaryotic Cell*, 13(7), 919–932. <https://doi.org/10.1128/EC.00260-13>
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Wessel, D., & Flügge, U. I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical Biochemistry*, 138(1), 141–143. [https://doi.org/10.1016/0003-2697\(84\)90782-6](https://doi.org/10.1016/0003-2697(84)90782-6)

- WHO. (2022). Fact sheet: Chagas disease (also known as American trypanosomiasis). *Chagas Disease Fact Sheets*. Retrieved from [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis))
- Wiśniewski, J. R. (2017). Filter-Aided Sample Preparation: The Versatile and Efficient Method for Proteomic Analysis. In *Methods in Enzymology* (Vol. 585, pp. 15–27). <https://doi.org/10.1016/bs.mie.2016.09.013>
- Wiśniewski, Jacek R. (2018). *Filter-Aided Sample Preparation for Proteome Analysis* (D. Becher, Ed.). In (pp. 3–10). <https://doi.org/10.1007/978-1-4939-8695-8>
- Wiśniewski, Jacek R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, 6(5), 359–362. <https://doi.org/10.1038/nmeth.1322>
- Xing, G., Zhang, J., Chen, Y., & Zhao, Y. (2008). Identification of four novel types of in vitro protein modifications. *Journal of Proteome Research*, 7(10), 4603–4608. <https://doi.org/10.1021/pr800456q>
- Xu, C., & Ray, D. S. (1993). Isolation of proteins associated with kinetoplast DNA networks in vivo. *Proceedings of the National Academy of Sciences*, 90(5), 1786–1789. <https://doi.org/10.1073/pnas.90.5.1786>
- Xu, M., Tang, M., Chen, J., Yang, T., Zhang, X., Shao, M., ... Rao, Z. (2020). PII Signal Transduction Protein GlnK Alleviates Feedback Inhibition of N -Acetyl- I -Glutamate Kinase by I -Arginine in *Corynebacterium glutamicum*. *Applied and Environmental Microbiology*, 86(8), 1–19. <https://doi.org/10.1128/AEM.00039-20>
- Yamamoto, S., Suda, M., Niimi, S., Inui, M., & Yukawa, H. (2013). Strain optimization for efficient isobutanol production using *corynebacterium glutamicum* under oxygen deprivation. *Biotechnology and Bioengineering*, 110(11), 2938–2948. <https://doi.org/10.1002/bit.24961>
- Yang, C.-I., Kim, J., & Shan, S. (2022). Ribosome-nascent Chain Interaction Regulates N-terminal Protein Modification. *Journal of Molecular Biology*, 434(9), 167535. <https://doi.org/10.1016/j.jmb.2022.167535>
- Yao, C. (2010). Major Surface Protease of Trypanosomatids: One Size Fits All? *Infection and Immunity*, 78(1), 22–31. <https://doi.org/10.1128/IAI.00776-09>
- Zheng, S., & Doucette, A. A. (2016). Preventing N - and O -formylation of proteins when incubated in concentrated formic acid. *PROTEOMICS*, 16(7), 1059–1068. <https://doi.org/10.1002/pmic.201500366>
- Zhong, J., Sun, Y., Xie, M., Peng, W., Zhang, C., Wu, F.-X., & Wang, J. (2021). Proteoform characterization based on top-down mass spectrometry. *Briefings in Bioinformatics*, 22(2), 1729–1750. <https://doi.org/10.1093/bib/bbaa015>
- Zhou, Q., Hu, H., & Li, Z. (2014). New Insights into the Molecular Mechanisms of Mitosis and Cytokinesis in Trypanosomes. In *International Review of Cell and*

Molecular Biology (1st ed., Vol. 308, pp. 127–166).
<https://doi.org/10.1016/B978-0-12-800097-7.00004-X>

Zimmer, J. S. D., Monroe, M. E., Qian, W. J., & Smith, R. D. (2006). Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrometry Reviews*, 25(3), 450–482.
<https://doi.org/10.1002/mas.20071>

Zuma, A. A., dos Santos Barrias, E., & de Souza, W. (2020). Basic Biology of *Trypanosoma cruzi*. *Current Pharmaceutical Design*, 27(14), 1671–1732.
<https://doi.org/10.2174/1381612826999201203213527>

Anexos

5.2.5 Anexo 1: Artigo “Revealing *Corynebacterium glutamicum* proteoforms through top-down proteomics”.



OPEN Revealing *Corynebacterium glutamicum* proteoforms through top-down proteomics

Reynaldo Magalhães Melo¹, Jaques Miranda Ferreira de Souza¹, Thomas Christopher Rhys Williams², Wagner Fontes¹, Marcelo Valle de Sousa¹, Carlos André Ornelas Ricart¹ & Luis Henrique Ferreira do Vale¹✉

Corynebacterium glutamicum is a bacterium widely employed in the industrial production of amino acids as well as a broad range of other biotechnological products. The present study describes the characterization of *C. glutamicum* proteoforms, and their post-translational modifications (PTMs) employing top-down proteomics. Despite previous evidence of PTMs having roles in the regulation of *C. glutamicum* metabolism, this is the first top-down proteome analysis of this organism. We identified 1125 proteoforms from 273 proteins, with 60% of proteins presenting at least one mass shift, suggesting the presence of PTMs, including several acetylated, oxidized and formylated proteoforms. Furthermore, proteins relevant to amino acid production, protein secretion, and oxidative stress were identified with mass shifts suggesting the presence of uncharacterized PTMs and proteoforms that may affect biotechnologically relevant processes in this industrial workhorse. For instance, the membrane proteins mepB and SecG were identified as a cleaved and a formylated proteoform, respectively. While in the central metabolism, OdhI was identified as two proteoforms with potential biological relevance: a cleaved proteoform and a proteoform with PTMs corresponding to a 70 Da mass shift.

The bacterium *Corynebacterium glutamicum* is an industrial workhorse capable of producing a broad range of biomolecules from a large pool of substrates¹ and is currently the best option for production of L-alpha amino acids². Amino acid production represents a multi-billion dollar market³, with an annual production of approximately 10 million tons worldwide², with L-glutamate and L-lysine corresponding to 3.3 million tons/year and 2.2 million tons/year, respectively². Besides amino acid production, *C. glutamicum* is also utilized in other biotechnological processes, such as the production of heterologous proteins⁴, organic acids⁵, carotenoids, isobutanol⁶, polymers⁷, and more recently, its potential in bioremediation processes has been investigated⁸.

Recent developments in proteomics have demonstrated the importance of post-translational modifications (PTMs) in several species of bacteria⁹. In *C. glutamicum*, the importance of PTMs was investigated previously revealing the crucial effect of oxoglutarate dehydrogenase inhibitor (OdhI) phosphorylation in the production of L-glutamate^{10,11}. More recently, mass spectrometry proteomics of *C. glutamicum* showed the influence of L-glutamate-producing conditions on the succinylation and acetylation of *C. glutamicum* metabolic proteins, including OdhI¹².

There are currently two main proteomics approaches that are referred to as bottom-up and top-down. Briefly, in the bottom-up approach proteins are cleaved by a protease (usually trypsin) and the subsequent peptides are analyzed by mass spectrometry, whilst in the top-down procedure, proteins are examined in their intact forms¹³. As a consequence of protease digestion, the information from several protein regions can be lost in bottom-up approaches, including PTMs. Furthermore, in this approach protein identity is inferred based on peptide identification, and this may cause ambiguity in cases in which peptides may belong to more than one protein¹³. In contrast, the top-down approach allows the identification of the full sequence of proteins. Moreover, analysis of intact forms of proteins permits the identification of proteoforms, defined as different forms of proteins derived from the same gene¹⁴. Such proteoforms can be generated by amino acid residue substitution, proteolytic cleavage, alternative splicing, and several types of PTMs¹⁵. Furthermore, different proteoforms of the same protein can have different function and affect biological processes of an organism¹⁶. The power of

¹Laboratory of Protein Chemistry and Biochemistry, Department of Cell Biology, Institute of Biology, University of Brasilia, Brasilia, Brazil. ²Laboratory of Plant Biochemistry, Department of Botany, Institute of Biology, University of Brasilia, Brasilia, Brazil. ✉email: luisvale@unb.br

top-down proteomics has been applied to other bacteria¹⁷. For instance, top-down proteomics of *Escherichia coli* and the identification of proteoforms has been performed¹⁸. Despite the evidence of the relevance of PTMs in *C. glutamicum* metabolism, no top-down proteomic studies have been published to date. Here we carried out the first top-down proteomics analysis of *C. glutamicum*, using a precursor tolerant approach to identify and reveal PTMs and proteoforms.

