



**Universidade de Brasília**

**Instituto de Ciências Biológicas  
Departamento de Biologia Celular  
Programa de Pós-Graduação em Biologia Microbiana**

Proteômica *top-down* da bactéria produtora  
de aminoácidos *Corynebacterium*  
*glutamicum*

Reynaldo Magalhães Melo

Brasília, fevereiro de 2019



**Universidade de Brasília**

**Instituto de Ciências Biológicas  
Departamento de Biologia Celular  
Programa de Pós-Graduação em Biologia Microbiana**

**Proteômica *top-down* da bactéria produtora  
de aminoácidos *Corynebacterium  
glutamicum***

**Reynaldo Magalhães Melo**

**Orientador: Prof. Dr. Luis Henrique Ferreira do Vale  
Laboratório de Bioquímica e Química de Proteínas  
(CEL/UnB)**

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Microbiana, do Departamento de Biologia Celular, do Instituto de Ciências Biológicas da Universidade de Brasília como parte dos requisitos para obtenção do título de Mestre em Biologia Microbiana.

Brasília, fevereiro de 2019

# Banca Examinadora

**Prof. Dr. Luis Henrique Ferreira do Vale (presidente, orientador)**

Universidade de Brasília

Departamento de Biologia Celular

Laboratório de Bioquímica e Química de Proteínas

**Prof. Dra. Eliane Ferreira Noronha**

Universidade de Brasília

Departamento de Biologia Celular

Laboratório de Enzimologia

**Prof. Dr. Wagner Fontes**

Universidade de Brasília

Departamento de Biologia Celular

Laboratório de Bioquímica e Química de Proteínas

**Prof. Dra Consuelo Medeiros Rodrigues de Lima (suplente)**

Universidade de Brasília

Departamento de Biologia Celular

Laboratório de Bioquímica e Química de Proteínas

# Agradecimentos

A oportunidade de agradecer a todos os que colaboraram com essa pesquisa possibilita a demonstração de que não se faz ciência sozinho, lição árdua que aprendi ao longo dessa etapa da minha vida acadêmica. Vida essa que só foi possível ter início devido a uma pessoa a qual tenho profunda admiração e respeito, o Professor Carlos André Ornelas Ricart, quem me aceitou como estagiário, mesmo tendo reprovado em sua matéria de bioquímica. Como se não bastasse ter aberto as portas e me aceitado foi sempre orientador ímpar, disponibilizando seu tempo para resolver conflitos científicos e pessoais de todos que o buscavam em sua sala, independentemente de suas tarefas sempre encontrou tempo para auxiliar a todos. Felizmente todas essas regalias se perpetuaram mesmo após minha mudança de coordenação ao adentrar no mestrado, tenho certeza que levo comigo vários aprendizados acadêmicos e para a vida a partir da minha convivência com o professor Carlos André. Anterior à transição ao mestrado, foram extremamente importantes no meu desenvolvimento científico os colegas de laboratórios, que se tornaram amigos importantes, Agenor de Castro Moreira Santos Junior e Arthur Henriques Pontes, a quem devo boa parte dos meus conhecimentos laboratoriais e boas risadas em bares.

Ao findar da graduação fui acolhido pelo programa de pós-graduação em biologia microbiana, sendo orientado pelo professor Edivaldo Ximenes Ferreira Filho, o qual agradeço pela constante atenção e disponibilidade ao longo do primeiro semestre do mestrado. Durante orientação do professor Edivaldo Ximenes também agradeço pelo apoio e auxílio dos colegas do laboratório de enzimologia, Caio de Oliveira Gorgulho Silva e Helder Andrey Rocha Gomes. Ao findar do primeiro semestre na biologia microbiana me fora proposto mudar o tema de minha dissertação e começar a trabalhar sob a orientação do Professor Luis Henrique Ferreira do Vale. Tenho muito a agradecer ao professor Luis, que sempre me orientou e apoiou na aventura que foi desenvolver esse projeto, tanto cientificamente quanto emocionalmente. Agradeço também ao professor Marcelo Valle de Souza que diversas vezes me auxiliou em questões acadêmicas além de ser o coordenador de projeto financiado pela FAP-DF, responsável por parte dos fomentos a materiais utilizados nesse trabalho. Além da FAP-DF, também agradeço à

CAPES por fornecer a minha bolsa de mestrado, possibilitando minha dedicação integral à pesquisa.

Agradeço também a todos os professores e colegas do laboratório de bioquímica e química de proteínas que tornaram a experiência do mestrado excepcional em todos os aspectos, em especial os colegas e amigos Alexandre Ferreira Sant Ana, Raphaela Menezes de Oliveira, Jaques Henrique Miranda, Marcos Rodrigo Alborghetti, Lucas Silva de Oliveira, Samuel Coelho Mandacaru e os professores Wagner Fontes, Aline Martins Araújo e Ana Flávia Parente aos quais eu devo grande aprendizado decorrente de nossas discussões científicas.

Esse trabalho não seria possível sem a ajuda do técnico Nuno Manoel, a quem agradeço pelos reparos e manutenções dos mais diversos equipamentos do laboratório, desde a chaleira elétrica, fornecedora do café diário, ao analisador de aminoácidos utilizado nesse projeto, além das ajudas técnicas os momentos de descontração com suas piadas tremendas também foram essenciais para um ambiente propício no laboratório. Também agradeço de forma especial o técnico Jaques Henrique Miranda, quem se tornou um grande amigo, pela assistência na manutenção e utilização do LTQ orbitrap Elite, além de sua ajuda durante os passeios ocasionais pela UnB para descontrair em semanas difíceis.

Tenho muito a agradecer à minha companheira e namorada Roberta Sousa, por seu apoio e compreensão nos dias difíceis desse mestrado, quando não pude vê-la e prestar a devida atenção ao nosso relacionamento, mesmo assim ela sempre me ajudou de forma incondicional, servindo como um porto seguro em meio as turbulências acadêmicas. Além dela, sua família como um todo foi de grande ajuda, em especial agradeço aos meus sogros, Nélia Rosa e Roberto Carlos, por sempre me receberem com amor em seu lar, que durante o mestrado se tornou minha casa mais próxima à UnB.

Para finalizar deixo meu agradecimento mais profundo aos meus familiares, minha mãe Rozangela Rodrigues Magalhães, meu pai, Geraldo da Silva Melo, e minhas irmãs, Rayana Magalhães Melo e Daine da Silva Melo. Em especial agradeço à minha mãe, a pessoa mais bondosa que conheci e provavelmente conhecerei em minha vida, quem sempre apoiou minhas decisões e possibilitou emocionalmente esse mestrado, e meu pai, com quem sempre tenho oportunidade de aprender a partir de seus constantes

exemplos de superação, além de seu apoio financeiro sem o qual teria sido muito difícil permanecer no mestrado.

# Apoio financeiro

Agradeço ao fomento providenciado pela CAPES, responsável pelo pagamento da minha bolsa. Financiamento feito pela FINEP providenciado para a sala de espectrometria de massa e financiamento da FAP-DF que por meio de projetos de demanda espontânea e pronex fomentou custos variados ao longo do desenvolvimento desse projeto.

*"I do not know what I may appear to the world; but to myself I seem to have been only like a boy playing on the seashore and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me."*

(Isaac Newton)

"Eu prefiro ser essa metamorfose  
ambulante do que ter aquela velha  
opinião formada sobre tudo"  
(Raul Seixas)



## RESUMO

*Corynebacterium glutamicum* é uma bactéria utilizada em indústrias para produção de biomoléculas, existindo cepas capazes de produzir isobutanol, succinato, pantotenato, biopolímeros, entre outros. Apesar desse potencial, sua aplicação industrial mais relevante é na produção de aminoácidos, que é o terceiro grupo de moléculas mais produzido por fermentação no mundo. O sequenciamento do genoma dessa bactéria em 2003 possibilitou o desenvolvimento de novas cepas com mutações dirigidas. A proteômica apresenta grande potencial para a caracterização de funcionamento de vias metabólicas, sendo capaz de identificar modificações pós-traducionais (PTMs), conhecidas por seu papel na regulação do metabolismo. Nesse contexto, a proteômica *top-down* está ganhando espaço como abordagem para identificação de proteoformas, por ser capaz de identificar várias PTMs em proteomas complexos. Esse trabalho teve como objetivo caracterizar o proteoma de *C. glutamicum*, principalmente proteoformas de proteínas reguladoras da biossíntese de aminoácidos e propor mecanismos de regulação não descritos que podem melhorar a eficiência na produção de aminoácidos. Foram realizadas análises proteômicas *top-down* em amostras de *C. glutamicum* cultivadas em meio BT e CGXII. Em meio BT foram detectados 1873 PrSMs e identificadas 492 proteínas não redundantes com 681 proteoformas diferentes. As modificações identificadas foram clivagens e adições de grupamentos formil, acetil ou fosfato. Além disso pôde-se estabelecer condição de degradação proteolítica nessa amostra a partir da contagem de espectros. A análise de cultivo CGXII foi feita com tween 40, para estimular a produção de glutamato, e sem tween 40. A produção de glutamato e aspartato pode ser evidenciada na condição estimulante após 30 hs de estímulo. Apesar da produção de aminoácidos, não foi possível detectar diferenças estatísticas na quantidade de modificações entre as duas condições. Ademais, foi possível verificar efeito de degradação proteolítica muito menor nessas amostras quando comparadas à cultivadas em meio BT. No metabolismo de aminoácidos, foram detectadas proteoformas das enzimas lactato desidrogenase e treonina desidratase fosforiladas que sugerem mecanismo de regulação dessas enzimas por PTMs na biossíntese de L-glutamato, L-lisina e L-leucina. A identificação de vários níveis de PTM não caracterizada no banco de dados foi executada em proteína anotada por similaridade à subunidade da acetolactato sintase e em subunidade pequena da 3-isopropilmalato desidratase, ambas enzimas presentes na biossíntese de L-valina e L-leucina. As massas dessas PTMs sugerem a presença de estearoilção, possível mecanismo de regulação enzimática. Além da adição de grupamentos, foram detectadas proteoformas com pequenas clivagens em suas sequências, dentre elas a corismato mutase apresentou processamento não descrito no banco de dados. Essa clivagem em corismato mutase pode abalar a formação de supercomplexo entre unidades de corismato mutase e DAHP sintase e conseqüentemente afetar a produção de L-fenilalanina e L-tirosina.

**Palavras-chave:** *Corynebacterium glutamicum*, produção de aminoácidos, modificação pós-traducional, proteômica *top-down*.

## Abstract

*Corynebacterium glutamicum* is a workhorse bacterium used to produce many biomolecules, among them there are strains capable to produce isobutanol, succinate, panthothenate, biopolimers and other compounds. Amino acids are the most explored group of molecules for industrial production in *C. glutamicum*. These are the third most produced biomolecules through fermentation worldwide. The *C. glutamicum* genome sequencing was published in 2003, and it increased our capacity to genetically modify it. Among the “omics”, proteomics presents great potential to characterize metabolic pathways, once it is able to identify PTMs that are required for metabolism regulation. Regarding proteomics, top-down proteomics is an approach which is growing within this field due to its capacity to identify many PTMs in complex proteomes. The objective of this project was to characterize the proteome of *C. glutamicum*, mainly those proteins related to amino acids metabolism. Also, propose regulations of mechanisms that can enhance amino acid production in this bacterium. Top-down proteomics experiments were performed in *C. glutamicum* samples cultivated in BT medium and CGXII medium. In BT medium were detected 1873 PrSMs having identified 492 were non-redundant proteins with 681 proteoforms. Those proteoforms were related to cleavages and addition of groups such as formyl, acetyl and phosphates. These analyses were also able to detect sample proteolysis degradation through spectral counting. The CGXII medium was submitted to different conditions: i) with glutamate production, induced by addition of tween 40 and, ii) control condition, without tween 40. Glutamate production was observed after 30 hs of cultivation with tween 40 in contrast to the sample without this molecule, where glutamate production was not detected. Despite this phenotype, no significant changes in modifications patterns was observed by means of spectral counting. Furthermore, in this sample, the degradation effect observed was much less than the one detected in BT medium samples. Phosphorylated proteoforms identified suggest a potential amino acid metabolism regulation. Such proteins as lactate dehydrogenase and threonine dehydratase detected with phosphate addition group might infer a regulation in the biosynthesis of L-glutamate, L-lysine and L-Leucine. Moreover, it was identified mass shift patterns, suggesting different amounts of the same PTM into two proteins (3-isopropylmalate dehydratase and acetolactate synthase, annotated by similarity). These two proteins catalyze biosynthesis steps of L-valine and L-leucine. The mass shifts present in these suggest different levels of stearylation modification, possibly a new mechanism of regulation in this amino acid synthesis. Beyond group additions, proteoforms were detected with small sequence cleavages. Among them, chorismate mutase drives special attention because this protein is part of a supercomplex formed by chorismate mutase and DAHP synthase. Consequently, this cleavage could disturb the production of L-phenylalanine and L-tyrosine.

**Keywords:** *Corynebacterium glutamicum*, amino acid production, post-translational modification, top-down proteomics

## LISTA DE ABREVIATURAS E SIGLAS

- AM.** *Absolute mass*
- BM.** *Biomarker*
- BT.** Meio de cultura basal
- CGXII.** Meio de cultura *Corynebacterium glutamicum* XII
- CoA.** Coenzima A
- E.B.** Extrato Bruto
- GELFrEE.** *Gel-eluted liquid fraction entrapment electrophoresis*
- GRAS.** *Generally Recognized as Safe*
- HH.** *High-High*
- KEGG.** *Kyoto Encyclopedia of Genes and Genomes*
- KO.** *Kegg orthology*
- LC-MS/MS.** Cromatografia líquida acoplada a espectrometria de massas
- LH.** *Low-High*
- LDH.** Lactato desidrogenase
- MM.** Marcador de massa molecular
- OD.** *Optical density*
- ODHC.** 2-oxoglutarato desidrogenase
- P.O.** Precursor observado
- P.T.** Precursor teórico
- PrSM.** *Protein Spectrum Match*
- PTM.** *Pos-translationa modification*
- SDS-PAGE.** *Sodium dodecyl sulfate polyacrylamide gel electrophoresis*
- ST.** *Sequence tag*
- TSA.** *Tryptic Soy Agar*
- TSB.** *Tryptic Soy Broth*