Results and discussion

As expected, a wide range of proteoform molecular masses could be seen in the pre-fractionated intracellular proteins, but GELFrEE was efficient in separating large proteoforms from smaller forms (Fig. 1A). Fractions 0, 1, 2, 3, 4, 5, 7, 9 and 11 displayed proteoforms below 50 kDa, and for this reason were chosen for subsequent LC-MS/MS analysis. Despite the clear presence of proteins in the 30–50 kDa range, as shown by SDS-PAGE (Fig. 1A), LC-MS analysis detected proteoforms predominantly below 30 kDa (supplementary Fig. S1). The reduced number of observed proteoforms larger than 30 kDa may be due to known challenges for the identification of denatured large proteoforms such as the signal-to-noise ratio reduction occurring with increases in proteoform molecular weight¹⁹.

Global post-translational modifications profile. The top-down proteomic analysis of *C. glutamicum* generated 5127 proteoform spectrum matches (PrSMs), providing the identification of 1125 different proteoforms related to 273 different proteins (supplementary Table 1). Approximately 65% (177 proteins) of the total protein number and 47% of PrSMs (2423 PrSMs) were identified with mass shifts (Δm), mass differences between the expected precursor mass and observed precursor mass, which suggests the presence of PTMs. Putative PTMs were annotated in supplementary Table 1 using the entire Unimod database, however, this information should be used with caution, since spectra were not manually verified and the software considered rounded mass shifts for comparisons, generating a tolerance of approximately 0.5 Da. There are software available for spectra visualization and further evaluation of PTMs, which includes ProSight Lite²⁰ and TopMSV²¹. Analysis of all PrSMs with Δm revealed a broad diversity of Δm , suggesting the presence of different PTMs (Fig. 1B).

Some of the most frequent Δm identified here were also reported as highly present in *Escherichia coli* through bottom-up proteomics using an unbiased identification strategy²². In *C. glutamicum*, the mass shifts most frequently identified were 16 Da (putative oxidation, 10.15% of modified PrSMs), -18 Da (putative dehydration, 2.97% of modified PrSMs), and 32 Da (possible double oxidation, 1.32% of modified PrSMs). Interestingly, the proportion of *E. coli* identified peptide spectrum matches (PSMs) with -18 Da (3.8%) and 32 Da (0.8%) Δm were very similar to those found in this study. In contrast, the proportion of PSMs identified with 16 Da (24%) Δm in *E. coli* was greater than in *C. glutamicum*. However, it is worth mentioning that the high number of PrSMs identified with 15 Da Δm (6.11% of modified PrSMs) may be caused by miss identification of the 16 Da Δm , as will be further discussed in the next topics.

Proteins identified with Δm were analyzed using the String Cytoscape app, after clustering proteins according to their interaction nodes, each cluster was submitted to overrepresentation analysis and the most representative

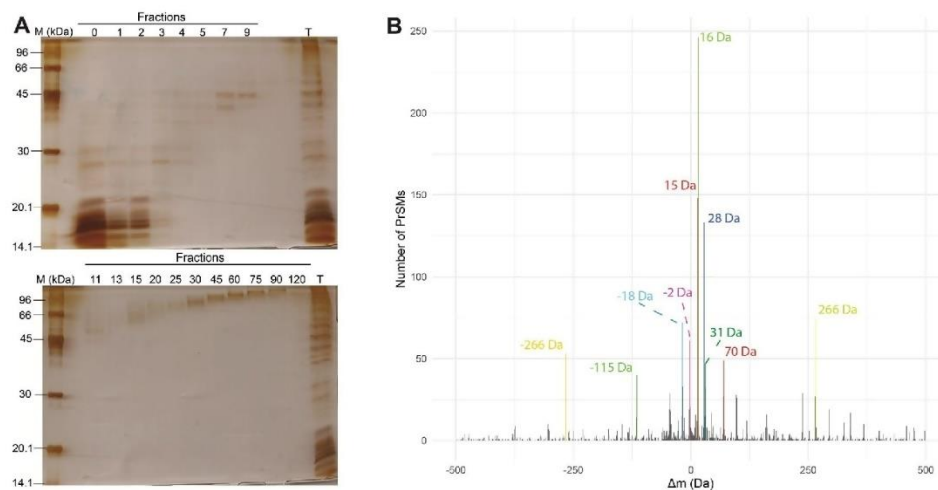


Figure 1. SDS-PAGE of GELFrEE fractions and global profile of identified mass shifts (Δm). (A) SDS-PAGE of fractions resulting from GELFrEE (represented by the numbers) and pre-fractionated sample (T). (B) Difference of observed and theoretical precursor masses (mass shifts, Δm) PrSMs count histogram, where the ten most frequent Δm are highlighted in different colors. The bin was set to 1 and x limits to -500 Da and 500 Da.

terms were related to different gene ontology terms such as protein-containing complexes, electron transfer activity and CYTH domain (CyaB, thiamine triphosphatase), pyrimidine metabolism, transmembrane helix, biosynthesis of secondary metabolites and thioredoxin domain. Moreover, eight ribosomal PrSMs presented two Δm , indicating the presence of at least two concomitantly occurring PTMs (supplementary Fig. S2).

Beyond functional annotation, we analyzed the number of N-terminal acetylation and some frequently identified Δm : 16 Da, 15 Da, 28 Da, 266 Da, -18 Da, and 32 Da (Fig. 2A). These Δm suggest the presence of oxidation (UnimodAC: 35, $\Delta m = 15.994915$ Da), deamidation followed by a methylation (UnimodAC: 528, $\Delta m = 14.999666$ Da), formylation (UnimodAC: 122, $\Delta m = 27.994915$ Da), sodium dodecyl sulfate (SDS) adducts²³, dehydration (UnimodAC: 23, $\Delta m = -18.010565$ Da), and persulfide (UnimodAC: 421, $\Delta m = 31.972071$ Da). The 28 Da mass shift may also correspond to di-methylation (Unimod: 36, $\Delta m = 28.031300$ Da). Despite being difficult to differentiate between formylation and di-methylation, the higher number of Δm close to 27.99 Da and 27.98 Da (supplementary Table 2) leans the inference towards formylation events. The amino acid residue localization of the 28 Da mass shift was not possible without ambiguity, as can be seen by the absence of this mass shift in Fig. 2B. A possible reason for this is that the suggested formylation events were identified mainly at the N-terminal of proteins, and its assignment was always between a few N-terminal residues of the protein, as in the case of the protein MslL (supplementary Fig. 7). When the assignment of the mass shift was in more than one amino acid residue, the information was not plotted to avoid ambiguity. The suggested oxidation events related to the 16 Da Δm is further supported by a large number of methionine residues modified by this mass shift (Fig. 2B). On the other hand, the 15 Da Δm seems to be sometimes misinterpreted, since the most frequently identified residue with this modification was also methionine. However, some glutamic acid residues were identified with this Δm , suggesting that deamidation followed by methylation can still be present in some cases, but are easily confused with oxidations (Fig. 2B). The 32 Da mass shift identified in some proteins could be related to persulfide PTM, however, this modification occurs in cysteine and aspartic acid, and none of these could be observed in the unambiguously identified residues (Fig. 2B). Other possibility would be two oxidations, supported by the methionine residues identified with this mass shift. On the other hand, the possibility of persulfide modification should not be excluded, mainly in proteins in which the modified amino acid residue could not be unambiguously identified. Protein oxidation may have great relevance in biological processes, however, it is important to note that it may also be an artefact caused by sample handling²⁴. Moreover, several Δm were approximately 266 Da, suggesting the presence of adducts resulting from sodium dodecyl sulfate (SDS). The presence of SDS artifacts was also reported in an intact proteomic analysis of *E. coli*, and suggested to be caused by incomplete removal of the detergent²⁵. Regarding the -18 Da Δm , it may result from dehydration events and was predominantly identified in serine residues (Fig. 2B). These predominant mass shifts may also be related to other putative PTMs, such as tyrosine oxidation to 2-aminotyrosine (UnimodAC: 342, $\Delta m = 15.010899$ Da) for the mass shifts of 15 Da. These possibilities can be observed separately for each identified PrSM in supplementary Table 1. N-terminal acetylation was the most frequently identified PTM (Fig. 2A); on the other hand, the 42 Da mass shift, which suggests lysine acetylation (UnimodAC: 1, $\Delta m = 42.010565$ Da) was not frequently detected (supplementary Table 2). In identifying proteoforms, the N-terminal acetylation was used as a variable modification. Therefore, proteoforms with lysine acetylation near the N-terminal without fragments to explain the 42 mass shift to a lysine residue could be misidentified with a N-terminal acetylation.