# Lista de figuras

FIGURA 1. CULTURA DE <i>C. GLUTAMICUM</i> ATCC 13032 EM MEIO TSA.....	18
FIGURA 2. VIA METABÓLICA GERAL DO CICLO DO ÁCIDO CÍTRICO, DESTACANDO A ENZIMA ODHC.....	21
FIGURA 3. DIFERENCIAÇÃO DAS ABORDAGENS <i>BOTOM-UP</i> E <i>TOP-DOWN</i> . DEMONSTRAÇÃO DE COMO PODEM OCORRER PERDAS DE INFORMAÇÃO SOBRE MODIFICAÇÕES PÓS TRADUCIONAIS (PTMs).....	23
FIGURA 4. ESQUEMA DE FRACIONAMENTO POR GELFREE E ELETROFORESE EM GEL DE POLIACRILAMIDA DAS FRAÇÕES OBTIDAS. ....	25
FIGURA 5. REPRESENTAÇÃO DAS POSSÍVEIS FORMAS QUE PODEM SER GERADAS A PARTIR DE UM MESMO GENE.. ....	26
FIGURA 6. MAPEAMENTO DAS 15 PTMS MAIS COMUNS ENCONTRADOS NO UNIPROT.....	33
FIGURA 7. FIGURA REPRESENTANDO O TCS DE <i>C. GLUTAMICUM</i> PARA A RESPOSTA A VARIAÇÃO EXTRACELULAR NA CONCENTRAÇÃO DE CITRATO.....	35
FIGURA 8. ESQUEMA DEMONSTRANDO AS DUAS AMINAS POSSÍVEIS DE SEREM ACETILADAS. ....	37
FIGURA 9. FLUXOGRAMA DE TRABALHO CONTENDO DE FORMA SIMPLIFICADA AS ETAPAS REALIZADAS DURANTE ESSE TRABALHO .....	43
FIGURA 10. FLUXOGRAMA APRESENTANDO DETALHES DOS DOIS TIPOS DE CULTIVOS DE BACTÉRIAS UTILIZADOS. ....	45
FIGURA 11. ESQUEMA ILUSTRATIVO DE QUANTIFICAÇÃO DE PROTEÍNAS POR QUBIT.....	50
FIGURA 12. FOTO DE ESTRUTURA UTILIZADA PARA REALIZAÇÃO DE GELFREE. A .....	51
FIGURA 13. FLUXOGRAMA DAS ETAPAS REALIZADAS EM PROCESSAMENTO BIOINFORMÁTICO DOS DADOS ADQUIRIDOS EM MS .....	54
FIGURA 14. CURVAS DE CRESCIMENTO DA BACTÉRIA <i>C. GLUTAMICUM</i> EM MEIO BT E CGXII.....	58
FIGURA 15. CURVA DE CRESCIMENTO DE <i>C. GLUTAMICUM</i> EM MEIO CGXII EM CONDIÇÃO CONTROLE E COM TWEEN 40.....	59
FIGURA 16. CONCENTRAÇÃO DE AMINOÁCIDOS DETECTADOS NO SOBRENADANTE APÓS H DE CULTIVO EM CONDIÇÃO CONTROLE E COM TWEEN 40.....	60

FIGURA 17. SDS-PAGE 12% DAS FRAÇÕES DE F.0 A F.9 PROVENIENTES DE GELFREE E O EXTRATO BRUTO.....	62
FIGURA 18. SDS-PAGE 12% REALIZADO COM AS FRAÇÕES DE F.11 A F.75 PROVENIENTES DE GELFREE .....	62
FIGURA 19. REPRESENTAÇÃO DOS ACESSOS DE PROTEÍNAS NÃO REDUNDANTES DE ACORDO COM A TOLERÂNCIA DE PRECURSOR UTILIZADA EM AMOSTRAS DE BT ADQUIRIDAS POR HH.....	64
FIGURA 20. DIGRAMA DE VENN REPRESENTANDO OS ACESSOS ÚNICOS IDENTIFICADOS E SUAS RELAÇÕES COM AS TOLERÂNCIAS DE PRECURSOR UTILIZADA NAS BUSCAS POR AM ADQUIRIDAS POR LH. ....	65
FIGURA 21. QUANTIDADE DE ACESSOS PROTEICOS NÃO REDUNDANTES IDENTIFICADOS EM CADA TIPO DE AQUISIÇÃO. ....	68
FIGURA 22. DIAGRAMA DE VENN COM A QUANTIDADE DE ACESSOS NÃO REDUNDANTES, PARA CADA TIPO DE BUSCA.....	69
FIGURA 23. TIPOS E QUANTIDADE DE MODIFICAÇÕES IDENTIFICADAS EM TODOS OS PRSMs PROVENIENTES DE CULTIVO EM MEIO BT.....	70
FIGURA 24. QUANTIDADE DE ACESSOS ANOTADOS PARA <i>GENE ONTHOLOGIES</i> (GO) DE FUNÇÃO MOLECULAR DAS PROTEÍNAS IDENTIFICADAS COM N <sup>A</sup> -ACETILAÇÃO.....	71
FIGURA 25. ESPECTRO DE MASSA MS1 APRESENTANDO DOIS PRECURSORES RESPONSÁVEIS PELA IDENTIFICAÇÃO DE Q8NMI3 E SUPOSTAS METILAÇÕES. ....	73
FIGURA 26. ESPECTRO DE MASSA MS1 APRESENTANDO DOIS PRECURSORES RESPONSÁVEIS PELA IDENTIFICAÇÃO DE Q8NSS1 E SUPOSTAS METILAÇÕES.....	74
FIGURA 27. QUANTIDADE DE PRSMs EM RELAÇÃO AOS TIPOS DE MODIFICAÇÕES DETECTADAS E CLASSIFICADAS COMO REFERENTES AO METABOLISMO DE AMINOÁCIDOS.....	78
FIGURA 28. RELAÇÃO ENTRE A MASSA DOS PRECURSORES OBSERVADOS E DOS TEÓRICOS DE ACORDO COM A QUANTIDADE DE PRSMs (%) DE DETECÇÕES VÁLIDAS EM MEIO BT.....	79
FIGURA 29. RELAÇÃO ENTRE A MASSA DOS PRECURSORES OBSERVADOS E DOS TEÓRICOS DE ACORDO COM A QUANTIDADE DE PRSMs VÁLIDAS (%) DA CONDIÇÃO CONTROLE CULTIVADA EM MEIO CGXII. ....	80
FIGURA 30. RELAÇÃO ENTRE A MASSA DOS PRECURSORES OBSERVADOS E DOS TEÓRICOS DE ACORDO COM A QUANTIDADE DE PRSMs VÁLIDAS (%) DA CONDIÇÃO COM TWEEN 40 CULTIVADA EM MEIO CGXII.....	81

FIGURA 31. SCANS REFERENTES À IDENTIFICAÇÃO DE UMA DAS PRSMs DA PROTEÍNA P42457.....	83
FIGURA 32. SDS-PAGE 12% REALIZADO COM 10 µL DAS FRAÇÕES COLETADAS A PARTIR DOS FRACIONAMENTOS EM GELFREE DAS AMOSTRAS PROVENIENTES DE CULTIVO CGXII.....	85
FIGURA 33. RELAÇÃO ENTRE A QUANTIDADE DE ACESSOS ÚNICOS IDENTIFICADOS EM CADA FERRAMENTA DE BUSCA DAS PROTEÍNAS PROVENIENTES DE MEIO CGXII.....	86
FIGURA 34. RELAÇÃO DE ACESSOS ÚNICOS IDENTIFICADOS A PARTIR DE PROTEÍNAS DE CULTIVOS EM MEIO CGXII EM CADA CONDIÇÃO.....	88
FIGURA 35. RELAÇÃO DA QUANTIDADE DE PRSMs DETECTADOS PARA CADA TIPO DE MODIFICAÇÃO NAS CONDIÇÕES CONTROLE E COM ADIÇÃO DE TWEEN 40.....	90
FIGURA 36. QUANTIDADE DE PRSMs DETECTADOS EM PROTEÍNAS RELEVANTES NA PRODUÇÃO DE AMINOÁCIDOS DE ACORDO COM A MODIFICAÇÃO IDENTIFICADA.....	91
FIGURA 37. ESPECTRO DE MASSA COM A RELAÇÃO MASSA CARGA (M/Z) E ABUNDÂNCIA RELATIVA DOS ENVELOPES ISOTÓPICOS RESPONSÁVEIS PELA IDENTIFICAÇÃO DE DUAS FORMAS DA PROTEÍNA Q79VG4.....	96
FIGURA 38. ESPECTRO DE MASSA COM M/Z E ABUNDÂNCIA RELATIVA DE UM DOS ÍONS IDENTIFICADORES DA PROTEÍNA Q79VG4 E ENVELOPES DE CARGAS DEMONSTRANDO POSSÍVEIS METILAÇÕES.....	97
FIGURA 39. MAPA DA BIOSSÍNTESE DE VALINA, ISOLEUCINA E LEUCINA GERADO PELO KEGG.....	98
FIGURA 40. ESPECTRO DE MASSA DO ÍON PRECURSOR DA IDENTIFICAÇÃO DA PROTEÍNA FOSFORIBOSIL-ATP PIROFOSFATASE.....	100
FIGURA 41. QUANTIDADE DE PRSMs DETECTADOS PARA CADA TIPO DE MODIFICAÇÃO PRESENTE NAS PROTEÍNAS RELACIONADAS AO METABOLISMO DE AMINOÁCIDOS DA CONDIÇÃO COM TWEEN 40.....	102
FIGURA 42. ESPECTRO DE MASSA DE DUAS PRSMs IDENTIFICADAS COMO Q8NS29, APRESENTANDO DIFERENÇA DE 42 DA ENTRE SI.....	105

## LISTA DE TABELAS

TABELA 1. QUANTIDADE DE OCORRÊNCIAS DAS 15 PTMS MAIS COMUNS ANOTADOS NO UNIPROT.....	32
TABELA 2. COMPOSIÇÃO DOS MEIOS DE CULTURA BT E CGXII. ....	44
TABELA 3. COMPOSIÇÃO DOS TAMPÕES DE LISE UTILIZADOS NO ESTUDO.....	49
TABELA 4. QUANTIDADE DE PRSMS DETECTADOS, PROTEÍNAS NÃO REDUNDANTES E PROTEOFORMAS NÃO REDUNDANTES IDENTIFICADAS .....	67
TABELA 5. PRSMS DE Q8NMI3 E Q8NSS1 IDENTIFICADOS COM VARIAÇÃO DE MASSA MAIOR OU IGUAL A 14 DA.....	72
TABELA 6. CÓDIGOS DE ACESSOS DETECTADOS EM MEIO BT COM FUNÇÃO ANOTADAS DENTRO DAS VIAS DE BIOSÍNTESE DE AMINOÁCIDOS (AA), CICLO DO ÁCIDO CÍTRICO (TCA), GLICÓLISE/GLICONEOGÊNESE (GLI) E VIA DAS PENTOSAS FOSFATOS (PPP)	76
TABELA 7. QUANTIDADE DE PRSMS, PROTEÍNAS NÃO REDUNDANTES E PROTEOFORMAS DIFERENTES IDENTIFICADAS NAS DUAS CONDIÇÕES DE CULTIVO EM MEIO CGXII...	86
TABELA 8. PROTEÍNAS RELACIONADAS AO METABOLISMO DE AMINOÁCIDOS POSSUINDO RELAÇÃO P.O./P.T $\leq$ 50 %. P.....	92
TABELA 9. PRSMS DETECTADOS COM RELAÇÃO DE DIFERENÇA ENTRE PRECURSOR OBSERVADO E TEÓRICO (P.O./P.T.) MAIOR DO QUE 101%.....	94
TABELA 10. ACESSOS NÃO REDUNDANTES DETECTADOS COM MAIOR SIMILARIDADE (MAIS PRÓXIMO DE 100%) NA RELAÇÃO DE DIFERENÇA ENTRE PRECURSOR OBSERVADO E TEÓRICO (P.O./P.T.). ....	103
TABELA 11. TABELA DE PRSMS DO ACESSO Q8NS29.....	105

# Sumário

RESUMO .....	viii
Abstract.....	ix
LISTA DE ABREVIATURAS E SIGLAS .....	x
LISTA DE TABELAS .....	xiv
1. INTRODUÇÃO.....	17
1.1. <i>Corynebacterium glutamicum</i> .....	17
1.2. <i>C. glutamicum</i> em indústria.....	18
1.3. Produção de aminoácidos .....	19
1.4. Abordagens proteômicas .....	21
1.4. Proteômica de <i>C. glutamicum</i> .....	27
1.5. PTMs em procaríotos .....	30
1.5.1. Fosforilação .....	34
1.5.2. Acetilação .....	36
1.5.3. Formilação .....	39
1.5.4. Metilação .....	39
1.5.5. Proteólise como PTM .....	40
2. Justificativa.....	41
3. Objetivos gerais .....	42
3.1. Objetivos específicos .....	42
4. Material e Métodos .....	43
4.1. Microrganismo e cultivo de bactérias .....	43
4.1.1. Cultivo em meio BT .....	45
4.1.2. Cultivo em meio CGXII .....	45
4.2. Curva de crescimento comparativa entre cultivo em CGXII e BT .....	46
4.3. Análise de aminoácidos .....	47
4.4. Análises estatísticas .....	48
4.5. Análise proteômica .....	48
4.5.1. Lise Celular .....	48
4.5.2. Quantificação de proteínas .....	49



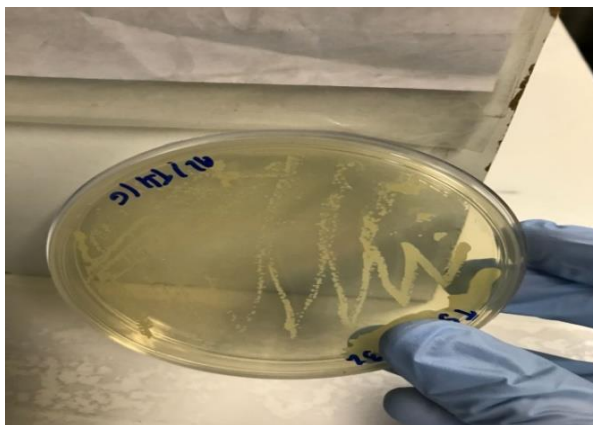
. 4.4.3. GELFrEE .....	50
4.4.4 SDS-PAGE .....	51
4.4.5. Remoção do SDS.....	51
4.4.6. LC-MS/MS .....	52
4.4.6. Bioinformática .....	53
5. Resultados e Discussão.....	57
5.1. Crescimento bacteriano.....	57
5.2. Análise de aminoácidos .....	59
5.3. Análise proteômica de <i>C. glutamicum</i> em meio BT .....	61
5.4. Análise proteômica de <i>C. glutamicum</i> em meio CGXII.....	84
5.4.1. Anotação de proteínas em condição controle .....	90
5.4.2. Anotação de proteínas em condição com tween 40.....	101
5.4.3. Anotação de proteínas presentes na condição controle e tween 40.....	105
6. Conclusões.....	107
7. Perspectivas .....	108
8. Referências bibliográficas .....	109

# 1. INTRODUÇÃO

## 1.1. *Corynebacterium glutamicum*

A subordem Corynebacterineae é caracterizada pela presença de bactérias com ácido micólico na parede celular (TRYFONA; BUSTARD, 2005). 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, e *Mycobacterium tuberculosis* (GEORGI; RITTMANN; WENDISCH, 2005). O gênero *Corynebacterium*, o qual possui 90 espécies, compreende organismos relevantes à saúde humana, animal e também de interesse industrial (TAUCH; BURKOVSKI, 2015). *C. glutamicum* é o representante com maior importância biotecnológica desse grupo taxonômico, além de ser considerado organismo modelo. Essa variedade de organismos com genomas sequenciados e bem estudados foi utilizada para realizar a anotação do genoma de *C. glutamicum* (KALINOWSKI et al., 2003a). Além disso foram feitos estudos em organismos patogênicos a partir de dados coletados para *C. glutamicum* (TAUCH; BURKOVSKI, 2015), demonstrando a característica dessa bactéria como um modelo.

*C. glutamicum* possui a forma de cajado, não é esporulante, geralmente é encontrada no solo e quando cultivada em meio sólido, forma colônias amareladas (Figura 1) (TRYFONA; BUSTARD, 2005).



**Figura 1. Cultura de *C. glutamicum* ATCC 13032 em meio TSA. Cultivo por 12h a 30°C a partir de inóculo realizado com colônia isolada.**

*C. glutamicum* foi isolada pela primeira vez em 1956 por Kinoshita et al., em um zoológico do Japão, sendo denominada inicialmente de *Micrococcus glutamicus*. Na época, a companhia Kyowa Hakko buscava microrganismos selvagens capazes de produzir aminoácidos, quando se deparou com essa bactéria. Posteriormente, *M. glutamicus* e várias outras espécies isoladas como produtoras de aminoácidos tiveram suas identificações alteradas para a espécie *C. glutamicum* (YANG; YANG, 2017). A descoberta de microrganismos produtores de aminoácidos foi marcante para aquela década, porque 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 lhe foi conferido o *status* de “geralmente reconhecida como segura” ou *status* GRAS (*Generally Recognized as Safe*), o que aumenta seu potencial de aplicação biotecnológica (LEE et al., 2016).

## **1.2. *C. glutamicum* em indústria**

*C. glutamicum* possui um sistema metabólico com grande capacidade para produção de biomoléculas de interesse industrial (LEE et al., 2016). Ela já foi utilizada

para a produção de diversos aminoácidos (HIRASAWA; SHIMIZU, 2016), para produzir isobutanol (YAMAMOTO et al., 2013), lactato (OKINO; INUI; YUKAWA, 2005), pantotenato (HÜSER et al., 2005), succinato (LITSANOV; BROCKER; BOTT, 2012), cadaverina (BUSCHKE et al., 2013), alguns polímeros (LIU et al., 2007; MATSUMOTO et al., 2011), entre outras biomoléculas (BECKER; WITTMANN, 2012).

O interesse das indústrias pela bactéria *C. glutamicum* pode ser demonstrado pelo sequenciamento de seu genoma por dois grupos concomitantemente (IKEDA; NAKAGAWA, 2003; KALINOWSKI et al., 2003b) e pelo desenvolvimento de várias cepas por meio de técnicas de melhoramento genético (HIRASAWA; SHIMIZU, 2016). Além disso, a quantidade de registros de patentes, 1700, e estudos, 2700, publicados sobre essa bactéria até o ano de 2015 mostram crescente interesse em entender e aplicar conhecimentos sobre essa bactéria para aprimorar sua produção industrial (LEE et al., 2016).

### **1.3. Produção de aminoácidos**

Atualmente o principal método para a produção industrial de aminoácidos é a fermentação (LEUCHTENBERGER; HUTHMACHER; DRAUZ, 2005), sendo que esse processo fermentativo industrial não se limita a reações anaeróbias, como geralmente é definido o termo, reações fermentativas, em âmbito bioquímico. A fermentação industrial é realizada por meio de pré-cultivo de um microrganismo, esse é então inoculado em um tanque de fermentação esterilizado, também conhecido como biorreator, contendo os nutrientes necessários para o desenvolvimento do organismo, sob condições de temperatura, pH e agitação controlados (HERMANN, 2003).

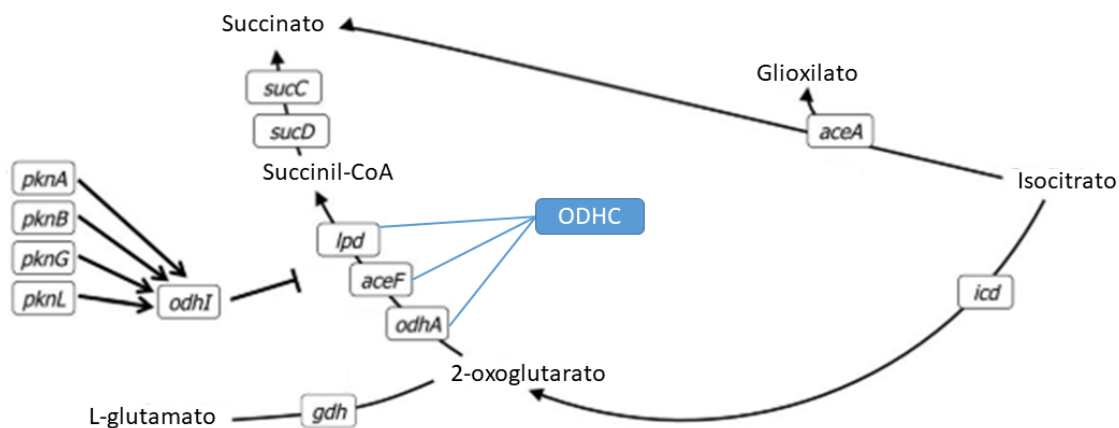
Aminoácidos são o terceiro grupo de moléculas mais produzido pelo método de fermentação industrial, atrás apenas da produção de etanol e antibióticos, respectivamente primeiro e segundo colocados (LEUCHTENBERGER; HUTHMACHER; DRAUZ, 2005). Em 2006 a produção global de aminoácidos foi registrada em 1,5 milhão de toneladas/ano (KATAOKA et al., 2006) e em 2016 alcançou o marco de 2,5 milhões de toneladas/ano (HIRASAWA; SHIMIZU, 2016). 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; HUTHMACHER; DRAUZ, 2005).

A superprodução de ácido glutâmico por *C. glutamicum* pode ser estimulada pela adição de detergentes ao meio de cultura, a utilização de antibióticos durante o crescimento da bactéria ou a utilização de concentrações sub-ótimas de biotina ao longo do crescimento celular. Os efeitos da adição de detergente e a utilização de concentração baixa de biotina foram relacionados ao decréscimo na produção da proteína DtsR1, possuindo similaridade com um dos componentes da Acetil-CoA carboxilase, enzima que catalisa a síntese de malonil-CoA, precursor para produção de ácidos graxos. Esse suposto efeito na produção de ácidos graxos pode alterar a composição e permeabilidade da membrana celular, o que fez pesquisadores acreditarem que mudanças estruturais na membrana seriam a principal causa para secreção ácido glutâmico (SHIMIZU et al., 2003). No entanto, foi descoberto que somente mudanças estruturais na parede celular não são o suficiente para ocasionar a superprodução de aminoácidos (LUBITZ; WENDISCH, 2016). Esse modelo já foi refutado por diversos estudos, como a demonstração da importância do transporte ativo para a secreção de ácido glutâmico (BECKER et al., 2013) e o efeito da utilização de da produção desse aminoácido sobre o complexo 2-oxoglutarato desidrogenase, também conhecido como complexo  $\alpha$ -cetoglutarato desidrogenase (ODHC) (BOULAHYA et al., 2010).

Ao longo dos anos, estudos foram evidenciando a importância da enzima ODHC (EC: 1.2.4.2) sobre o mecanismo de superprodução de glutamato por *C. glutamicum* (BOULAHYA et al., 2010; 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 (figura 2), foi provado que a atividade específica de ODHC está suprimida quando é estimulada a produção de ácido glutâmico (SHIMIZU et al., 2003). Essa enzima é um complexo proteico constituído por três subunidades diferentes, E1 (codificado por *odhA*), E2 (codificado por *aceF*) e E3 (Codificado por *lpd*). (KIM et al., 2009). 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 (figura 2) (NIEBISCH et al., 2006).



**Figura 2. Parte do ciclo do ácido cítrico, destacando a enzima ODHC.** São demonstrados nos quadros brancos os genes codificadores das proteínas responsáveis por cada etapa dessa via. A relação de inibição do complexo ODHC é representada pela presença das cinases codificadas pelos genes *pknA*, *pknB*, *pknG* e *pknL*, responsáveis pela fosforilação de *OdhI*. Adaptado de Wendisch & Lubitz (2016).

#### 1.4. Abordagens proteômicas

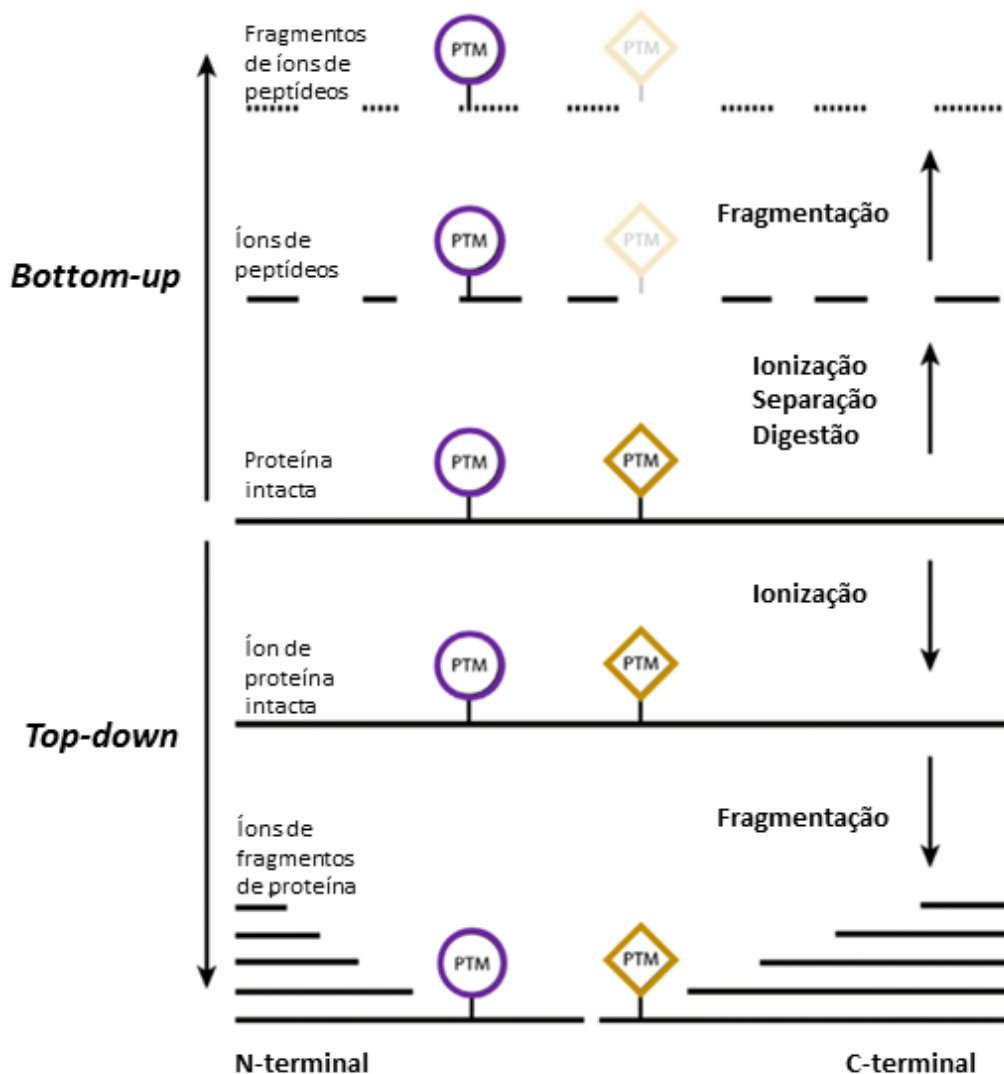
A Proteômica trata do estudo de um conjunto de proteínas relacionando-as a uma função celular (AEBERSOLD; MANN, 2003). É basicamente o estudo dos proteomas, que por sua vez são definidos como o conjunto de proteínas de determinado organismo ou fração de um organismo expressos sob influência externa em um determinado tempo (ZIMMER et al., 2006).

As proteínas são moléculas constituídas por resíduos de aminoácidos codificados a partir de material genético. Apesar de ser possível deduzir a sequência de seus resíduos com base nos genes, nem sempre é factível determinar quando essas proteínas serão produzidas, nem a sua abundância. Essa produção é controlada pelas condições locais e temporais em que o organismo se encontra, como: condição hídrica, interação de

organismo com patógenos, reações a efetores ou inibidores, entre outros. Dessa forma, a proteômica é apresentada como abordagem para estudar diferentes condições sendo capaz de quantificar as proteínas e conseqüentemente interpretar a influência dos estímulos no fenótipo de organismos.

Além de possibilitar a quantificação relativa entre condições, a proteômica é capaz de observar modificações pós-traducionais (PTMs), 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 direta das proteínas para interpretar o fenótipo de organismos.

A espectrometria de massas (MS) tornou-se um dos métodos mais utilizados em estudos proteômicos. Atualmente existem duas principais abordagens para análises proteômicas por MS, uma chamada de *bottom-up* e outra de *top-down*. O fluxograma padrão para análise proteômica *bottom-up* começa com o preparo do organismo/fração de organismo nas condições desejadas e extração das proteínas. Após a extração, as proteínas são clivadas em processo de digestão por protease e tem seus peptídeos ionizados e fragmentados em MS. Na abordagem *top-down* os passos são semelhantes, no entanto as proteínas são ionizadas intactas, não passam por processo de digestão, e sofrem fragmentação em MS (KELLIE et al., 2012). Essa abordagem é favorecida quando se trata da análise de PTMs, pois durante etapa de digestão algumas PTMs podem atrapalhar no processo de clivagem, afetando a identificação de peptídeos, ademais, os peptídeos responsáveis pela identificação de proteínas podem não conter PTMs, no entanto essa informação pode estar presente em outras sequências de resíduos não identificados. Já 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 (CATHERMAN; SKINNER; KELLEHER, 2014) (figura 3).



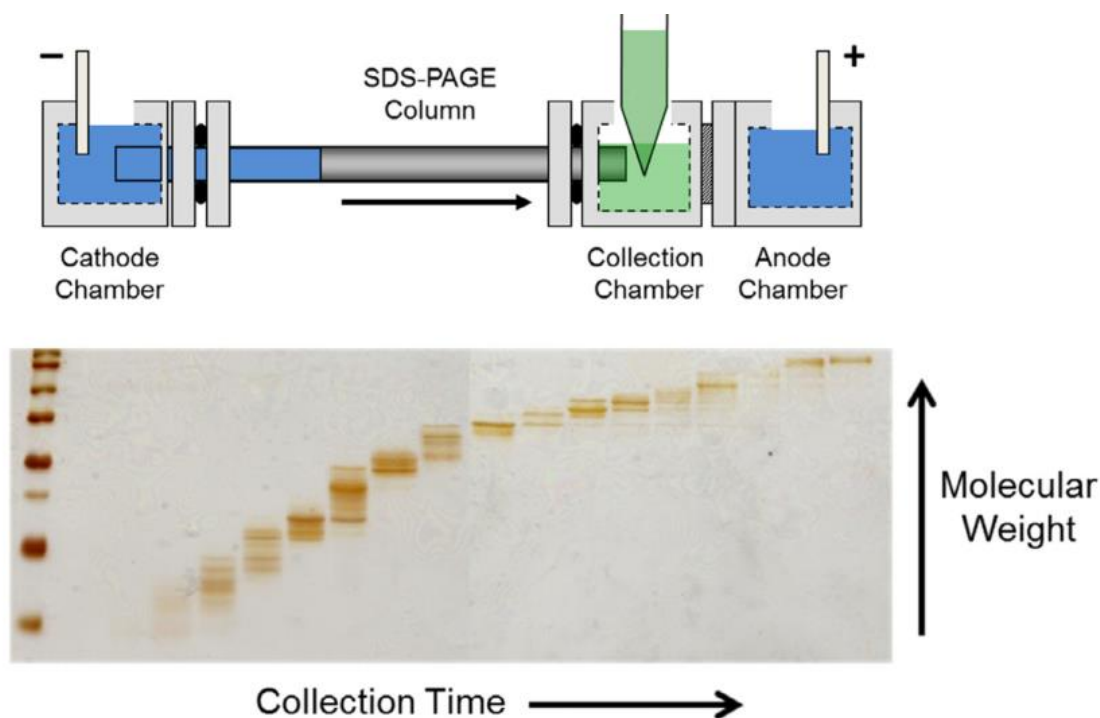
**Figura 3. Ilustração das abordagens *bottom-up* e *top-down*.** Indicação de como podem ocorrer perdas de informação sobre modificações pós traducionais (PTMs) na abordagem proteômica *bottom-up* em contraste à permanência dessas em abordagem *top-down*. Adaptado de Kelleher et al. 2014.