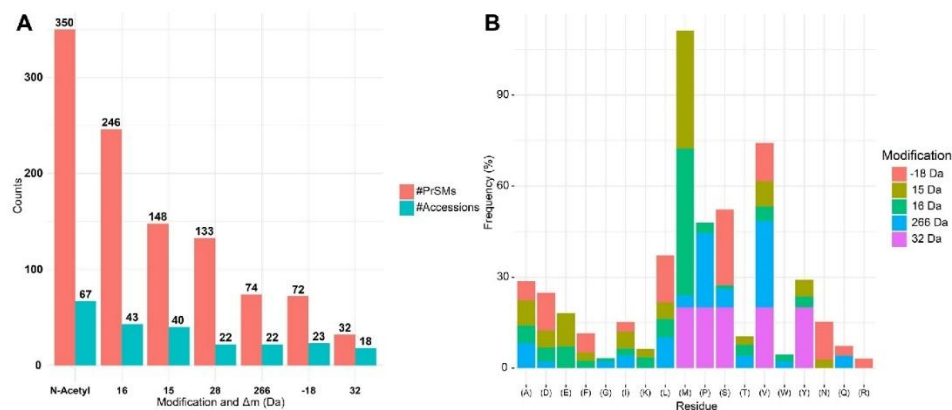


Figure 2. Number of identified PrSMs and protein accessions of the most frequent modifications and ratio of unambiguously identified residues of each mass shift. **(A)** Number of proteoform spectrum matches (PrSMs) and proteins identified with most frequent mass shifts and N-terminal acetylation. **(B)** Ratio of the number PrSMs of an amino acid residue modified by each mass shifts (Δm) and the total PrSMs identified with the same Δm . For example, the number of methionine identified with mass shift of 16 Da (41 PrSMs)/number of all residues identified with 16 Da Δm (85 PrSMs) = 48% (represented by the light green bar above the (M) in the x axis).

Overrepresentation analysis of proteins belonging to the six most frequent modifications allowed the identification of terms related to ribosomal proteins, predominant in proteins with Δm of -18 Da, 15 Da, 16 Da, and N-terminal acetylated proteins (Fig. 3). Prokaryotic ribosomal proteins are described to be commonly methylated and acetylated²⁶. On the other hand, Δm of 28 Da and 266 Da were overrepresented for terms related to membrane proteins. In addition, proteins identified with 28 Da mass shift showed overrepresentation for the secretion system and protein export (Fig. 3).

We identified 13 proteins that presented more than 15 proteoforms and most of them were ribosomal (supplementary Fig. S3). Moreover, when considering proteins with more than 9 proteoforms, the cellular component annotation also revealed membrane proteins (supplementary Fig. S3). In spite of the great number of proteoforms, it is worth mentioning that some of them may be boosted by SDS adducts. For example, 50S ribosomal protein L7/L12 (Q8NT28) proteoforms were identified with Δm of 266 Da, 532 Da, and 401 Da, probably representing the addition of one SDS adduct, two SDS adducts, and two SDS adducts plus the loss of the initiator methionine residue. The presence of SDS adducts in proteins is not desirable, however, the identification of these proteins is important to avoid misidentifications and to assign proteins that were only identifiable through adducted forms²⁵.

In order to contribute with more accurate data regarding *C. glutamicum* biotechnological and metabolic processes, some PrSMs related or potentially involved with these functions were manually characterized through inspection of MS1 and MS2 Spectra (Table 1).

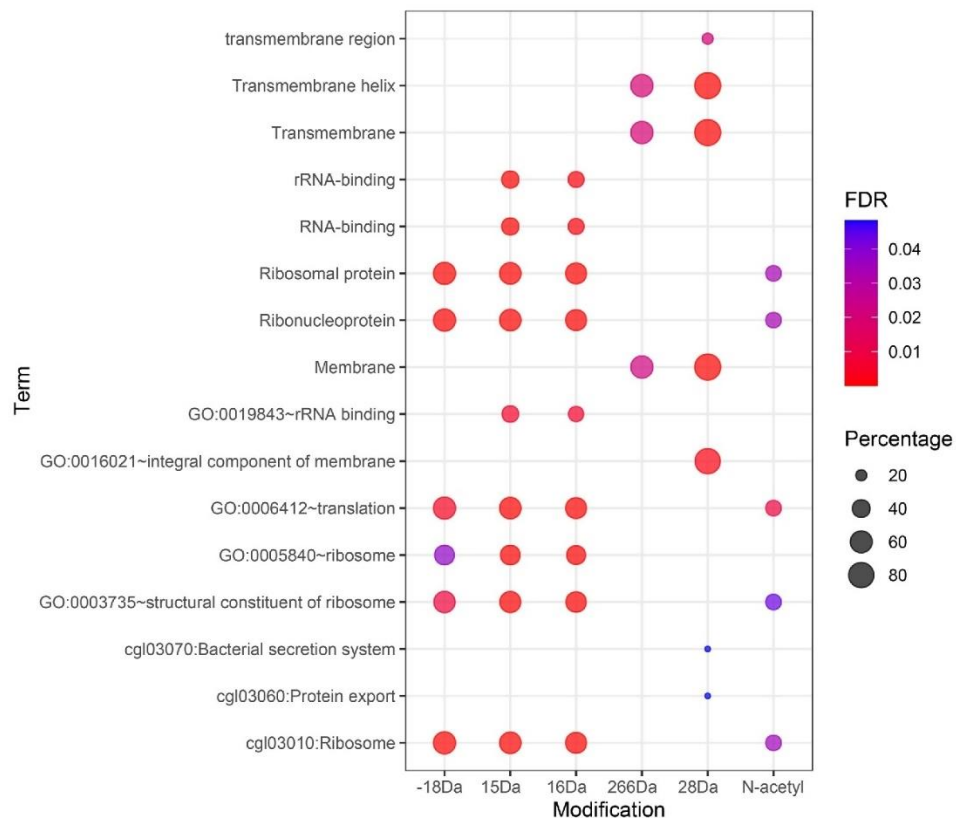


Figure 3. Overrepresentation analysis of proteins according to identified modification and mass shifts. The proteins accession codes of each modification or mass shift were separately submitted to overrepresentation analysis on DAVID against all encoded proteins of the *C. glutamicum* genome. The resulting FDR, representing the score of enrichment, and percentage, representing the proportion of identified proteins to all proteins present in the related *gene-set*, were plotted using ggplot2 bubble plot representation.

UniprotAC	Protein name	Biological Process	Proteoforms	Putative PTM
Q8Z469	SecG	Protein export	28 Da	N-formylation
Q8NS24	mepB	Peptidase/cell wall metabolism	N-terminal cleavage	N-terminal cleavage
Q8NQJ3	OdhI	Glutamate production	70 Da/N-terminal cleavage	Crotonaldehyde lipid peroxidation or butyrylation
Q8NL68	HMADP	Glutamate production/stress response	30 Da/-2 Da	Methylation + oxidation/disulfide bond
Q8NMS6	Peroxioredoxin	Stress response	154 Da/186 Da	ONE lipid peroxidation/ONE lipid peroxidation + oxidation
Q8NLG6	Thioredoxin	Stress response	30 Da/-2 Da	Methylation + oxidation/disulfide bond
Q8NS07	MscL	Metabolite efflux	28 Da	N-formylation

Table 1. Proteoforms with potential metabolism regulation function and of biotechnological interest. Proteoforms that presented a potential involvement in the regulation of metabolism or important biotechnological processes are represented by their corresponding protein names, Uniprot accession code (AC) and identified mass shift or sequence cleavage. Putative post-translational modifications (PTM) were defined based on identified mass shifts using the unimod database (<https://www.unimod.org>).

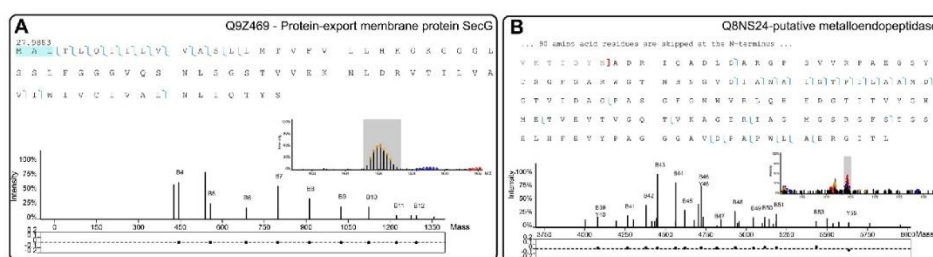


Figure 4. PrSMs of Protein-export membrane protein SecG (Q8Z469) and putative metalloendopeptidase mepB (Q8NS24). In the top region of both images (A and B) can be seen the sequence of the identified protein, where possible sites of the Δm are represented by residues highlighted in blue background and matched fragments are represented by blue lines between amino acid residues. In the bottom region, the MS1 and isolation window of the identified precursor are depicted (right side of each figure) as well as the mass representation of the MS2 spectrum attached to the mass error of matched fragments. (A) Inspection of the 28 Da mass shift identified in SecG, where we could see that 12 fragments (all *b*-ions) support the identification of the 27.9883 mass shift near the N-terminal. The mass error of great part of these fragments were close to 0 ppm, as depicted in the bottom of the image. (B) Inspection of mepB cleaved proteoform (cleavage site represented by the red bracket), which demonstrates a great number of fragments supporting the cleaved proteoform (all *b*-ions) with mass error close to 0 ppm.