A abordagem *bottom-up* está melhor estabelecida atualmente 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 comparação quantitativa entre condições, fato evidenciado pelas várias estratégias de quantificação relativa de proteínas já desenvolvidas para essa abordagem (XIE et al., 2011).



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*), o sSEC (*serial Size Exclusion Chromatography*) e o CZE (*Capillary Zone Electrophoresis*). 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, justamente pela 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).

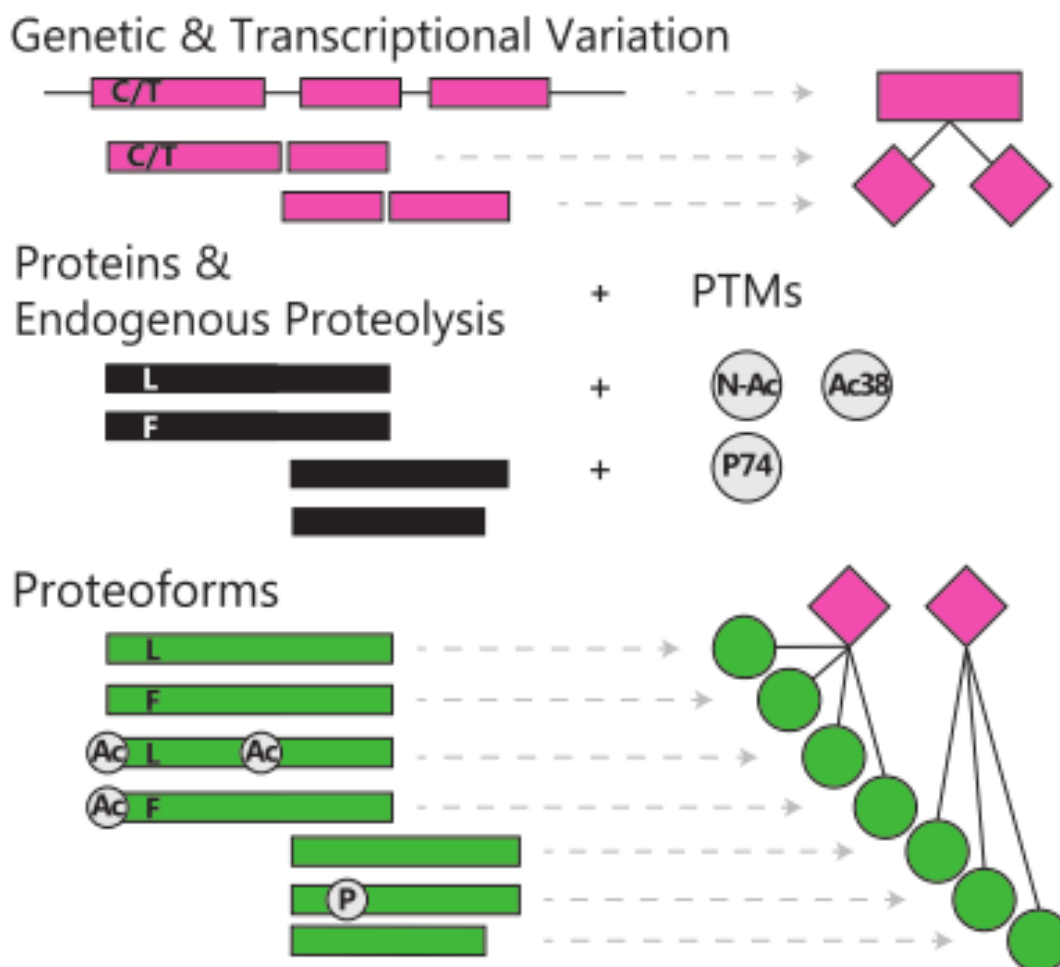
O GELFrEE é uma técnica utilizada para fracionar amostras complexas de proteínas baseada nos tamanhos dessas moléculas. Esse fracionamento pode ser feito de maneira desnaturante, utilizando um detergente para carregar e desnaturar as proteínas (geralmente SDS), quanto não desnaturante, empregando moléculas que não deturpam a conformação das proteínas para carregá-las, como Coomassie. Essa metodologia é baseada na aplicação de corrente elétrica em polos de cargas positiva e negativa, com um gel de poliacrilamida tubular entre eles. A carga das proteínas é uniformizada pela adição de alguma molécula com carga negativa, seja ela SDS (*Sodium Dodecyl Sulfate*) ou *Coomassie blue-G* e aplicadas no topo do gel de poliacrilamida. Ao aplicar a corrente, as proteínas migram pelo gel de acordo com o seu tamanho e são coletadas ao final do tubo em uma câmara de coleta (figura 4) (SKINNER et al., 2015).



**Figura 4. Esquema de fracionamento por GELFrEE e eletroforese em gel de poliacrilamida das frações obtidas.** A câmara indicada pelo símbolo negativo é o anodo e o positivo, o catodo, fazendo as proteínas migrarem do anodo para o catodo devido a carga presente no tampão de amostra e serem coletadas na câmara de coleta de acordo com a eluição da malha do gel, representada em cinza. Adaptado de Catherman et al., 2014.

Quando a análise por MS é feita em nível de peptídeos (*bottom-up*), informações de diversos fragmentos da proteína não são analisadas no espectrômetro de massa, pois são perdidas, seja por não terem sido ionizados, por terem sido perdidos na preparação da amostra ou por não estarem em abundância suficiente para serem detectados. Esses fragmentos não analisados podem conter importantes informações sobre a proteína, principalmente informações sobre PTMs, *alternative splicing*, e isoformas (CHAIT, 2006). Por meio da abordagem *top-down*, muitas dessas informações são mantidas, pois a proteína é analisada em seu estado intacto.

Vocábulo alcinhado há pouco tempo referente a PTMs e outros tipos de variações em proteínas é o termo proteoformas, formas de proteínas provenientes de um mesmo gene, englobando PTMs, proteínas derivadas de *splicing* e isoformas (figura 5)(SMITH, KELLEHER, 2014).



**Figura 5. Representação das possíveis formas que podem ser geradas a partir de um mesmo gene.** Em rosa são indicadas variações de nucleotídeos, em preto e círculos cinzas são possíveis modificações em nível de proteína (N-Ac = Acetilação do N-terminal; Ac38 = Acetilação no resíduo 38; P74 = Fosforilação do resíduo 74) e em verde são mostradas as proteoformas geradas pelas combinações de mais de um tipo de modificações. Adaptado de Cesnik, et al., 2018.

Outra desvantagem da abordagem proteômica *bottom-up* em relação à *top-down* é 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 et al., 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 nesse método não se tem informação experimental da proteína precursora dos peptídeos analisados (NESVIZHSKII; AEBERSOLD, 2005).

#### 1.4. Proteômica de *C. glutamicum*

Apesar dos diversos esforços nos estudos para compreensão de *C. glutamicum*, como análises de fluxos metabólico (WENDISCH et al., 2006), análises metabolômicas (ZHANG et al., 2018a), análises transcriptômicas (PARK et al., 2015), análises por deleção, silenciamento e estímulos de transcrição (BECKER; WITTMANN, 2012), poucos estudos proteômicos, principalmente de alta eficiência, sobre o metabolismo de aminoácidos dessa bactéria foram feitos.

Vários estudos sobre *C. glutamicum* foram realizados utilizando eletroforese bidimensional (2-DE) seguido de identificação de proteínas por abordagem *bottom-up*. Por meio dessas técnicas foi criado o primeiro mapa fosfoproteômico de *C. glutamicum* (BENDT et al., 2003), também foi realizado estudo que identificou 14 proteínas diferencialmente abundantes na utilização de citrato e de glicose por *C. glutamicum* (POLEN et al., 2007). Utilizando abordagem semelhante fora analisada a mutação genética de uma cepa de *C. glutamicum* com depleção parcial da expressão do gene *ftsI* e foi possível observar 22 proteínas mais abundantes na condição com depleção parcial em relação à cepa controle. No entanto, apenas 8 dessas proteínas foram identificadas, dentre as proteínas identificadas 7 são proteínas do metabolismo e uma é a proteína DivIVA, responsável por parte da divisão celular (VALBUENA et al., 2006).

Continuando com abordagem 2-DE, Lüdke et al. estudou as frações citoplasmáticas e membranares de *C. glutamicum* comparando uma cepa selvagem à cepa com deleção do gene *ftsh*, responsável pela produção de uma protease. Essa análise foi realizada para investigar a influência dessa proteína no metabolismo de nitrogênio dessa bactéria. Foram encontradas 6 proteínas com abundâncias diferentes na fração membranares enquanto na citoplasmática foram observadas 43 proteínas. No entanto, nas duas frações estudadas somente 10 proteínas apresentaram diferença maior do que 2 vezes entre as duas condições, a maioria sendo relacionadas à via central do metabolismo de carbono e algumas relacionadas à produção de aminoácidos. Uma observação interessante feita nesse estudo foi a identificação de dois *spots* da proteína gliceraldeído-3-fosfato desidrogenase (GAPDH) com dois pontos isoelétricos diferentes, um deles sendo considerado com abundância relativa maior na cepa mutada do que na cepa selvagem, indicando a presença de uma proteoforma relevante nesse processo, no entanto

essa não pôde ser identificada por limitações técnicas. Como conclusão desse estudo foram apresentadas poucas relações ao esperado metabolismo de nitrogênio, justificada pelo próprio como uma possível limitação da cobertura obtida das proteínas de membrana de apenas 10% do total devido à técnica utilizada. Apesar da baixa relação nesse metabolismo, a deleção apresentou influência relevante no metabolismo central de carbono e na síntese de aminoácidos (LÜDKE et al., 2007).

Em estudo que demonstra o caráter de organismo modelo de *C. glutamicum*, Fanous et al (2010) investigou essa bactéria para entender os mecanismos de estresse ao cobalto e prata por meio de análise proteômica baseada em 2-DE com intuito de propor esse organismo como indicador ambiental. Nesse estudo foram analisadas três condições: controle, adaptada ao estresse de cobalto e adaptada ao estresse de prata. Foram identificadas 19 das 21 proteínas diferencialmente abundantes na condição de estresse a prata e 17 das 22 na condição adaptada à cobalto. Por meio dessas identificações e análise das abundâncias, o estudo chegou à conclusão de que muitas proteínas relacionadas a mecanismos antioxidativos são mais produzidas em ambas condições, estresse à prata e ao cobalto. Além disso, foi possível identificar perfis proteicos únicos de cada condição, propondo assim biomarcadores de condições sub-letais para essas moléculas. No caso de estresse por cobalto, foram propostas duas proteínas envolvidas no processo de biossíntese da parede celular, enzima ramificadora de 1,4-alfa-glucana e epimerase de nucleosídeo difosfato-açúcar. Para estresse por prata foram propostas as enzimas 2-metil-citrato desidratase 1 e 2-metil-citrato desidratase 2 como biomarcadoras (FANOUS et al., 2010).

Reforçando a utilização de *C. glutamicum* como organismo modelo, foi realizado um estudo de otimização da técnica *single-reaction monitoring* (SRM) utilizando *isotopic dilution mass spectrometry* (IDMS) para detecção e quantificação de 14 enzimas em cultivos contendo acetato ou glicose como fontes de carbono (VOGES; NOACK, 2012). O estudo mostrou a linearidade de seus aferimentos por meio de IDMS, onde obtiveram  $R^2 > 0,98$  na maioria dos casos. Por meio da quantificação relativa das proteínas obtidas nas duas condições supracitadas foi possível observar uma via alternativa para o metabolismo de acetato, possuindo maior abundância das enzimas referentes ao ciclo do glioxilato e menor abundância das enzimas glicolíticas.

Outro estudo que demonstra o interesse nessa bactéria como organismo modelo fez a quantificação absoluta de proteínas por espectrometria de massa com QconCAT (*quantitative concatemers*), sendo o primeiro estudo a quantificar de forma absoluta enzimas citoplasmáticas de procariotos. Essa abordagem é baseada na expressão heteróloga de sequências de DNA referentes a determinados digestos proteotípicos, marcados com um átomo pesado. Esses digestos proteotípicos são então utilizados com concentração conhecida como padrão para a quantificação absoluta em MS (BEYNON et al., 2005). Por meio de QconCAT, Voges, et al. (2015) quantificou de forma absoluta 10 enzimas das vias centrais do metabolismo de carbono em células de *C. glutamicum*, quantificando a massa dessas enzimas como correspondente a 10% do peso seco da bactéria. Além disso, a quantificação absoluta foi utilizada em conjunto com dados de taxa de conversão de substrato dessas enzimas, fornecendo valores de atividade máxima teórica para essas enzimas. Foi necessário considerar todas as cópias de proteínas presentes ativas. Esses dados de atividade máxima teórica das enzimas foram comparados com informações de atividade obtidas a partir de uma rede de fluxo metabólicos *in-silico*. Por meio dessa comparação foi possível observar dados congruentes entre as comparações, como no caso da enzima frutose bifosfato aldolase, que apresentou atividade máxima da rede de fluxos similar aos dados obtidos por taxa de conversão e quantificação absoluta. Para PK (piruvato cinase) e PEPCK (fosfoenolpiruvato carboxicinase) os dados de atividade máxima proveniente da rede de fluxos foram muito menores quando comparados à atividade máxima prevista pela inferência da quantificação absoluta, indicando a presença de mecanismos reguladores nessas enzimas, como PTMs ou regulação por ATP (VOGES et al., 2015a).

De acordo com busca na literatura, em 2017 foi realizado o último estudo publicado de *C. glutamicum* utilizando abordagem proteômica. Nele Zhang et al. aborda por meio de análise transcriptômica e proteômica a diferença entre duas cepas de *C. glutamicum*, VWB-1 e ATCC13869. A cepa VWB-1 é capaz de produzir e secretar valina em altas concentrações e foi criada a partir de mutagênese aleatória, enquanto a ATCC13869 é uma cepa selvagem sem essa capacidade. O estudo foi feito para entender quais são as modificações responsáveis pela maior produção e secreção de valina. Para análise proteômica o método utilizado foi mais uma vez a 2-DE. Com essa abordagem foi possível identificar 96 proteínas diferencialmente abundantes entre as cepas, sendo 50 reguladas positivamente e 46 reguladas negativamente comparando a cepa VWB-1 à

ATCC13869. Parcialmente concordante com os dados proteômicos, a análise transcriptômica detectou 1155 transcritos diferencialmente expressos entre as duas cepas, sendo 619 regulados positivamente e 536 regulados negativamente, mostrando a complexidade do mecanismo utilizado por VWB-1 na produção de L-valina.

Estudo que explorou as condições de PTMs em *C. glutamicum* na produção de aminoácidos foram identificados os succiniloma e acetiloma dessa bactéria nas condições de produção de glutamato e controle por abordagem *bottom-up*. Frações foram enriquecidas para essas modificações em resíduos de lisina com anticorpos específicos, e submetidas à cromatografia líquida acoplada a espectrometria de massa (LC-MS/MS). Além disso foram observados padrões de succinilação e acetilação por meio de *Western blotting* com mesmos anticorpos usados no para análise em MS. O experimento evidenciou de forma semi-quantitativa a redução da acetilação e aumento da succinilação na condição de produção de glutamato quando comparada ao controle (MIZUNO et al., 2016). Esse estudo demonstrou possível importância das PTMs na regulação metabólica de aminoácidos em *C. glutamicum*.