Membrane proteins. Secretion system protein (SecG) (Fig. 4A) and large-conductance mechanosensitive channel (MscL) (supplementary Fig. S4) were identified with 28 Da mass shifts. The PrSMs fragments of these proteoforms presented ions supporting the identification of a mass shift of 28 Da, always near the N-terminus, and the mass errors of fragments containing this Δm were extremely low. An example of MS1 and MS2 inspection is demonstrated with SecG (Q8Z469) (Fig. 4A). In this figure, the large number of *b*-ions supporting the 27.9883 mass shift (*b*₄, *b*₅, *b*₆, *b*₇, *b*₈, *b*₉, *b*₁₀, *b*₁₁ and *b*₁₂) represented by the blue lines in the protein sequence is shown, and it can be seen that the mass shift is localized to a very narrowed region of three amino acid residues, due to the several *b*-ions contained in this region.

The 28 Da mass shift near the protein N-terminus of these proteins suggests the presence of N-terminal formylation (Unimod: 122, $\Delta m = 27.994915$ Da) in the identified proteoforms. Recent evidence has suggested the N-terminal formylation of methionine as a signal for protein degradation. This mechanism has been hypothesized as a quality control of protein translation in bacteria²⁷. Both the SecG and MscL formylated PrSMs mentioned above were identified with the presence of N-terminal methionine. This suggests a possible mechanism of degradation of membrane proteins, including some with promising biotechnological applications. For example, the SecG protein is part of the Sec protein export pathway. There is a growing interest in the capacity of *C. glutamicum* to express and secrete heterologous proteins of biotechnological interest²⁸. Furthermore, MscL and MscS are mechanosensitive channels, known for reacting to osmotic stress. Another mechanosensitive channel of *C. glutamicum* (MscCG; P42531) plays a major role in L-glutamate efflux²⁹.

Top-down proteomics is efficient in identifying cleaved proteoforms. For example, mepB, a membrane protein related to the metalloendopeptidase (Q8NS24) was identified by a portion of its sequence with precursor mass of 14.464 kDa. This mass corresponds to the loss of 97 amino acid residues from its N-terminal region (Fig. 4B).

In contrast, a *mepB* cleavage site was supposed to be between Ala43 and Ala44, after a putative signal peptide or transmembrane helix³⁰. Furthermore, according to pfam (<https://pfam.xfam.org/>), *mepB* has a domain belonging to the M23 metallopeptidase family (MEROPS) [130–226]. A well characterized member of MEROPS is the LytM protein from *Staphylococcus aureus*, a metallopeptidase involved in autolysis³¹. It was demonstrated that cleavage in the N-terminal chain causes its peptidase activity to be activated³². Moreover, MEROPS proteins have specificity to peptidoglycan polyglycine regions, some with suggested cell wall metabolism activity³³. In agreement, the *C. glutamicum mepB* gene was described as part of the MtrAB regulon, a two component system implicated in osmoregulation and cell wall metabolism control³⁴. Although it is not clear whether *mepB* is secreted or membrane bound, it has been suggested that its activity would occur extracitoplasmically³⁰. Considering this, *C. glutamicum mepB* may have an important role in cell wall metabolism and/or heterologous protein secretion integrity, and its activity is likely regulated by cleavage of its N-terminal.

Tricarboxylic acid cycle and glutamate metabolism. *C. glutamicum* is widely used in the industrial production of amino acids, especially L-glutamate, which is produced in millions of tons per year³⁵. The tricarboxylic acid (TCA) cycle is an important step in L-glutamate production by *C. glutamicum*. It is well established that the decrease of 2-oxoglutarate dehydrogenase complex (ODHC) activity occurs in conditions that induce L-glutamate production^{36,37}. Furthermore, ODHC activity was found to be regulated by the phosphorylation status of oxoglutarate dehydrogenase inhibitor (OdHI, Q8NQJ3). Thus, the phosphorylated proteoform of OdHI is unable to interact with ODHC, whilst, unphosphorylated OdHI interacts with ODHC inhibiting 2-oxoglutarate conversion to succinyl-CoA¹¹ (Fig. 5A). Moreover, the decreased ODHC activity causes an enhanced production of L-glutamate from 2-oxoglutarate³⁷ (Fig. 5A). Congruently, OdHI phosphorylation status is affected by methods that induce L-glutamate production¹¹.

In this study, seven proteoforms of OdHI were identified. Some of them seemed to be caused by sample preparation or ionization process, such as oxidation ($\Delta m = 16$ Da) and dehydration ($\Delta m = -18$ Da). Conversely, two proteoforms were identified with potential biological relevance: one of them with N-terminal cleavage of three residues and another one with Δm of 70 Da (supplementary Fig. S5). Although these two proteoforms were confidently identified by TopPic, a low number of matching fragments covered the modified regions, lowering our confidence regarding their identifications, principally concerning their location in the sequence. On the other hand, in agreement with one of the identified proteoforms, a bottom-up proteomic analysis in a preliminary study conducted by our group suggested the presence of Δm of 70 Da in the N-terminal peptide of OdHI (data not shown).

Recently, the 70 Da mass shift was identified in the GM-CSF heterologous expressed protein in *Escherichia coli* system. This modification was hypothesized to be a result of crotonaldehyde formed during oxidative stress by lipid peroxidation. The aldehyde reacts with the protein N-terminus or lysine residues, resulting in the Δm of 70 Da³⁸. Another possible PTM for this mass shift is the butyrylation of lysine (UnimodAC: 1289, $\Delta m = 70.041865$ Da). In the present study, the exact site of the modification resulting in the 70 Da mass shift

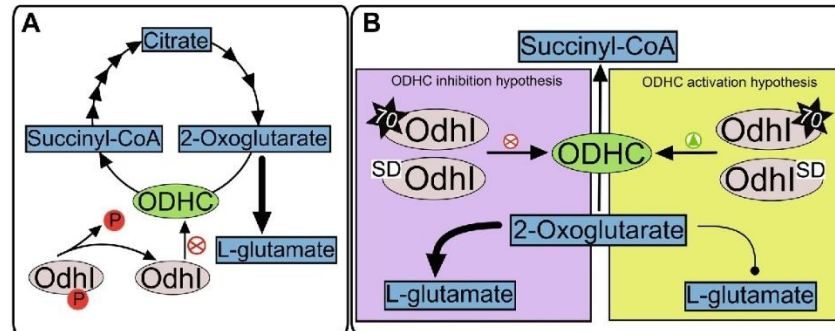


Figure 5. Representation of known and putative OdHI interactions with ODHC and its impact on L-glutamate production. (A) Interaction of OdHI and ODHC hampered by OdHI phosphorylation. The absence of the phosphate group on OdHI results in its interaction and inhibition of ODHC (represented by the red symbol), and flux to L-glutamate production is enhanced (represented by the large arrow from 2-oxoglutarate to L-glutamate)^{11,37}. (B) Depiction of possible metabolic effects and protein interactions affected by OdHI proteoforms described in this study: Δm of 70 Da (represented by the star symbol), and N-terminal dipeptide cleavage (represented by the SD letters on OdHI). The effect of identified proteoforms on OdHI function remains unclear, however potential affected mechanisms are represented through the inhibition of ODHC (purple box; enhancement of L-glutamate production symbolized by the large arrow), or activation of ODHC (green box; decrease of L-glutamate production symbolized by the line with a dot). Both hypothesis may influence L-glutamate production. *OdHI* oxoglutarate dehydrogenase inhibitor, *ODHC* oxoglutarate dehydrogenase complex.

in OdhI could not be identified due to the complexity of the spectra and limited fragmentation by MS/MS. Therefore, other modifications such as 5 methyl group additions (14 Da), and acetylation (42 Da) followed by formylation (28 Da) could not be excluded. As aforementioned, ODHC inactivation arises from the interaction of unphosphorylated OdhI and the OdhA subunit of ODHC. The T14 OdhI phosphorylation inhibits this interaction, resulting in ODHC activation¹⁰. OdhI has a fork-head associated (FHA) domain responsible for the recognition of a phosphothreonine residue. Furthermore, this domain is the region that interacts with the OdhA subunit of ODHC, resulting in its inactivation. Moreover, it was demonstrated that the phosphorylation of the OdhI threonine residue causes drastic changes in its conformation, leading to an auto inhibition of its function, consequently activating ODHC³⁹. More recently, it was reported that K142 succinylation also affects the OdhI–ODHC interaction, hampering the inhibition of ODHC with impacts on *C. glutamicum* L-glutamate production^{12,40}.

Furthermore, as an acetylation site at K52 of OdhI was also described^{12,40}, we investigated such a modification in addition to N-terminal formylation (27.9949 Da + 42.0106 Da) as a possible cause of production of the 70 Da OdhI proteoform. However, these modifications resulted in a considerable loss of matched fragment peaks (data not shown). Therefore, it is improbable that this was the source of such a mass shift. Moreover, the suggested region for the 70 Da Δm is near to the known T14 phosphorylation site of OdhI, consequently raising the question if it could affect its interaction with ODHC. OdhI phosphorylation at T14 is mainly catalyzed by *C. glutamicum* PknG¹⁰. Considering the recognition capacity of the OdhI FHA domain of its phosphothreonine and the close location of this phosphorylation site with the 70 Da Δm proteoform, this modification may affect OdhI inhibition of ODHC, resulting in its activation or inactivation (Fig. 5B). A possible mechanism is the obstruction of OdhI phosphorylation by the presence of the 70 Da PTM, allowing the FHA domain interaction and consequently inhibition of ODHC (Fig. 5B, purple box). Another hypothesis is that the 70 Da modification of OdhI, even having T14 phosphorylated residue, interferes with the FHA domain recognition of the phosphothreonine, resulting in the inhibition ODHC by OdhI (Fig. 5B, purple box). Moreover, the 70 Da mass shift may even behave similarly to the phosphorylated residue, inducing the conformational change of OdhI, inhibiting it and activating ODHC (Fig. 5B, green box). It is less likely that the N-terminal tripeptide truncation causes similar effects, however its influence in OdhI cannot be discarded (Fig. 5B). Furthermore, these alterations may influence L-glutamate production in these bacteria, as a consequence of ODHC regulation (Fig. 5B). Despite these potentially relevant effects of OdhI putative proteoforms, further studies must be performed to investigate the importance and validity of these modifications in OdhI function.

Another protein with relevance to the glutamate production process is the heavy-metal-associated domain (HMA) containing protein (Q8NL68, HMADP). Its corresponding transcript was identified as up-regulated in a series of glutamate overproduction conditions, however its function in this process remains unclear⁴¹. Here we identified three different proteoforms of this protein. One of them presented Δm of -2 Da (supplementary Fig. S6), suggesting the presence of a disulfide bond (UnimodAC: 2020, $\Delta m = -2.015650$ Da), amino acid residue substitution (UnimodAC: 1217, $\Delta m = -2.015650$ Da, Val- > Pro or UnimodAC: 1145, $\Delta m = -1.997892$ Da, Met- > Glu) or didehydro (UnimodAC: 401, $\Delta m = -2.015650$ Da) in the modified region. Moreover, several PrSMs were identified with Δm of approx. 30 Da. Inspection of its spectra resulted in the confirmation of this mass shift, but it is more likely that it resulted from two PTMs, one of 14 Da and other of 16 Da (supplementary Fig. S6), possibly corresponding to methylation and oxidation, respectively. Furthermore, an isotopic envelope could be detected near this identified proteoform ($\Delta m = 30$ Da) with the intact mass of the -2 Da modification of HMADP (supplementary Fig. S6, precursor in blue).

The two putative modifications of HMADP proteoforms are close together in the protein sequence (supplementary Fig. S6). Moreover, no canonical proteoform could be identified. These suggest that the presence of methylation may affect disulfide bond formation. The HMA sequence is found in bacterial proteins conferring toxic heavy metal resistance⁴². To clarify *C. glutamicum* HMADP function, the Q8NL68 protein sequence was blasted against the UniprotKB reference proteomes plus Swiss-Prot database with default parameters in Uniprot, resulting in similarity to copper chaperone and heavy metal transport/detoxification proteins (supplementary Table 3). The transcripts for the copper-responsive two-component system, CopRS, were shown to be up-regulated in *C. glutamicum* during penicillin induction of glutamic acid production⁴³. Recently, it was demonstrated that copper can induce glutamic acid production by *C. glutamicum*, however in a lower amount compared to typical treatments, such as penicillin or biotin limitation⁴⁴. Despite this, the role of HMADP in L-glutamate production remains unclear, however it appears to be a stress response related protein, and it may therefore be regulated or regulate other proteins through PTMs associated with oxidative stress.

Stress response related proteins. Less common Δm values were observed in a peroxiredoxin (Q8NMS6). One proteoform with a mass shift of 154 Da was identified and N-terminal acetylation was also detected (supplementary Fig. S7). Inspection of the MS/MS spectrum allowed us to identify fragments that support the 154 Da mass shift, however, the N-acetylation could not be narrowed down to a shorter sequence region, making it unclear if it was a series of methylation events or indeed an acetylation (supplementary Fig. S7). Considering only one PTM, the mass shift of 154 Da may be caused by three events: addition of glycerophosphate (UnimodAC: 419, $\Delta m = 154.003110$ Da), decanoyl (UnimodAC: 449, $\Delta m = 154.135765$ Da) or 4-oxo-2-nonenal (ONE) (UnimodAC: 721, $\Delta m = 154.099380$ Da). The ONE protein modification is a lipid peroxidation product, caused by reactive species interactions with membrane lipids, suggesting this protein may be localized near the cell membrane of *C. glutamicum*. This molecule can be added to nucleophile amino acid residues⁴⁵. Observing the proposed region of modification, the presence of ONE at suggested site is unlikely, however, near this location are three possible modification sites, C51, K48, and C46 (supplementary Fig. S7). This peroxiredoxin was described as a peroxiredoxin Q with an important role in the oxidative stress response, and

capable of reducing H_2O_2 , t-BOOH, cumene hydroperoxide, and peroxyxynitrite. Interestingly, the C46 proposed to suffer the 154 Da mass shift is the catalytic cysteine of this peroxiredoxin⁴⁶. This suggests this modification may have an important impact on its catalytic activity, and consequently in several *C. glutamicum* oxidative stress responses. Moreover, proteins involved in oxidative stress regulation are often susceptible to oxidative modifications as a process to regulate redox activity in the cell²⁴. This evidence supports the 154 Da mass shift as a ONE modification in *C. glutamicum* peroxiredoxin. However, other possibilities cannot be discarded. Another peroxiredoxin proteoform with a Δm of 186 Da supports the identification of the 154 mass shift in this protein and is likely caused by a second PTM in its sequence of approximately 32 Da. In agreement, near the precursor fragmented for the identification of the 186 Da Δm proteoform, there was an isotopic envelope corresponding to the loss of around 32 Da (supplementary Fig. S7). This suggests the presence of a PTM of 32 Da in addition to the putative ONE modification. Mass differences of 32 Da are usually due to two oxidations in different methionines, but the methionines in this protein sequence are closer to the C-terminal, where several fragments were identified without this mass shift. Another possibility for this mass shift would be a dihydroxylation of cysteine (UnimodAC: 425, $\Delta m = 31.989829$ Da). There are two cysteines near the modified region (supplementary Fig. S7). Despite the pKa of free cysteine being around 8.6, the presence of positively charged residues near the cysteine decreases it by 3–4 units, supporting the oxidation of its thiol group⁴⁷. In agreement, there is an arginine (R54) and lysine (K47) near C51. Moreover, the C51 is the resolving residue and was suggested to be very important in the catalytic process of this peroxiredoxin⁴⁶. Two oxidations of cysteines result in the formation of sulfinic acid, which is an irreversible modification and signal for protein degradation⁴⁷. Therefore, there is a good chance that the peroxiredoxin protein of *C. glutamicum* undergoes oxidative modifications which could regulate its activity and integrity. Furthermore, the suggested modification by a ONE peroxidation suggests a possible role of this protein in cell membrane repair.