## 1.5. PTMs em procariotos

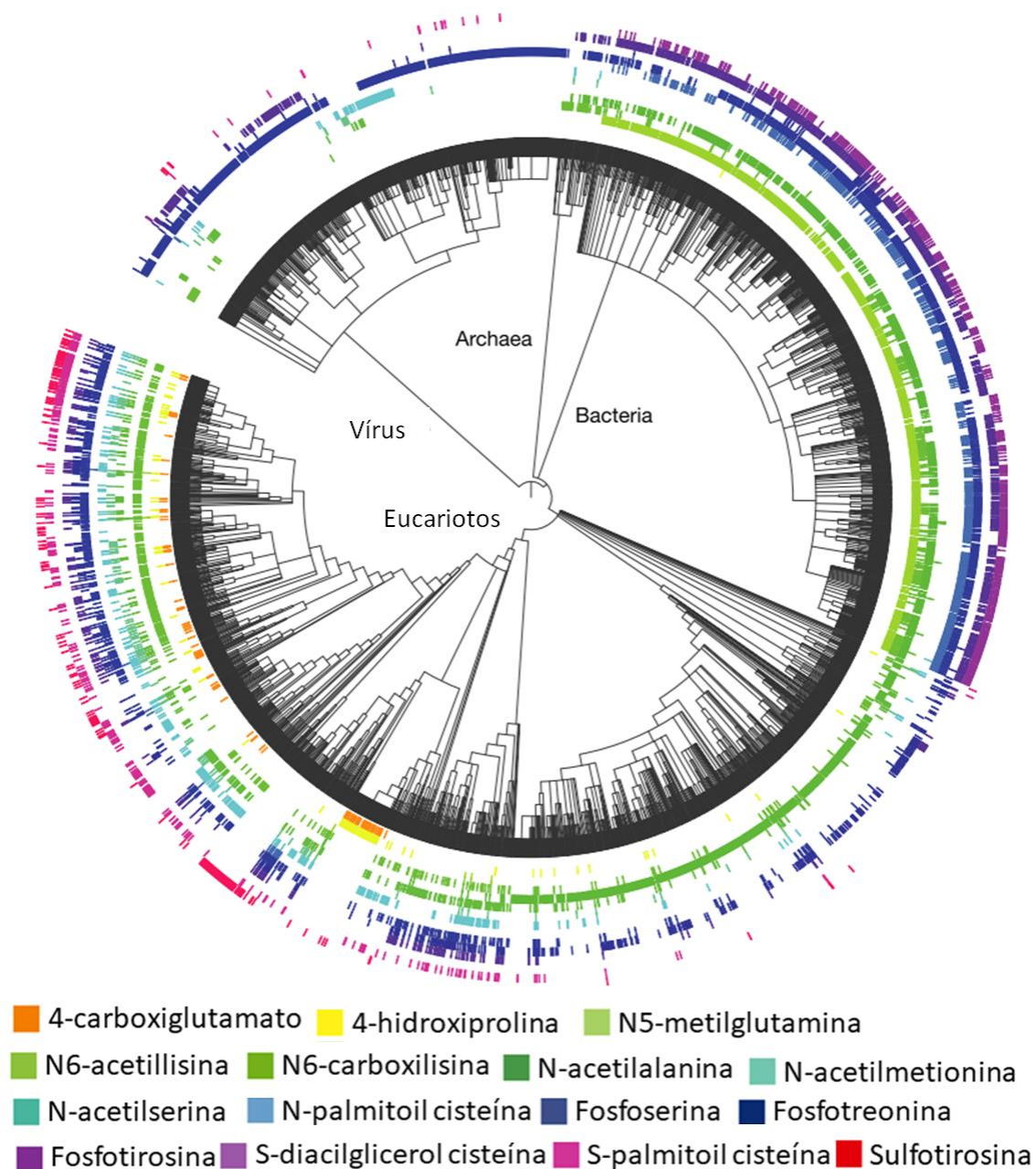
PTMs são capazes de modificar função, estrutura, atividade catalítica, além de alterar interação entre proteínas e outras biomoléculas. Inicialmente com funções subestimadas para procariotos, recentemente estão sendo identificadas cada vez mais PTMs com diferentes funções nesses organismos, avanço oriundo do desenvolvimento em proteômica por espectrometria de massa (CAIN; SOLIS; CORDWELL, 2014). Essas modificações apresentam possibilidade de resposta rápida e de baixa utilização energética, quando comparada à regulação genética. Em bactérias, respostas ágeis são de extrema importância, por essas estarem sujeitas a constantes alterações do meio externo (KOBIR et al., 2011). Funções importantes como transdução de sinais (KOBIR et al., 2011), fator de virulência (SARAVIA et al., 2011), regulação de vias metabólicas (PISITHKUL; PATEL; AMADOR-NOGUEZ, 2015), alteração da interação entre proteínas (BOULAHYA et al., 2010), entre outras (CAIN; SOLIS; CORDWELL, 2014), já foram descritas para PTMs em procariotos.

Alguns tipos de PTMs comuns são fosforilação, metilação, acetilação, carboxilação e clivagens enzimáticas por proteases. Diversas outras modificações já foram descritas e caracterizadas em procariotos e revisões completas foram realizadas nos últimos anos (CAIN; SOLIS; CORDWELL, 2014; GRANGEASSE; STÜLKE; MIJAKOVIC, 2015). Beltrao, et al (2013) contabilizou as 15 PTMs mais comuns anotadas no Uniprot quanto à quantidade de anotações e número de organismos em que foram anotadas. Foi notada grande quantidade de fosforilações, acetilações, carboxilações, metilações, adição de ácidos graxos, glicosilações, dentre outras (tabela 1). Além disso, as 15 PTMs mais comuns foram mapeadas de acordo com suas classificações em uma árvore filogenética (figura 6). Por meio da árvore filogenética é possível observar a grande quantidade de acetilação, fosforilação e metilação anotadas em bactérias (BELTRAO et al., 2013). Dessa forma, essas PTMs serão abordadas com mais profundidade em relação a suas funções e características em procariotos. Juntamente a essas também será abordada a formilação, que recentemente foi proposta como mecanismo de sinalização para degradação de proteínas (PIATKOV et al., 2015).



**Tabela 1. Quantidade de ocorrências das 15 PTMs mais comuns anotados no Uniprot.** Pesquisa realizada com dados de 2013, importante ressaltar que elas estão divididas por resíduos, então adição de um mesmo grupo em resíduos diferentes fora considerada como PTMs diferentes. Adaptado de Beltrao (2013).

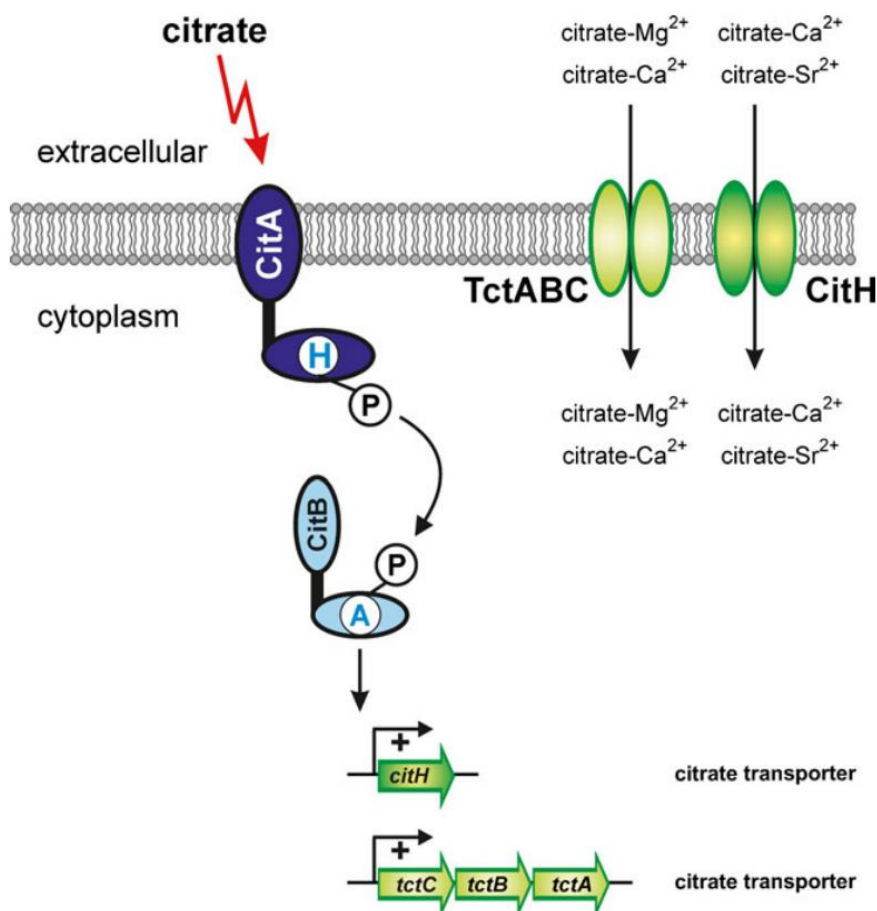
<b>PTM</b>	<b>Número de espécies</b>	<b>Total de ocorrências</b>
Fosfoserina	1701	71744
Fosfotreonina	505	14604
Fosfotirosina	1023	8280
N-acetilalanina	347	2971
N-acetilmetionina	161	1711
N-acetilserina	357	1767
N6-acetilisina	664	13570
N-palmitoil cisteína	577	2056
S-palmitoil cisteína	295	2628
N6-carboxylisina	1178	1600
4-carboxyglutamato	97	1252
N5-metilglutamina	776	1298
4-hidroxiprolina	115	1796
S-diacilglicerol		
cisteína	551	1940
Sulfotirosina	188	1015
Glicil lisina		
isopeptídeo (com NEDD8)	13	47
Glicil lisina		
isopeptídeo (com SUMO)	106	1234
Glicil lisina		
isopeptídeo (com ubiquitina)	252	2815



**Figura 6. Mapeamento das 15 PTMs mais comuns encontrados no Uniprot.** Dados coletados em 2013 de acordo com sua classificação taxonômica, em uma árvore filogenética, criada com a ferramenta *Interactive Tree of Life* do EMBL. Os dados de PTMs considerados foram apenas aqueles anotados em pelo menos 10 espécies e com pelo menos 1000 ocorrências. Adaptado de Beltrao 2013.

### 1.5.1. Fosforilação

A fosforilação de proteínas é a adição de um grupamento fosforil à cadeia lateral de um resíduo proteico. Já foram identificados resíduos de serina, treonina e tirosina fosforilados em procariotos e eucariotos. Ademais, resíduos fosforilados de lisina, arginina, histidina, aspartato, glutamato e cisteína também já foram identificados *in-vivo* em procariotos (CAIN; SOLIS; CORDWELL, 2014). Resíduos de aspartato e histidina fosforilados em procariotos comumente fazem parte da via de sinalização celular em bactérias denominada *two-component system* (TCS) (KOBIR et al., 2011). Essa via funciona como resposta a variações externas, consistindo em uma histidina cinase, capaz de realizar autofosforilação a depender de um estímulo, essa sendo chamada de sensor, que a partir de sua autofosforilação, por meio de ATP, é capaz de doar o grupamento fosfato a um resíduo aspartato de outro componente do TCS, chamado de regulador. Por sua vez o regulador ao ser fosforilado sofre mudança conformacional, induzindo inibição ou ativação da transcrição (figura 7) (BOTT; BROCKER, 2012).



**Figura 7.** Figura representando o TCS de *C. glutamicum* para a resposta a variação extracelular na concentração de citrato. A proteína sensora é representada por CitA e a reguladora por CitB. Em cada uma delas estão destacados seus respectivos domínios com resíduos de fosforilação, representados pelas letras H, de histidina e A, de aspartato. Adaptado de Bott & Brocker (2012).

A fosforilação de resíduos de serina e treonina, normalmente realizadas por proteínas pertencentes à família de cinases de Hanks, também já foi descrita em procariontes, sendo relacionada à regulação metabólica do catabolismo do carbono ou da via do glioxilato. Além disso, muitas vezes é utilizado em processos de enfrentamento a hospedeiros, durante infecção, provavelmente por ser um mecanismo presente em diversos tipos celulares, e muito semelhante ao de eucariotos (KOBIR et al., 2011). Outro mecanismo da fosforilação da arginina é demonstrado em *B. subtilis*, onde essa PTM é capaz de regular mecanismo de proteólise. Essa regulação pode ocorrer tanto por meio da inibição de CtsR, inibidor da protease ClpC, quanto por ativação da própria ClpC por

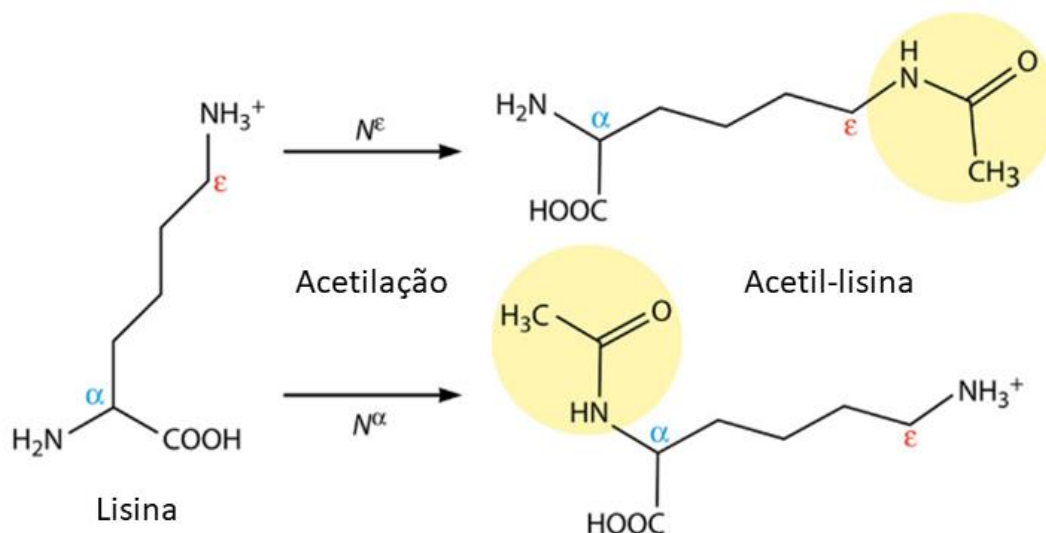
fosforilação de dois sítios de arginina, por meio de McsB (ELSHOLZ et al., 2012; FUHRMANN et al., 2009).

A fosforilação de cisteínas é pouco estudada em procariotos e eucariotos, possivelmente por esse tipo de modificação ser considerado extremamente raro. Acredita-se que essa atue na conformação estrutural das proteínas em que ocorre afetando na formação de pontes dissulfeto (CAIN; SOLIS; CORDWELL, 2014).

A tirosina é um resíduo comumente encontrado fosforilado em procariotos, diferentemente do processo em eucariotos a fosforilação desse resíduo em procariotos normalmente não é feita por cinases de Hanks. Em bactérias essas fosforilações estão relacionadas à produção de exopolissacarídeos, ativação de enzimas intracelulares e até mesmo determinação de localização de proteínas (KOBIR et al., 2011).

### **1.5.2. Acetilação**

A acetilação está presente em procariotos e eucariotos (MA; WOOD, 2011), e a compreensão do impacto biológico dessa modificação pós-traducional está em crescimento (VERDIN; OTT, 2013). Essa PTM consiste na adição de um grupo acetil a um grupo amino, podendo ocorrer nas aminas  $\epsilon$  de resíduos de lisinas ( $N^\epsilon$ -acetilação) ou nas aminas  $\alpha$  de extremidades N-terminais ( $N^\alpha$ -acetilação) (figura 8) (CAIN; SOLIS; CORDWELL, 2014; HENTCHEL; ESCALANTE-SEMERENA, 2015).