Another thioredoxin (Q8NGLG6, trxB1), described as thiol-disulfide isomerase and thioredoxins, was identified by two proteoforms, both with N-terminal cleavage of 26 amino acid residues. Interestingly, one proteoform was identified with approximately -2 Da mass shift in a region near two cysteines (supplementary Fig. S8), suggesting the formation of a disulfide bond. Another proteoform was identified with a Δm of 29.88 Da with ambiguous possibility of modifications (supplementary Fig. S8), such as two oxidations plus a disulfide bond, or methylation followed by oxidation. Moreover, both proteoforms were identified by good quality MS/MS spectra with several fragments representing the two modified forms (supplementary Fig. S8). This thioredoxin was demonstrated to be responsive to disulfide stress, regulated by the SigM sigma factor in *C. glutamicum*⁴⁸. Thioredoxins are known to act as a repair system of oxidized cysteine residues, through its CxxC motif. Its cysteine residues undergo oxidation, forming a disulfide bond, producing the reduced form of cysteine residues in the target protein⁴⁷. The presence of a -2 Da Δm identified near this motif of trxB1 reinforces that it is a disulfide bond. Moreover, the identification of oxidized and reduced proteoforms of a thioredoxin indicates the possibility of comparative studies in quantifying these proteoforms and estimate the redox status of the cell.

Top-down proteomic analysis of the industrial workhorse *Corynebacterium glutamicum* revealed several new putative PTMs of this bacterium related to different biological processes. More precisely, 1125 proteoforms were identified, from 273 proteins. Moreover, membrane proteins and proteins involved in translation seem to be heavily susceptible to PTMs. Proteins relevant to biotechnological and metabolic processes were identified by new proteoforms, which may imply new regulation strategies. For example, the OdhI protein is involved in amino acid production and was identified by a new proteoform suggesting a putative butyrylation or crotonaldehyde peroxidation, that may affect its inhibition of ODHC, therefore influencing glutamate production. The endoproteinase mepB protein was also identified with a new proteoform, cleavage of 93 amino acids residues of its N-terminal, suggesting an undescribed mechanism of its activation in *C. glutamicum* with possible effects on cell wall metabolism. SecG membrane protein, a subunit of protein secretion system had a putative N-terminal formylation, suggesting a degradation signaling with possible consequences in protein secretion. Another membrane protein was identified with a putative N-terminal formylation, MscL, which may influence metabolite efflux. A peroxiredoxin was identified with a putative ONE modification near its catalytic cysteine residue, suggesting a role in oxidative stress responses. Another stress response protein was identified with N-terminal cleavage, that may influence disulfide stress response. The HMADP protein, implicated in glutamate producing conditions and metal transport/detoxification, was identified by two proteoforms, one with a putative disulfide bond and another with putative methylation and oxidation, suggesting possible effects in glutamate production and/or metal detoxification. The influence of these proteoforms in *C. glutamicum* biological processes should be validated and further investigated. For instance, a recent study used genetic engineering of *C. glutamicum* OdhI, introducing mimic amino acid residues, to investigate OdhI acetylation and succinylation influence on its interaction with ODHC⁴⁹. Another possibility would be comparative analysis coupling quantitative top-down proteomics with metabolomics, observing differences in proteoform abundances correlated with metabolite abundance changes.

Methods

Strain and growth conditions. *C. glutamicum* ATCC 13032 was grown in tryptic soy agar (17 g/L pancreatic digest of casein, 3 g/L papaic digest of soybean, 5 g/L NaCl, 2.5 g/L K_2HPO_4 , 2.5 g/L glucose monohydrate, 1.5 g/L agar) and pre-culture was performed overnight in tryptic soy broth under 170 rpm agitation at 30 °C. Afterwards, the pre-culture was used to inoculate three flasks containing CGXII media [20 g/L of $(NH_4)_2SO_4$, 5 g/L of Urea, 40 g/L of glucose, 1 g/L of KH_2PO_4 , 1 g/L of K_2HPO_4 , 0.25 g/L of $MgSO_4 \cdot 7H_2O$, 42 g/L of MOPS, 10 mg/L of $CaCl_2$, 10 mg/L of $FeSO_4 \cdot 7H_2O$, 10 mg/L of $MnSO_4 \cdot 7H_2O$, 1 mg/L of $ZnSO_4 \cdot 7H_2O$, 0.2 mg/L of $CuSO_4$, 0.02 mg/L of $NiCl_2 \cdot 6H_2O$, 0.2 mg/L of biotin, 0.03 mg/L of protocatechuic acid]⁴⁹ [start optical density (OD) = 1.0]. The samples were grown for 36 h under 170 rpm agitation at 30 °C. Growth measurements were assessed by OD_{600} using a spectrophotometer CLARIOstar (BMG LABTECH, Germany).

Sample preparation for LC–MS/MS. *C. glutamicum* cultures were harvested and centrifuged at 12,000×g for 10 min at 21 °C. Pellets were washed with CGXII medium minus glucose, by resuspending the pellet and centrifuging at 12,000×g for 10 min at 21 °C. Cells were then resuspended in lysis buffer (100 mM Tris–HCl, 2 mM DTT, SDS 4%, cOmplete™ EDTA-free Protease Inhibitor Cocktail from Roche, Basel, Switzerland), followed by physical disruption through maceration in liquid nitrogen. After maceration, the solution was centrifuged at 20,000×g for 15 min at 21 °C to separate insoluble particles, and the supernatant with soluble intracellular proteins was stored at –80 °C until further use. Proteins were then quantified by Qubit™ (Invitrogen). A pooled sample was prepared using equal protein quantities of each replicate and 500 µg of the pooled sample was fractionated by Gel-eluted Liquid Fraction Entrapment Electrophoresis—GELFrEE⁵⁰, using a homemade cartridge with 12% resolving gel and 4% stacking gel. First, the sample was diluted in sample buffer 2× (Tris–HCl 0.125 M, SDS 4%, glycerol 20%, DTT 0.1 M, bromophenol blue 0.01%), then it was submitted to electrophoresis using a running buffer (Tris–HCl 0.025 M, Glycine 0.192 M, SDS 0.1%) and a constant current of 10 mA. Fractions were collected according to time (in minutes) after bromophenol blue elution (0 min). Subsequent collections were as follows: 1, 2, 3, 4, 5, 7, 9, 11, 13, 15, 20, 25, 30, 45, 60, 75, 90, and 120 min. The molecular mass range of proteins from each fraction was assessed by SDS-PAGE⁵¹. GELFrEE fractions were then submitted to methanol/chloroform/water precipitation for SDS removal⁵². Briefly, four volumes of methanol were added to the fractions, followed by addition of 1 volume of chloroform and 3 volumes of water, with 30 s of vortexing subsequent to the inclusion of each solution. After water addition and vortex, the solutions were submitted to centrifugation at 20,000×g for 10 min at 21 °C, resulting in the formation of two layers, with proteins floating between them. Then, the superior layer was removed, with care to prevent disturbing the protein pellet, and 3 volumes of methanol were added, followed by gentle mixing and centrifugation at 20,000×g for 10 min at 21 °C. Subsequently, the supernatant was discarded, followed by a second wash with 3 volumes of methanol as described before, and the resulting pellet was dried in a laminar flow bio-hood. After drying, pellets were resuspended in 5% acetonitrile and 0.1% formic acid and stored in –80 °C before LC–MS/MS.

LC–MS/MS. Three technical replicates of GELFrEE fractions 0, 1, 2, 3, 4, 5, 7, 9 and 11 were submitted to top-down analysis on a nano-UHPLC Dionex Ultimate 3000 system coupled to Orbitrap Elite™ mass spectrometer (Thermo Scientific, Bremen, Germany) (LC–MS/MS). Analytical (30 cm, 75 µm I.D.) and trap (4 cm, 100 µm I.D.) columns packed with PLRPS 1000 A, 5 µm (Agilent, California, USA) were used for reverse phase separation. In the analytical column a tip of approximately 1 cm was pulled using a P-2000 instrument (Sutter, California, USA), to be used as emitter in the mass spectrometer. Both columns were kept at room temperature, controlled at approximately 22 °C. Samples were loaded using a flowrate of 3 µL/min for 10 min, under the isocratic condition of 5% Acetonitrile and 0.1% formic acid, then a gradient under flow of 0.230 µL/min was used to elute proteoforms from the column for MS analysis. The gradient was composed of solutions A (formic acid 0.1%) and B (acetonitrile, formic acid 0.1%) and was created by: 5% B (0–10 min), 20% B (10–55 min), 55% B (55–60 min), 85% B (60–80 min) and 5% B (80–90 min). Acquisitions were made during the 90 min gradient in positive mode, and the top-2 most abundant ions were fragmented by stepped high-energy collisional dissociation (HCD) applying normalized collision energy of 25, using two steps, with collision energy width of 10, resulting in two separated fragmentations of ions with normalized collision energy of 20 and 30, and analysis of all resulting fragments together in the Orbitrap. Isolation width of precursors was 25 *m/z*, using a dynamic exclusion duration of 30 s. Source-induced dissociation (SID) was used with 15 eV in MS1 and the first MS2, whilst for the second most intense ion, it was set to 35 eV. Resolutions were 240,000 and 120,000 for MS1 and MS2, respectively. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁵³ partner repository with the dataset identifier PXD038038.

Proteoform identification and characterization. Thermo raw files were converted to MzML format using the MSConvert tool⁵⁴. Identification and characterization of proteoforms were done using the TopPic suite software tool, version 1.4⁵⁵. Briefly, spectra were deconvoluted using TopFD with default parameters followed by identifications against a *C. glutamicum* ATCC 13032 reference proteome (11–2020), obtained from UniProt (<https://www.uniprot.org/>). Two mass shifts between –500 Da and 500 Da were allowed in the identification and the N-terminal forms were set as “NONE” (No modifications), “NME” (N-terminal methionine excision), “NME-ACETYLATION” (N-terminal methionine excision and N-terminal acetylation), and “M-ACETYLATION” (N-terminal acetylation). A false discovery rate strategy was adopted with a decoy database, resulting in the identification of only proteoforms with FDR below 1% at the proteoform spectrum matches (PrSMs) level. Proteoform characterization and annotation were followed by visual interpretation of TopMSV²¹ files. Detected proteins’ MW evaluation and retention times were performed using VisioProt-MS⁵⁶.

Data analysis and functional annotation. Proteoforms identification data were analyzed in R 4.0.2 (R Core Team 2020) using the packages tidyverse⁵⁷ and ggplot2⁵⁸. Briefly, the number of identified mass shifts (Δm) was counted and the frequency of the most common Δm were defined based on rounded values. Then, the number of amino acid residues modified by each Δm was defined. Functional annotation and overrepresentation analysis were performed using the String Cytoscape app^{59,60}, and DAVID⁶¹. Visual interpretation of DAVID overrepresented terms was also carried out using R computer language through the creation of a bubble plot with the tools present in ggplot2.

Data availability

Mass spectra raw files are available via the Proteomics IDENTifications Database Archive (PRIDE, <https://www.ebi.ac.uk/pride/archive/>, ID: PXD038038).

Received: 7 December 2022; Accepted: 11 February 2023

Published online: 14 February 2023

References

1. Becker, J. & Wittmann, C. Bio-based production of chemicals, materials and fuels—*Corynebacterium glutamicum* as versatile cell factory. *Curr. Opin. Biotechnol.* **23**, 631–640 (2012).
2. Sanchez, S., Rodriguez-Sanoja, R., Ramos, A. & Demain, A. L. Our microbes not only produce antibiotics, they also overproduce amino acids. *J. Antibiot. (Tokyo)* **71**, 26–36 (2018).
3. Becker, J. & Wittmann, C. Industrial Microorganisms: *Corynebacterium glutamicum*. *Ind. Biotechnol.* <https://doi.org/10.1002/9783527807796.ch6> (2016).
4. Freudl, R. Beyond amino acids: Use of the *Corynebacterium glutamicum* cell factory for the secretion of heterologous proteins. *J. Biotechnol.* **258**, 101–109 (2017).
5. Okino, S., Inui, M. & Yukawa, H. Production of organic acids by *Corynebacterium glutamicum* under oxygen deprivation. *Appl. Microbiol. Biotechnol.* **68**, 475–480 (2005).
6. Yamamoto, S., Suda, M., Niimi, S., Inui, M. & Yukawa, H. Strain optimization for efficient isobutanol production using *Corynebacterium glutamicum* under oxygen deprivation. *Biotechnol. Bioeng.* **110**, 2938–2948 (2013).
7. Matsumoto, K., Kitagawa, K., Jo, S. J., Song, Y. & Taguchi, S. Production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) in recombinant *Corynebacterium glutamicum* using propionate as a precursor. *J. Biotechnol.* **152**, 144–146 (2011).
8. Ray, D. *et al.* The soil bacterium, *Corynebacterium glutamicum*, from biosynthesis of value-added products to bioremediation: A master of many trades. *Environ. Res.* **213**, 113622 (2022).
9. Carabetta, V. J. & Hardouin, J. Editorial: Bacterial post-translational modifications. *Front. Microbiol.* **13**, 1–2 (2022).
10. Niebisch, A., Kabus, A., Schultz, C., Weil, B. & Bott, M. Corynebacterial protein kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the Odh1 protein. *J. Biol. Chem.* **281**, 12300–12307 (2006).
11. Kim, J., Hirasawa, T., Saito, M., Furusawa, C. & Shimizu, H. Investigation of phosphorylation status of Odh1 protein during penicillin- and Tween 40-triggered glutamate overproduction by *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* **91**, 143–151 (2011).
12. Mizuno, Y. *et al.* Altered acetylation and succinylation profiles in *Corynebacterium glutamicum* in response to conditions inducing glutamate overproduction. *Microbiol. Biotechnol.* **5**, 152–173 (2016).
13. Catherman, A. D., Skinner, O. S. & Kelleher, N. L. Top Down proteomics: Facts and perspectives. *Biochem. Biophys. Res. Commun.* **445**, 683–693 (2014).
14. Smith, L. M. & Kelleher, N. L. Proteoform: A single term describing protein complexity. *Nat. Methods* **10**, 186–187 (2013).
15. Schaffer, L. V. *et al.* Identification and quantification of proteoforms by mass spectrometry. *Proteomics* **19**, 1–15 (2019).
16. Carbonara, K., Andonovski, M. & Coorsen, J. R. Proteomes are of proteoforms: Embracing the complexity. *Proteomes* **9**, 38 (2021).
17. Gault, J. *et al.* Complete posttranslational modification mapping of pathogenic *Neisseria meningitidis* pili requires top-down mass spectrometry. *Proteomics* **14**, 1141–1151 (2014).
18. McCool, E. N. *et al.* Deep top-down proteomics using capillary zone electrophoresis-tandem mass spectrometry: Identification of 5700 proteoforms from the *Escherichia coli* proteome. *Anal. Chem.* **90**, 5529–5533 (2018).
19. Fornelli, L. & Toby, T. K. Characterization of large intact protein ions by mass spectrometry: What directions should we follow?. *Biochim. Biophys. Acta-Proteins Proteom.* **1870**, 140758 (2022).
20. Fellers, R. T. *et al.* ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *Proteomics* **15**, 1235–1238 (2015).
21. Choi, I. K., Jiang, T., Kankara, S. R., Wu, S. & Liu, X. TopMSV: A web-based tool for top-down mass spectrometry data visualization. *J. Am. Soc. Mass Spectrom.* <https://doi.org/10.1021/jasms.0c00460> (2021).
22. Brown, C. W. *et al.* Large-scale analysis of post-translational modifications in *E. coli* under glucose-limiting conditions. *BMC Genomics* **18**, 1–21 (2017).
23. Fridriksson, E. K., Baird, B. & McLafferty, F. W. Electrospray mass spectra from protein electroeluted from sodium dodecylsulfate polyacrylamide gel electrophoresis gels. *J. Am. Soc. Mass Spectrom.* **10**, 453–455 (1999).
24. Lennicke, C. *et al.* Redox proteomics: Methods for the identification and enrichment of redox-modified proteins and their applications. *Proteomics* **16**, 197–213 (2016).
25. Dai, Y. *et al.* Elucidating *Escherichia coli* proteoform families using intact-mass proteomics and a global PTM discovery database. *J. Proteome Res.* **16**, 4156–4165 (2017).
26. Suh, M.-J., Hamburg, D.-M., Gregory, S. T., Dahlberg, A. E. & Limbach, P. A. Extending ribosomal protein identifications to unsequenced bacterial strains using matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics* **5**, 4818–4831 (2005).
27. Piatkov, K., Vu, T., Hwang, C.-S. & Varshavsky, A. Formyl-methionine as a degradation signal at the N-termini of bacterial proteins. *Microb. Cell* **2**, 376–393 (2015).
28. Lee, M. J. & Kim, P. Recombinant protein expression system in *Corynebacterium glutamicum* and its application. *Front. Microbiol.* **9**, 2523 (2018).
29. Nakamura, J., Hirano, S., Ito, H. & Wachi, M. Mutations of the *Corynebacterium glutamicum* NcG1221 gene, encoding a mechanosensitive channel homolog, induce L-glutamic acid production. *Appl. Environ. Microbiol.* **73**, 4491–4498 (2007).
30. Brocker, M., Mack, C. & Bott, M. Target genes, consensus binding site, and role of phosphorylation for the response regulator MtrA of *Corynebacterium glutamicum*. *J. Bacteriol.* **193**, 1237–1249 (2011).
31. Ramadurai, L. & Jayaswal, R. K. Molecular cloning, sequencing, and expression of *lytM*, a unique autolytic gene of *Staphylococcus aureus*. *J. Bacteriol.* **179**, 3625–3631 (1997).
32. Odintsov, S. G., Sabala, I., Marcyjaniak, M. & Bochtler, M. Latent *LytM* at 1.3 Å resolution. *J. Mol. Biol.* **335**, 775–785 (2004).
33. Grabowska, M., Jagielska, E., Czapińska, H. & Bochtler, M. High resolution structure of an M23 peptidase with a substrate analogue. *Nat. Publ. Gr.* <https://doi.org/10.1038/srep14833> (2015).
34. Bott, M. & Brocker, M. Two-component signal transduction in *Corynebacterium glutamicum* and other corynebacteria: On the way towards stimuli and targets. *Appl. Microbiol. Biotechnol.* **94**, 1131–1150 (2012).
35. Hirasawa, T. & Shimizu, H. Recent advances in amino acid production by microbial cells. *Curr. Opin. Biotechnol.* **42**, 133–146 (2016).
36. Kawahara, Y., Takahashi-Fuke, K., Shimizu, E., Nakamatsu, T. & Nakamori, S. Relationship between the glutamate production and the activity of 2-oxoglutarate dehydrogenase in *Brevibacterium lactofermentum*. *Biosci. Biotechnol. Biochem.* **61**, 1109–1112 (1997).
37. Shimizu, H. *et al.* Effects of the changes in enzyme activities on metabolic flux redistribution around the 2-oxoglutarate branch in glutamate production by *Corynebacterium glutamicum*. *Bioprocess Biosyst. Eng.* **25**, 291–298 (2003).
38. Sandberg, M. W., Bunkenborg, J., Thyssen, S., Villadsen, M. & Kofoed, T. Characterization of a novel +70 Da modification in rhGM-CSF expressed in *E. coli* using chemical assays in combination with mass spectrometry. *Amino Acids.* <https://doi.org/10.1007/s00726-021-03004-9> (2021).
39. Barthe, P. *et al.* Dynamic and structural characterization of a bacterial FHA protein reveals a new autoinhibition mechanism. *Structure* **17**, 568–578 (2009).