**Figura 8. Esquema demonstrando as duas aminas possíveis de serem acetiladas.** As duas possibilidades de acetilação em proteínas são no N-terminal ( $N^\alpha$ -acetilação) ou em resíduos de lisina ( $N^\epsilon$ -acetilação). Adaptado de Hentchel & Escalante-Semerena (2015).

A adição do grupo acetil pode ocorrer de forma enzimática, com auxílio de acetiltransferases, ou não enzimática, sendo que o grupo acetil doado à lisina normalmente é proveniente da molécula acetil-CoA (PISITHKUL; PATEL; AMADOR-NOGUEZ, 2015). Outra possível forma de adição de grupamento acetil foi descrita em *E. coli*, onde essa ocorreu de forma não enzimática a partir de moléculas de acetil-fosfato (WEINERT et al., 2013).

As funções relacionadas a acetilações em procariotos englobam motilidade, transcrição, metabolismo e resposta a estresses (CAIN; SOLIS; CORDWELL, 2014), no entanto seus mecanismos de regulação ainda são pouco conhecidos (WEINERT et al., 2013). As enzimas mais bem caracterizadas responsáveis pela acetilação e deacetilação de lisinas em procariotos são as acetiltransferases Pat e YifQ, e a deacetilase CobB (CAIN; SOLIS; CORDWELL, 2014), as quais participam da regulação da Acetil-CoA sintetase em *Salmonella entérica*. Quando a Acetil-CoA sintetase está acetilada no resíduo lisina<sup>609</sup> torna-se inativa e ao ser deacetilada pela CobB torna-se funcional, regulando assim a produção de acetil-CoA durante crescimento em baixas concentrações de acetato. Além disso, nesse mesmo organismo, essas enzimas são responsáveis por

regular a acetilação em um grande número de outras enzimas metabólicas (CHAN et al., 2011).

Em *E. coli* foi demonstrado o potencial da acetilação como um mecanismo de aumento da resistência a estresses, as cepas com aumento na produção de deacetilase, CobB, apresentaram menor resistência a estresses oxidativos e de temperatura (MA; WOOD, 2011). Outro mecanismo sugerido para a acetilação é a regulação positiva da glicólise, favorecendo a degradação de glicose quando são deletadas deacetilases, além de reduzir o fluxo da glicólise quando acetilases são deletadas do genoma de *S. enterica* (PISITHKUL; PATEL; AMADOR-NOGUEZ, 2015).

Resíduos de lisina podem sofrer outras PTMs a partir da transferência de grupamentos acil-CoA, onde o grupamento acil transferido pode ser succinil, malonil, propionil, causando succinilação, malonilação ou propionilação (CAIN; SOLIS; CORDWELL, 2014). Estipula-se que cada uma dessas modificações, por compartilharem o mesmo resíduo a ser modificado, seja capaz de influenciar na taxa de acetilação de proteínas, mas ainda são necessários mais estudos para compreender como essas modificações podem influenciar na dinâmica da acetilação em resíduos de lisina (CAIN; SOLIS; CORDWELL, 2014).

A N<sup>α</sup>-acetilação é uma PTM muito comum em eucariotos, devido a presença de complexos N<sup>α</sup>-acetiltransferase durante a tradução, ocorrendo muitas vezes a adição do grupamento acetil, (POLEVODA; SHERMAN, 2003). Em procariotos, essa modificação é menos comum, sendo porém bastante observada em lipoproteínas (CAIN; SOLIS; CORDWELL, 2014). Ao contrário da N<sup>ε</sup>-acetilação, a acetilação do N-terminal normalmente é irreversível (HENTCHEL; ESCALANTE-SEMERENA, 2015).

A acetilação não enzimática está diretamente relacionada à concentração de Acetil-Fosfato presente nas células de *E. coli*. Weinert et al. demonstrou que as condições de acetilação de resíduos de lisina estão relacionadas à disponibilidade de Acetil-Fosfato, observando os padrões de acetilação em cepas com deleção nos genes *acs*, *pta* e *ackA*, genes codificadores de enzimas responsáveis pelo metabolismo de acetato e acetil-CoA. O gene *acs* codifica a proteína Acetil-CoA sintetase, responsável pela reação de conversão de acetato para acetil-CoA. Os genes *pta*, que codifica para a enzima fosfotransacetilase, e o *ackA*, que correspondente a enzima acetato cinase, são responsáveis pela formação reversível de acetil-fosfato a partir de acetato (AckA), e

acetil-CoA (Pta). Cepas com a deleção do gene *pta* não acumularam acetilação como esperado e cepa com deleção do gene *ackA* apresentou aumento na quantidade de acetilação. Além disso, nesse mesmo estudo, foi provada a capacidade de acetilação *in-vitro* de resíduos de lisina a partir de acetil-fosfato (WEINERT et al., 2013). Essas evidências sugerem forte influência da molécula acetil-fosfato nos padrões de acetilações em bactérias.

### 1.5.3. Formilação

Formilação é uma modificação na metionina do N-terminal de proteínas que ocorre em procariotos assim como em algumas organelas eucarióticas, como mitocôndrias e cloroplastos. Realizada por formiltransferases acopladas ao ribossomo, é considerada uma modificação co-traducional, ocorrendo durante a tradução. Apesar de ser uma modificação conhecida há algum tempo em proteínas de bactérias, a sua função biológica não era descrita até pouco tempo.

Em 2015 Piatkov et al. apresentou evidências robustas da função da N-formilação em *E. coli* como um indicador para degradação proteolítica, mecanismo proposto em 2010 pelo mesmo grupo. Foi proposto que se o grupamento formil não fosse retirado após a tradução por uma peptídeo-deformilase, esse seria reconhecido por uma enzima responsável pela degradação dessa proteína. As evidências para sustentar essa hipótese são baseadas em vários tipos de experimento, observando padrões de degradação da proteína D2, ou de mutações suas que afetariam na eficiência da formilação, além da deleção de genes de formilação e deformilação (PIATKOV et al., 2015).

### 1.5.4. Metilação

A metilação é caracterizada pela adição de grupamento metil e está amplamente presente procariotos e eucariotos (LANOUILLE et al., 2014), comumente afetando interações de proteína-proteína, podendo causar variações em funções biológicas (ZHANG et al., 2018b).



A incorporação de metil em proteínas pode ser feita de duas formas, N-metilação e O-metilação, a primeira ocorrendo em resíduos de lisina, arginina e glutamina, enquanto a segunda se forma na cadeia lateral de glutamato (CAIN; SOLIS; CORDWELL, 2014). Essa modificação já foi relacionada a proteínas com diversas funções biológicas (CAIN; SOLIS; CORDWELL, 2014).

A função clássica da metilação em procariotos é a regulação da quimiotaxia, por meio de O-metilação da proteína Tsr em seu resíduo de glutamato, em resposta a quimiotratores (RICE; DAHLQUIST, 1991). Mais especificamente, a partir da metilação de proteínas presentes na membrana externa de bactérias, essas afetam cinases sensoras que se fosforilam e possibilitam a transferência de grupamento fosfato a proteínas reguladoras capazes de localizar o aparato flagelar e gerar a modificação no fenótipo de motilidade (CAIN; SOLIS; CORDWELL, 2014).

N-metilações são realizadas em proteínas com diversas funções biológicas em procariotos. Já foi demonstrada uma correlação entre a capacidade de virulência e os padrões de metilação em OmpB (*outer membrane protein B*) (CAIN; SOLIS; CORDWELL, 2014).

### **1.5.5. Proteólise como PTM**

Proteólise é uma modificação irreversível causada pela quebra na ligação peptídica por hidrolases, conhecidas como proteases (CAIN; SOLIS; CORDWELL, 2014), gerando novos N ou C terminais, gerando novas funções ou modificação na função biológica da proteína (ROGERS; OVERALL, 2013) . Esse grupo de enzimas pode ser dividido em duas classes, exopeptidases e endopeptidases, a primeira sendo caracterizada por clivar próximo aos terminais proteicos e a segunda clivando com alta especificidade sequências reconhecidas por seus domínios (CAIN; SOLIS; CORDWELL, 2014). Exemplo de evento de exopeptidases são as clivagens de metioninas em N-terminais de proteínas que ocorrem durante ou logo após a tradução (PIATKOV et al., 2015).

Obviamente a proteólise está envolvida no processo de degradação de proteínas e ciclagem de aminoácidos, no entanto, esse não é seu único papel em organismos.

Mecanismos de replicação, proliferação e morte celular, além de diversos processos patogênicos também são regulados por proteases (ROGERS; OVERALL, 2013).

## 2. Justificativa

A relevância econômica de *C. glutamicum* na indústria é evidente pela quantidade de patentes e estudos relacionados a essa bactéria, além de seus números de mercado (LEE et al., 2016). A necessidade de compreensão sobre os mecanismos de regulação de vias metabólicas de *C. glutamicum* são demonstradas pelo contínuo empenho em entender essa cepa, seja por estudos pangenômicos (YANG; YANG, 2017), transcriptômicos ou proteômicos (ZHANG et al., 2018a). Além disso, alguns estudos indicam seu potencial e atual utilização como organismo modelo (FANOUS et al., 2010; VOGES et al., 2015b). Apesar de tantos esforços já terem sido feitos para o entendimento das vias metabólicas e de proteínas reguladoras de *C. glutamicum*, a caracterização de seu proteoma continua incompleta, principalmente referente a informações sobre PTMs (WENDISCH et al., 2006), embora a importância dessas modificações em vias metabólicas de *C. glutamicum* já tenham sido observada (BENDT et al., 2003). Além disso, PTMs foram descritas como mecanismos de controle de proteínas reguladoras de crescimento e de produção de aminoácidos, como a ODHC (BOULAHYA et al., 2010). Ademais, apenas 2 estudos sobre essa bactéria estão depositados em um dos maiores bancos de dados de análises proteômicas por espectrometria de massa, o PRIDE (<https://www.ebi.ac.uk/pride/archive/>), até o momento dessa publicação, indicando a necessidade de estudos com abordagens mais recentes sobre essa bactéria

A identificação de proteoformas presentes em *C. glutamicum* pode ser utilizada como um recurso para elucidar o mecanismo de enzimas reguladoras de vias metabólicas, dentre elas a de produção de aminoácidos, nessa bactéria. Neste projeto utilizou-se a abordagem proteômica *top-down* para caracterizar as proteoformas de *C. glutamicum*.

### 3. Objetivos gerais

Analisar o proteoma de *C. glutamicum* por abordagem *top-down* e identificar proteoformas e proteínas reguladoras de vias metabólicas dessa bactéria.

#### 3.1. Objetivos específicos

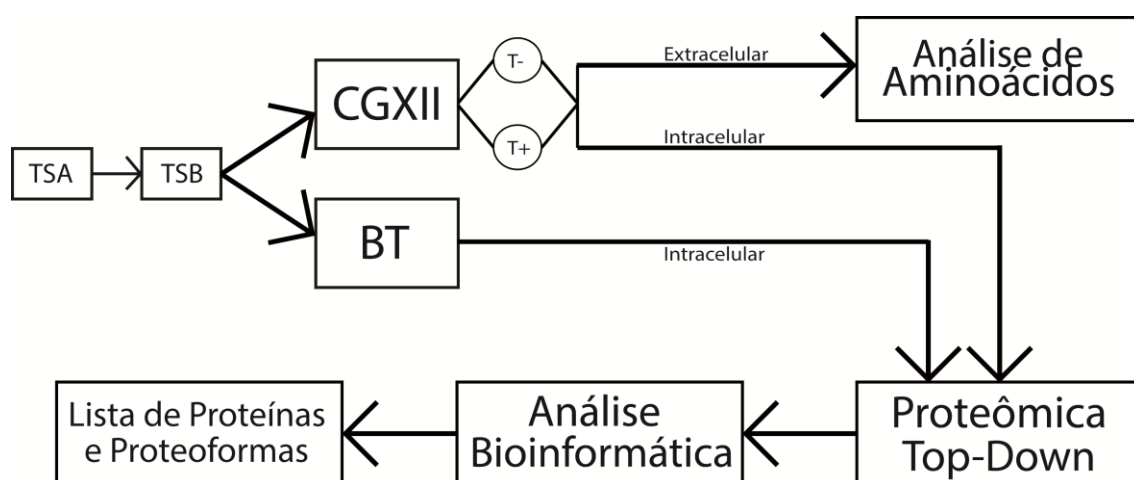
- I- Identificar proteoformas de proteínas no proteoma intracelular de *C. glutamicum* cultivado em meio BT.
- II- Identificar proteoformas presentes no proteoma intracelular de *C. glutamicum* sob influência da tween 40 e em condição controle, ambas cultivadas em meio CGXII.
- III- Identificar proteoformas reguladoras e possíveis alvos para futuros estudos de melhoramento na produção de aminoácidos em *C. glutamicum*.

## 4. Material e Métodos

### 4.1. Microrganismo e cultivo de bactérias

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

A bactéria *C. glutamicum* foi cultivada em dois meios de culturas, chamados BT e CGXII, sob 170 rpm de agitação e 30 °C. A amostra de meio BT que passou por análise proteômica foi coletada 18 h após o inoculo. O cultivo em meio CGXII foram coletadas amostras 6 h, 9 h, 12 h, 24 h e 36 h após o inóculo, sendo que foram submetidas para análise de aminoácidos os sobrenadantes dos tempos 6 h, 12 h e 36 h. Para análise proteômica de meio CGXII foi utilizado o intracelular obtido a partir da coleta de 36 h (figura 2).



**Figura 9. Fluxograma de trabalho contendo de forma simplificada as etapas realizadas durante esse trabalho.** *C. glutamicum* foi primeiramente cultivada em TSA (*Tryptone Soy Agar*) e posteriormente inoculada em TSB (*Tryptone Soy Broth*). As bactérias foram então inoculadas em meio de cultura basal (BT) ou no meio otimizado para *C. glutamicum* XII (CGXII) na presença (T+) ou ausência (T-) de tween 40. Os sobrenadantes dos meios de cultura foram submetidos a análise de aminoácidos e as frações intracelulares à proteômica *top-down*.

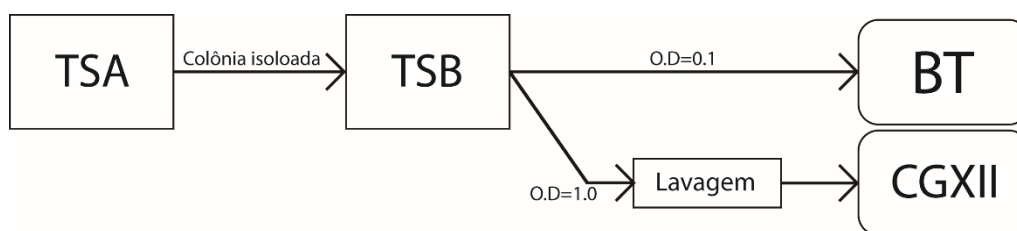
Os meios de cultura utilizados foram o meio TSB (Meio a base de trépticos de soja), proveniente da empresa MERCK, TSA (Meio a base de trépticos de soja com ágar), ambos preparados de acordo com protocolo do fabricante. Além deles, foram utilizados os meios de cultura CGXII e BT (Tabela 1) (KEILHAUER; EGGELING; SAHM, 1993).