40. Komine-Abe, A. *et al.* Effect of lysine succinylation on the regulation of 2-oxoglutarate dehydrogenase inhibitor, OdhI, involved in glutamate production in *Corynebacterium glutamicum*. *Biosci. Biotechnol. Biochem.* **81**, 2130–2138 (2017).
41. Kataoka, M. *et al.* Gene expression of *Corynebacterium glutamicum* in response to the conditions inducing glutamate overproduction. *Lett. Appl. Microbiol.* **42**, 471–476 (2006).
42. Bull, P. C. & Cox, D. W. Wilson disease and Menkes disease: New handles on heavy-metal transport. *Trends Genet.* **10**, 246–252 (1994).
43. Hirasawa, T., Saito, M., Yoshikawa, K., Furusawa, C. & Shmizu, H. Integrated analysis of the transcriptome and metabolome of *Corynebacterium glutamicum* during penicillin-induced glutamic acid production. *Biotechnol. J.* **13**, 1700612 (2018).
44. Ogata, S. & Hirasawa, T. Induction of glutamic acid production by copper in *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* **105**, 6909–6920 (2021).
45. Doorn, J. A. & Petersen, D. R. Covalent adduction of nucleophilic amino acids by 4-hydroxynonenal and 4-oxononenal. *Chem. Biol. Interact.* **143–144**, 93–100 (2003).
46. Su, T. *et al.* A thioredoxin-dependent peroxiredoxin Q from *Corynebacterium glutamicum* plays an important role in defense against oxidative stress. *PLoS ONE* **13**, e0192674 (2018).
47. Ezraty, B., Gennaris, A., Barras, F. & Collet, J. F. Oxidative stress, protein damage and repair in bacteria. *Nat. Rev. Microbiol.* **15**, 385–396 (2017).
48. Nakunst, D. *et al.* The extracytoplasmic function-type sigma factor SigM of *Corynebacterium glutamicum* ATCC 13032 is involved in transcription of disulfide stress-related genes. *J. Bacteriol.* **189**, 4696–4707 (2007).
49. Keilhauer, C., Eggeling, L. & Sahl, H. Isoleucine synthesis in *Corynebacterium glutamicum*: Molecular analysis of the *ilvB-ilvN-ilvC* operon. *J. Bacteriol.* **175**, 5595–5603 (1993).
50. Tran, J. C. & Doucette, A. A. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal. Chem.* **81**, 6201–6209 (2009).
51. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
52. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984).
53. Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).
54. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
55. Kou, Q., Xun, L. & Liu, X. TopPIC: A software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **32**, 3495–3497 (2016).
56. Locard-Paulet, M. *et al.* VisioProt-MS: Interactive 2D maps from intact protein mass spectrometry. *Bioinformatics* **35**, 679–681 (2019).
57. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
58. Wickham, H. *ggplot2*. <https://doi.org/10.1007/978-0-387-98141-3> (Springer, 2009).
59. Doncheva, N. T., Morris, J. H., Gorodkin, J. & Jensen, L. J. Cytoscape StringApp: Network analysis and visualization of proteomics data. *J. Proteome Res.* **18**, 623–632 (2019).
60. Szklarczyk, D. *et al.* STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
61. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

Acknowledgements

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Financiadora de Estudos e Projetos (FINEP) and Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF, Process N.: 00193.00001484/2021-91 and 0193001195/2016).

Author contributions

R.M.M. performed the bacterial growth and prepared the sample for mass spectrometry. R.M.M. and T.C.R.W. analyzed the data. J.M.F.S., W.F., M.V.S., and L.H.F.V. performed the mass spectrometer maintenance and injection of samples. C.A.O.R., R.M.M., T.C.R.W., and L.H.V.F. conceived the experiments. All authors revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29857-6>.

Correspondence and requests for materials should be addressed to L.H.F.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

5.2.6 Anexo 2: Produção bibliográfica e participações em eventos científicos ao longo do doutorado

Primeiro autor

- **Melo, R. M.**, de Souza, J. M. F., Williams, T. C. R., Fontes, W., de Sousa, M. V., Ricart, C. A. O., & do Vale, L. H. F. (2023). Revealing *Corynebacterium glutamicum* proteoforms through top-down proteomics. *Scientific Reports*, 13(1), 2602. <https://doi.org/10.1038/s41598-023-29857-6>.
- Vasconcellos, A. F.*, **Melo, R. M.***, Mandacaru, S. C., de Oliveira, L. S., de Oliveira, A. S., Moraes, E. C. dos S., ... Charneau, S. (2022). *Aedes aegypti* Aag-2 Cell Proteome Modulation in Response to Chikungunya Virus Infection. *Frontiers in Cellular and Infection Microbiology*, 12(June), 1–15. <https://doi.org/10.3389/fcimb.2022.920425>. (*These Authors Contributed Equally to the Completion of This Work).

Co-autor

- Moraes, E. C. dos S., Martins-Gonçalves, R., Silva, L. R. da, Mandacaru, S. C., **Melo, R. M.**, Azevedo-Quintanilha, I., ... Trugilho, M. R. O. (2022). Proteomic Profile of Procoagulant Extracellular Vesicles Reflects Complement System Activation and Platelet Hyperreactivity of Patients with Severe COVID-19. *Frontiers in Cellular and Infection Microbiology*, 12(July), 1–17. <https://doi.org/10.3389/fcimb.2022.926352>.
- Santos Júnior, A. de C. M. dos, **Melo, R. M.**, Ferreira, B. V. G., Pontes, A. H., Lima, C. M. R. de, Fontes, W., ... Ricart, C. A. O. (2021). Quantitative proteomics and phosphoproteomics of *Trypanosoma cruzi* epimastigote cell cycle. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1869(5), 140619. <https://doi.org/10.1016/j.bbapap.2021.140619>
- Vasconcellos, A. F., Mandacaru, S. C., de Oliveira, A. S., Fontes, W., **Melo, R. M.**, de Sousa, M. V., ... Charneau, S. (2020). Dynamic proteomic analysis of *Aedes aegypti* Aag-2 cells infected with Mayaro virus. *Parasites & Vectors*, 13(1), 297. <https://doi.org/10.1186/s13071-020-04167-2>.

Eventos científicos

- 2022 - Proteomics of axenic and intracellular amastigote lifeforms of *Trypanosoma cruzi* in replicative and non-replicative conditions presented at the 3rd Iberoamerican Conference on Mass Spectrometry. (Apresentação de pôster)
- 2021 - Unveiling new proteoforms of the industrial workhorse *Corynebacterium glutamicum* through top-down proteomics presented at the 2º Cycle of Webinars “My Friend is MASSA!”. (Apresentação oral)
 - Premiado com menção honrosa.