**Tabela 2. Composição dos meios de cultura BT e CGXII.**

<b>Meio BT glicose, pH=7.0</b>	
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	7 g/L
KH <sub>2</sub> PO <sub>4</sub>	0,5 g/L
K <sub>2</sub> HPO <sub>4</sub>	0,5 g/L
MgSO <sub>4</sub> · 7H <sub>2</sub> O	0,5 g/L
FeSO <sub>4</sub> · 7H <sub>2</sub> O	6 mg/L
MnSO <sub>4</sub> · H <sub>2</sub> O	4,2 mg/L
Biotina	0,2 mg/L
Glicose	10 g/L
<b>Meio CGXII, pH=7.0</b>	
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	20 g/L
Ureia	5 g/L
KH <sub>2</sub> PO <sub>4</sub>	1 g/L
K <sub>2</sub> HPO <sub>4</sub>	1 g/L
MgSO <sub>4</sub> · 7H <sub>2</sub> O	0,25 g/L
3-morpholinopropanesulfonic acid (MOPS)	42 g/L
CaCl <sub>2</sub>	10 mg/L
FeSO <sub>4</sub> · 7H <sub>2</sub> O	10 mg/L
MnSO <sub>4</sub> · H <sub>2</sub> O	10 mg/L
ZnSO <sub>4</sub> · 7H <sub>2</sub> O	1 mg/L
CuSO <sub>4</sub>	0,2 mg/L
NiCl <sub>2</sub> · 6H <sub>2</sub> O	0,02 mg/L
Biotina	0,2 mg/L
Glicose	40 g/L
Ácido protocatechuico (PCA)	0,03 mg/L

Os procedimentos de cultura realizados foram adaptados de Eggeling & Bott (EGGELING; BOTT, 2005) e Kawaguchi, et al. (KAWAGUCHI et al., 2006). A partir deles foram desenvolvidas duas abordagens com diferenças no meio de cultivo e no

método de inóculo. Brevemente, estoques de *C. glutamicum* armazenados em  $-80\text{ }^{\circ}\text{C}$  foram inoculados em meio TSA e cultivados *overnight* (aproximadamente 12 h), após esse tempo uma colônia isolada desse cultivo foi inoculada em meio TSB também cultivado *overnight*. A partir do aferimento de OD em espectrofotômetro fora calculado o volume necessário para inocular em meio BT (OD = 0,1) ou meio CGXII (OD = 1,0) (Figura 10).



**Figura 10. Fluxograma apresentando detalhes dos dois tipos de cultivos de bactérias utilizados.** O meio TSA significa *Tryptone Soy Agar*, TSB é o meio *Tryptone Soy Broth*, O.D significa *Optical density*, BT é um meio de cultura basal e CGXII é um meio de cultura otimizado para crescimento de *C. glutamicum*.

#### 4.1.1. Cultivo em meio BT

A partir do meio TSB foi feito o inóculo de meio BT com OD inicial de 0,1, esse cultivo foi feito em triplicatas biológicas. As condições utilizadas foram agitação de 170 rpm e  $30^{\circ}\text{C}$  (KAWAGUCHI et al., 2006). Após 18 horas de cultivo, 30 mL de cada replicata foi coletado e centrifugado a  $14000 \times g$  por 15 minutos e seus precipitados foram unidos e lisado em nitrogênio líquido para análise proteômica do intracelular.

#### 4.1.2. Cultivo em meio CGXII

Similar ao realizado para meio BT, em meio CGXII foi feito o inóculo a partir de células cultivadas de meio TSB, no entanto, foi calculado e inoculado o volume necessário para OD inicial = 1.0. Além disso, o volume extraído passou por procedimento

de lavagem adaptado de Neumeyer et al., 2013. Brevemente, o volume coletado foi submetido a centrifugação sob força de 5000Xg por 10 minutos a 4°C. Em seguida, o sobrenadante foi descartado e, então, a solução de sais (Meio CGXII sem fonte de carbono e PCA) foi adicionada e homogeneizada ao precipitado gerado. Após homogeneização, a solução passou por nova centrifugação a 5000 x g por 10 minutos a 4°C e o novo precipitado formado foi então ressuspensionado e homogeneizado em CGXII. Esse procedimento foi feito de forma homogênea com seis volumes iguais referentes ao inóculo de 1.0 OD e as soluções obtidas foram transferidas para seis Erlenmeyers com meio CGXII e incubadas a 30°C sob agitação constante de 170rpm, formando seis replicatas biológicas (NEUMEYER et al., 2013).

Na metade da fase log de crescimento da bactéria em meio CGXII, 6 h depois do inóculo, foi adicionado o detergente tween 40 a três replicatas em concentração final de 4 mg/mL, formando as condições com adição de tween 40 e controle (sem adição de tween 40). Posteriormente, foram coletadas amostras da fase log (6h, 9h, 12h) e estacionária (24h e 36 h) das duas condições para as análises de produção de aminoácidos e curva de crescimento. Os tempos de coletas foram selecionados baseado em estudo de Kim et al. (2011), onde foi observado produção de glutamato a partir da indução por tween 40 (KIM et al., 2011). As amostras coletadas foram analisadas em espectrofotômetro para dados de crescimento e 5 mL de cada cultivo passaram por centrifugação a 14000 x g por 15 minutos, o sobrenadante gerado foi armazenado em -80°C para as análises de quantificação e determinação de aminoácidos livres. O precipitado foi imediatamente submetido ao processo de lise celular de acordo com Fanous et al., 2008 (FANOUS et al., 2008), esse procedimento será detalhado na sessão de análise proteômica.

#### **4.2. Curva de crescimento comparativa entre cultivo em CGXII e BT**

A bactéria *C. glutamicum* foi cultivada em meio BT e CGXII, em microplacas de 16 poços com os inóculos pré-determinados e volume final de meio de 2 mL em cada poço. Foram realizadas triplicatas e cada uma foi considerada como uma réplica biológica. A incubação e a leitura de absorbância foram feitas em equipamento EPOCH 2 da BioTek®.

### 4.3. Análise de aminoácidos

Os sobrenadantes coletados nos tempos 6 h, 12 h e 36 h foram analisados para quantidade de aminoácidos. O preparo para a análise de aminoácidos passou por duas etapas: precipitação em metanol/clorofórmio/água (WESSEL; FLÜGGE, 1984) e troca catiônica. Brevemente, ao volume de 200 µL de cada réplica biológica foram adicionados 800 µL de metanol, 200 µL de clorofórmio e 600 µL de água, com homogeneização de 30 s em vortex entre cada adição. A solução foi então submetida a centrifugação a 21.000 x g e a fase polar foi coletada para passar por troca catiônica. Para o preparo da coluna de troca catiônica, realizada em resina *strongly acidic dowex 50w x8-200*, foram utilizadas ponteiros T-1000-B (Axygen) fechada com algodão de vidro em sua ponta.

Para ativar a resina foram adicionados 2 volumes de HCl 4M e a suspensão resultante foi aquecida em banho maria a 100°C por 1 h. A resina foi, então, lavada com água até o pH da lavagem se igualar ao da água utilizada. Após as lavagens com água, foram adicionados 2 volumes de NaOH 10% (p/v), em seguida a resina foi aquecida e lavada como descrito anteriormente e armazenada em HCl 2M até ser utilizada. Todos os procedimentos de remoção de solução da resina são feitos após a decantação da mesma em frasco de vidro, subsequentemente o líquido é vertido, evitando a saída da resina. Para realizar a troca catiônica a fase polar coletada da precipitação metanol/água/clorofórmio foi acidificada (pH = 2-3) carregada em 500 µL de coluna com resina ativada e eluída com 2 mL de NH<sub>4</sub>OH 1 M.

A quantificação e determinação dos aminoácidos livres secretados pelas bactérias foi realizada em cromatógrafo de troca catiônica de alta eficiência acoplado a detector 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. As amostras eluídas da coluna de troca catiônica foram diluídas em proporção 1:5 em água Milli-Q e aplicadas em triplicatas de 10 µL. O gradiente utilizado foi baseado em três fases móveis (A = H<sub>2</sub>O, B = Acetato e C= NaOH), montado da seguinte forma em uma corrida de 72 minutos, com fluxo constante de 250 µL/min: 24%B por 2 min; 36%B por 9 min; 90%B e 10%C por 6 min; 16% B e 40% C por 2 min; 16% B e 70 % C por 19,1 min; 80% B por 2,1 min; 24% B por 24,8 min.



#### **4.4. Análises estatísticas**

Foram realizadas análises estatísticas comparando as condições com tween 40 e controle, referentes às absorvâncias obtidas para a curva de crescimento e sobre as concentrações de aminoácidos, quando essas foram não nulas. Essas análises foram feitas utilizando o programa PAST (*PAleontological STatistics*). Os dados foram primeiramente submetidos ao teste Shapiro Wilk para observação da normalidade e, posteriormente, aos testes *One Way Anova* e *Tukey's Pairwise Comparisons*, todos com níveis de significância  $P < 0,05$ .

#### **4.5. Análise proteômica**

##### **4.5.1. Lise Celular**

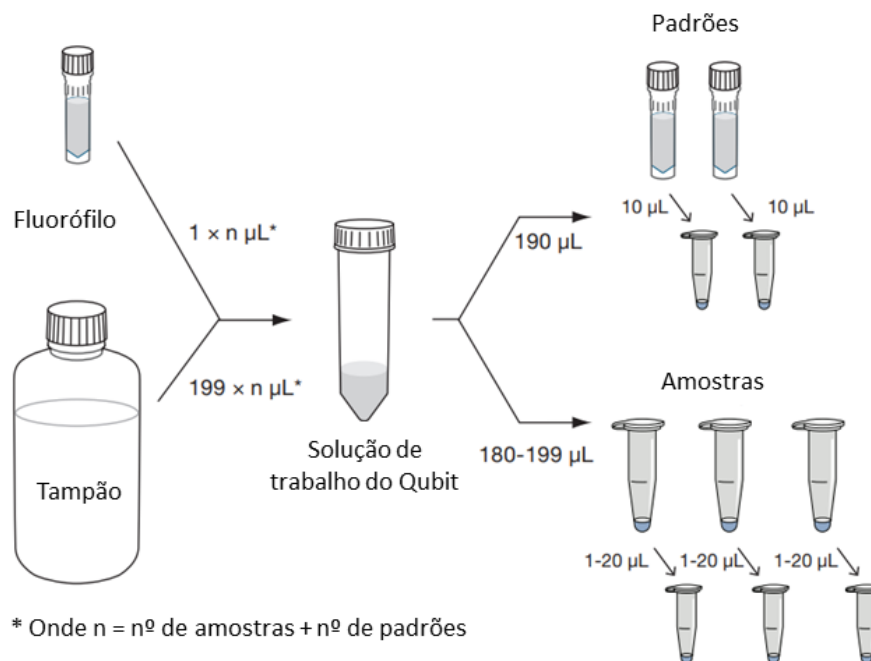
As células coletadas no tempo 18 h após o inoculo de cultivo em meio BT e 36 h de cultivo em meio CGXII foram submetidas a lise para análise do proteoma intracelular. Para as células obtidas em meio BT, foi utilizado o tampão de lise Triton (VOGES et al., 2015b) enquanto para as células adquiridas de meio CGXII foi utilizado o tampão de lise SDS (WIŚNIEWSKI et al., 2009). A composição dos dois tampões é descrita na tabela 2. Antes do procedimento de lise, foi feita uma lavagem das células com o tampão utilizado em cada lise. Após a lavagem, foi adicionado o tampão correspondente a cada meio e o rompimento físico foi realizado em cadinho após o congelamento das células em nitrogênio líquido.

**Tabela 3. Composição dos tampões de lise utilizados no estudo.** Onde IP\* é inibidor de proteases.

Tampões de lise (pH 7,9)			
Tampão Triton		Tampão SDS	
Reagente	Concentração	Reagente	Concentração
KPO <sub>4</sub>	50 mM	tris-HCl	100 mM
EDTA	2 mM	DTT	2 mM
DTT	2 mM	IP* Roche	1comp/10 mL
IP* Roche	1comp/10 mL	SDS	4%
triton-X100	0,20%		

#### 4.5.2. Quantificação de proteínas

A quantificação das amostras resultantes da lise celular foi feita a partir da aferição de fluorescência pela técnica de Qubit™ (Invitrogen). Foram considerados outras duas metodologias de quantificação, Bradford-lowry e micro BCA, no entanto essas não apresentaram reprodutividade quando a análise era feita com as amostras lisadas. Esse resultado era esperado considerando que no tampão de lise contendo SDS continha um interferente para cada uma dessas técnicas, SDS para Bradford-lowry e DTT para microBCA. Considerando esses resultados foi utilizada quantificação por Qubit™ (Invitrogen). Brevemente, preparou-se o reagente fluorescente a partir da mistura de fluoróforo e tampão, comercialmente vendidos pela marca, na proporção de 1:200. Em seguida, alíquotas de 5 µl de amostra diluídas em água miliQ (1:3) foram adicionadas ao reagente para atingir volume final de 200 µl. As misturas foram incubadas por 15 minutos e analisadas no equipamento portátil Qubit 2.0™ (Figura 11). Após a leitura de cada amostra, a concentração final foi obtida multiplicando o valor da leitura pelo fator de diluição escolhido.

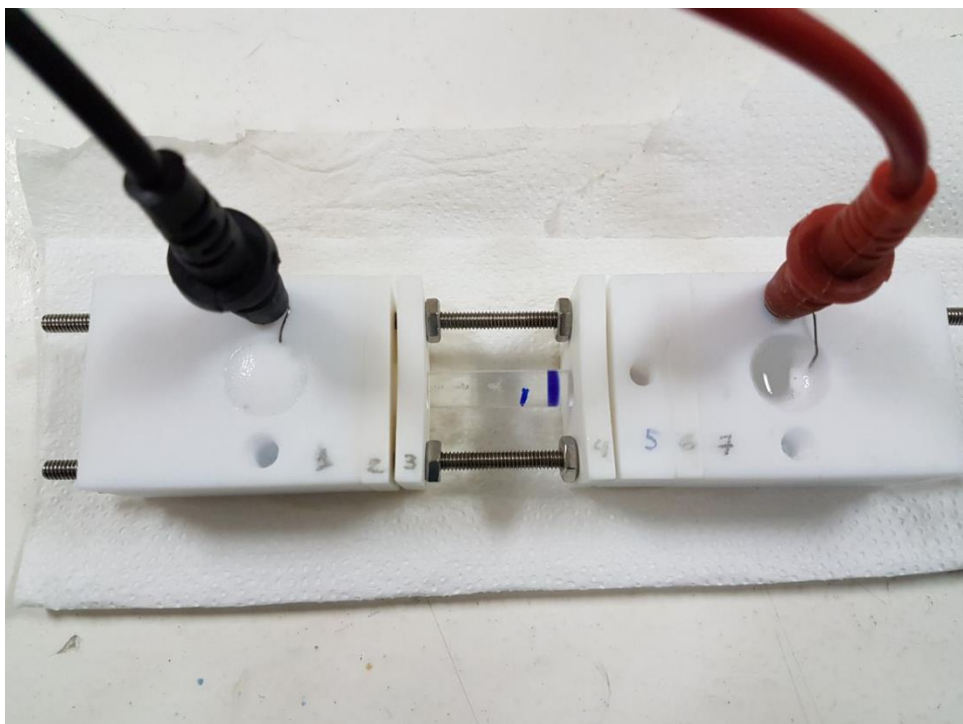


**Figura 11. Esquema ilustrativo de quantificação de proteínas por Qubit.**

Adaptado de Invitrogen®.

#### . 4.4.3. GELFrEE

A partir da quantificação proteica, 500 µg foram reduzidos e alquilados em tampão de amostra desnaturante (4% SDS, 20% glicerol, 0,05 M DTT, 0,004 % azul de bromofenol, 0,125 M Tris HCl) (LAEMMLI, 1970). Em seguida, essas foram fracionadas usando sistema GELFrEE, protótipo em teflon (Figura 12), como descrito anteriormente (TRAN; DOUCETTE, 2008). Brevemente, tubos preenchidos com acrilamida 12% (500 uL de volume) para o gel de separação e 4% para o gel concentrador (300 uL de volume) são acoplados ao sistema GELFrEE, onde as câmaras são completamente preenchidas com tampão de corrida (0,192 M glicina, 0,025 M Tris e 0,1% SDS). Nesse sistema foi aplicada corrente elétrica constante de 10 mA para a separação e as frações foram coletadas após a saída da frente azul do tampão de amostra da malha do gel (0 min, 1 min, 2 min, 3 min, 4 min, 5 min, 7 min, 9 min, 11 min, 13 min, 15 min, 20 min, 25 min, 30 min, 45 min, 60 min, 75 min, 90 min, 120 min). Em seguida, as frações foram submetidas a SDS-PAGE para aferimento da massa e a precipitação utilizando metanol/clorofórmio/água para remoção do SDS.



**Figura 12.** Foto de estrutura utilizada para realização de GELFrEE. As câmaras de carga são as próximas aos fios preto e vermelho, a câmara de coleta é indicada pelo número 5 e a frente de corrida é a linha em azul no tubo transparente.

#### 4.4.4 SDS-PAGE

Antes de serem precipitadas, as frações obtidas por GELFrEE foram submetidas a SDS-PAGE para confirmação e avaliação do fracionamento por massa. O processo utilizado foi baseado em Laemmli (LAEMMLI, 1970). O SDS-PAGE consistiu em malha de separação concentração final de acrilamida/bis-acrilamida (28.5%/1.5%) de 12% e malha de concentração com concentração final de 5%. A corrida foi realizada em sistema mini PROTEAN ® *Tetra Vertical Electrophoresis Cell* (Bio-rad) com amperagem constante a 25 mA para cada gel. Após a eletroforese os géis foram corados usando protocolo de coloração por nitrato de prata (BLUM; HILDBURG; GROSS, 1987).

#### 4.4.5. Remoção do SDS

A remoção do SDS das frações obtidas por GELFrEE foi realizada por precipitação com metanol/clorofórmio/água (WESSEL; FLÜGGE, 1984). Brevemente, as etapas iniciais foram feitas como descrito para a análise de aminoácidos. No entanto, após a centrifugação a fase polar é descartada com cuidado deixando a fase apolar e as proteínas precipitadas logo acima dessa. Então foram adicionados 3 volumes de metanol e a solução foi novamente centrifugada a 21.000 x g por 10 min para lavagem, processo esse que foi repetido 3 vezes. Após a precipitação, as amostras foram ressuspensas em tampão de injeção (ácido fórmico 0,1%, acetonitrila 5%) e armazenadas em -80°C para injeção em LC-MS/MS.

#### 4.4.6. LC-MS/MS

Todas análises do estudo 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™).

A cromatografia líquida de alta performance (HPLC) foi feita com resina PLRP-S 1000 A, 5 µM (Agilent) empacotada em colunas de 30 cm para coluna analítica e 4 cm para coluna de carregamento (*trap*). A fase móvel utilizada foi acetonitrila (B) e água (A) em corridas de 90 min com fluxo constante de 0,3 µL/min. O gradiente para eluição da coluna analítica foi usado de duas formas diferentes, de acordo com as frações injetadas. O gradiente utilizado nas análises de frações com proteínas menores que 50 kDa foi o seguinte: 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. Para as frações com proteínas maiores que 50 kDa o gradiente utilizado foi: 5% B por 10 min, 25% B por 45 min, 55%B por 5 min, 85%B por 20 min e por último 5% B por 10 min.

A aquisição dos espectros foi feita baseada na separação das proteínas em dois grupos, maiores e iguais a 50 kDa ou menores que 50 kDa (PEKAR; ZABROUSKOV; MAKAROV, 2010). Diferenciando o método de aquisição como *High resolution-High resolution* (HH) para proteínas menores que 50 kDa e *Low resolution -High resolution* (LH) para proteínas maiores que 50 kDa (CATHERMAN; SKINNER; KELLEHER,

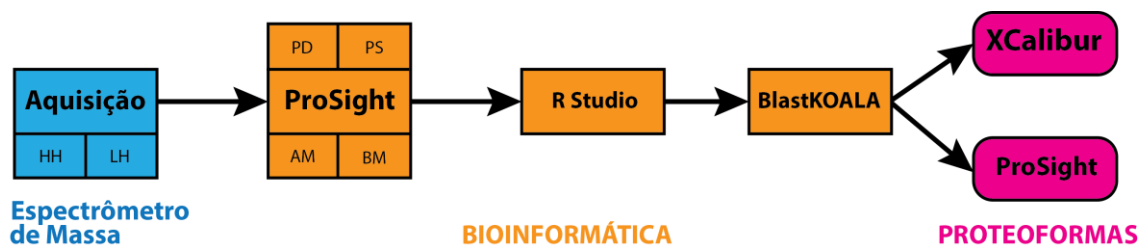
2014, MAKAROV; DENISOV, 2009). Mais especificamente, foram feitas aquisições ao longo dos 90 min de eluição da coluna analítica com SID (*source-induced dissociation*) =15 V, utilizando ESI (*electrospray ionisation*) em modo positivo para os dois métodos de aquisição. A aquisição de MS2 ou MS/MS foi feita de modo dependente de dados (*data dependent acquisition, DDA*) onde os dois íons mais intensos do MS1 são selecionados e fragmentados em HCD (*high collision dissociation*) para seus fragmentos serem analisados em MS2, sendo excluídos dessa seleção íons com sinal < 500, íons dentro de janela de isolamento de 25 m/z e íons com carga < 4. Para aquisição HH o MS1 e MS2 foram realizados por FTMS (orbitrap), com resoluções de 240.000 FWHM e 120.000 FWHM, respectivamente. As frações que passaram por aquisição LH foram submetidas à ITMS (*ion trap mass spectrometer*)-FTMS, com resolução de MS2 de 120.000 FWHM. Duas diferenças no protocolo padrão é a SID do segundo íon mais intenso da aquisição HH que foi modificada para 35 V e a janela de seleção de íons para MS2 em LH modificada para 200 nesse caso.

Na análise do proteoma de *C. glutamicum* cultivado em meio BT foi realizada apenas uma injeção de 20 µL, enquanto para as análises realizadas em meio CGXII foram feitas injeções em triplicatas técnicas de 10 µL cada.

#### 4.4.6. Bioinformática

Após a aquisição dos espectros de massa, esses foram processados e analisados por diversos programas, possuindo tratamentos diferentes no tipo de deconvolução e buscas utilizadas para os espectros provenientes de aquisição HH e LH. Brevemente, os dados adquiridos em espectrômetro de massa passaram por deconvolução em *software* ProSight PC, ou em nó do *Proteome Discoverer*. Após isso a identificação das massas neutras geradas foi realizada em ProSight PC e as listas de identificações geradas foram filtradas estatisticamente e tiveram redundâncias de eventos removidas em plataforma R Studio (Material\_suplementar: CG\_13032\_BTmediu\_DB\_M+F+A+allPTMs, GFC\_36.Rmd e GFT\_36.Rmd). As sequências de identificações não redundantes passaram por blast em plataforma online do KEGG (*Kyoto Encyclopedia of Genes and Genomes*) e os PrSMs (*Protein spectrum matches*) referentes a proteínas anotadas nas vias de glicólise/gliconeogênese, ciclo do ácido cítrico, biossíntese de aminoácidos e via

das pentoses fosfato foram anotados manualmente em XCalibur e ProSightPC 4.0 (figura 13). Todos esses tópicos são mais aprofundados nas páginas a seguir.



**Figura 13. Fluxograma das etapas realizadas em processamento bioinformático dos dados adquiridos em MS.** Os *softwares* utilizados ao longo do processo são demonstrados nos quadros com características específicas de cada um em quadros menores. HH= *High-High*, LH= *Low-High*, PD se refere ao nó presente no *proteome discoverer* capaz de fazer deconvolução de espectros de alta massa, PS se refere à ferramenta presente no ProSight capaz de fazer deconvolução de espectros com massa monoisotópica, AM (*Absolute mass*) e BM (*Biomarker*) são tipos de buscas presentes no ProSight.

#### 4.4.6.1. Deconvolução de espectros

Foram utilizados dois métodos de deconvolução devido a diferença de resolução entre os dois métodos de aquisição, LH e HH. Os dados de MS1 obtidos por LH foram deconvoluídos pelo cRAWler que utiliza o algoritmo ReSpect®, na plataforma Proteome discoverer 2.1 (Thermo Scientific™), através do nó ProSightPD 1.1, isso possibilitou a determinação da massa média de cada componente observado nos espectros. Para deconvolução dos dados de MS2 desse mesmo tipo de aquisição foi utilizado o Xtract® no software ProSightPC 4.0 (Thermo Scientific™), resultando em massas monoisotópicas para os espectros. Os dados adquiridos por metodologia HH tiveram ambos espectros MS1 e MS2 deconvoluídos pelo software Xtract®, proporcionando em ambos precursor e fragmentos dados de massa monoisotópicos.

#### 4.4.6.2. Identificação de proteínas

As identificações dos espectros deconvoluídos foram executadas no software ProSightPC 4.0 (Thermo Scientific™) (YATES; KELLEHER, 2013). O software ProSightPC 4.0 cria banco de dados a partir das sequências proteicas anotadas, nesse estudo foram utilizadas sequências obtidas a partir do UniprotKB (<https://www.uniprot.org/>), aonde os dados de sequências foram obtidos no formato \*.xml, por possuir mais informação sobre proteoformas do que os formatos \*.fasta e \*.txt. A partir desse arquivo, foi criado o banco de dados, utilizando os seguintes parâmetros para gerar proteoformas: clivagem de metionina no N-terminal, acetilação e formilação no N-terminal, além de considerar todas PTMs vindas do arquivo \*.xml (Pontes dissulfeto, O3-fosfopanteteína-L-serina, fosforilação, mono-metilação e oxidação). Ao final da adição de todas proteoformas o banco possuía 3.502 sequências básicas e 17.619 proteoformas.

No *software* ProSightPC 4.0, a busca pode ser feita por três métodos, *absolute mass* (AM), *Sequence tag* (ST) e *Biomarker* (BM). A busca feita por AM considera o íon precursor e compara-o contra o banco de dados selecionado para identificar possíveis íons precursores com  $m/z$  iguais, a partir dessa informação, define fragmentos teóricos para os precursores do banco de dados e compara esses com os fragmentos analisados experimentalmente, todo procedimento pode ser ajustado para tolerâncias definidas pelo usuário (tolerância de busca do precursor e dos fragmentos). Na busca utilizando ST, os dados obtidos são todos em nível de íons já fragmentados, onde o fragmento gerado a partir da proteína é comparado contra um banco de dados de fragmentos teóricos de STs, podendo ocorrer a identificação ou não desses fragmentos. Em casos onde fragmentos não são identificados eles são chamados de “*missing rungs*”, a partir dos dados de identificação os fragmentos analisados são compilados, de forma a gerar algo próximo visualmente de um gráfico, onde os fragmentos serão comparados entre si para verificar qual aminoácido que pode constituir os fragmentos categorizados como *missing rungs* e identificá-los (LEDUC et al., 2004). A metodologia de busca BM utiliza abordagem semelhante à de AM, mas considera possíveis clivagens no precursor teórico, essas clivagens ocasionam nas chamadas proteínas truncadas (CATHERMAN; SKINNER; KELLEHER, 2014). Para as identificações geradas por qualquer tipo de busca é dado um *P-Score* baseado na chance de ocorrer uma identificação aleatória. A ferramenta de busca ST não foi utilizada por não possuir padronização estatística descrita na literatura até o momento.



As buscas realizadas por AM em HH e LH possuíram diferenças na tolerância de massa dos seus precursores. Enquanto para HH foram utilizadas 4 tolerâncias de precursores diferentes (2,3 Da; 2 kDa; 20 kDa e 100 kDa), para LH foram usadas apenas 3 tolerâncias (2 kDa, 20 kDa e 100 kDa), para os fragmentos a tolerância usada foi sempre de 10 ppm (CATHERMAN et al., 2013a). As tolerâncias de identificação de precursor acima de 2,3 Da foram utilizadas para considerar variações de massa entre o precursor observado e o teórico, normalmente relacionadas a PTMs. Especificamente no caso da tolerância de 100 kDa, essa foi utilizada para identificar modificações grandes em proteínas pequenas e a buscar em todo o banco de dados de precursores.

Todos os resultados das buscas foram exportadas e analisados usando uma ferramenta do ProSighPC 4.0 chamada *best hit only*, que remove a redundância de espectros identificados em mais de uma busca. Para a análise em meio BT foram utilizados os dados adquiridos por HH e LH, no entanto para a análise dos dados em meio CGXII só foram utilizados os dados provenientes de aquisição HH.

Os parâmetros estatísticos utilizados para considerar a identificação de proteínas como confiável foram possuir  $E\text{-Value} \leq 10^{-4}$  no caso das buscas por AM ou  $E\text{-Value} \leq 6 \times 10^{-5}$  e  $C\text{-score} > 3$  para as buscas por BM (NTAI et al., 2014). Além disso foi utilizado um filtro de confiança para que somente fossem consideradas como válidas proteínas com pelo menos 4 fragmentos experimentais iguais aos teóricos. Por ser um estudo de descoberta, não de validação de biomarcadores ainda não foram utilizadas metodologias estatísticas baseadas em testes múltiplos. Apesar disso, o  $E\text{-Value}$  utilizado foi definido anteriormente considerando correção baseada em testes multivariados durante buscas em banco de dados muito maior do que o apresentado nesse trabalho (CATHERMAN et al., 2013b), reforçando o caráter confiável das identificações desse estudo.

#### **4.4.6.3. Processamento das identificações**

Após as identificações pelo ProSightPC 4.0, as sequências de aminoácidos das proteínas identificadas foram submetidas a BlastKOALA (<http://www.kegg.jp/blastkoala/>) (KANEHISA; SATO; MORISHIMA, 2016),

ferramenta do KEGG. Os parâmetros utilizados para o blast foram o ID taxonômico 196627 (*C. glutamicum* ATCC 13032), contra o banco de dados “*genus\_prokaryotes.pep*”, fornecido pelo próprio KEGG.

Para interpretação dos dados de identificações foi utilizada a plataforma R (<https://www.r-project.org/>), com auxílio da IDE RStudio (<https://www.rstudio.com/>). Toda análise feita foi salva em formato R *markdown* (\*.Rmd) e está disponível no material suplementar (CG\_13032\_BTmediu\_DB\_M+F+A+allPTMs, GFC\_36.Rmd e GFT\_36.Rmd). Brevemente, todas identificações exportadas pelo ProSightPC 4.0 foram filtradas com os filtros estatísticos, descritos acima. Foi feita verificação de redundâncias nas identificações de eventos devido à quantidade de buscas, ou seja, quando um mesmo evento de fragmentação era responsável por duas identificações diferentes. Sendo observada a presença dessas redundâncias as mesmas foram excluídas de acordo com *P-score*, prevalecendo como identificação válida a com menor valor, gerando assim os PrSMs estatisticamente válidos. Todos os PrSMs estatisticamente válidos de cada condição podem ser encontrados no material suplementar (“meioBT\_estatistica”, “MeioCGXII\_36h\_controle\_estatistica” e “MeioCGXII\_36h\_tween40\_estatistica”).

A partir dessas identificações foram quantificados os números de acessos e proteoformas não redundantes. A contabilização das proteoformas anotadas automaticamente pelo ProSightPC 4.0 foi feita considerando os códigos de acesso, a busca utilizada e as PTMs presentes em cada PrSM.

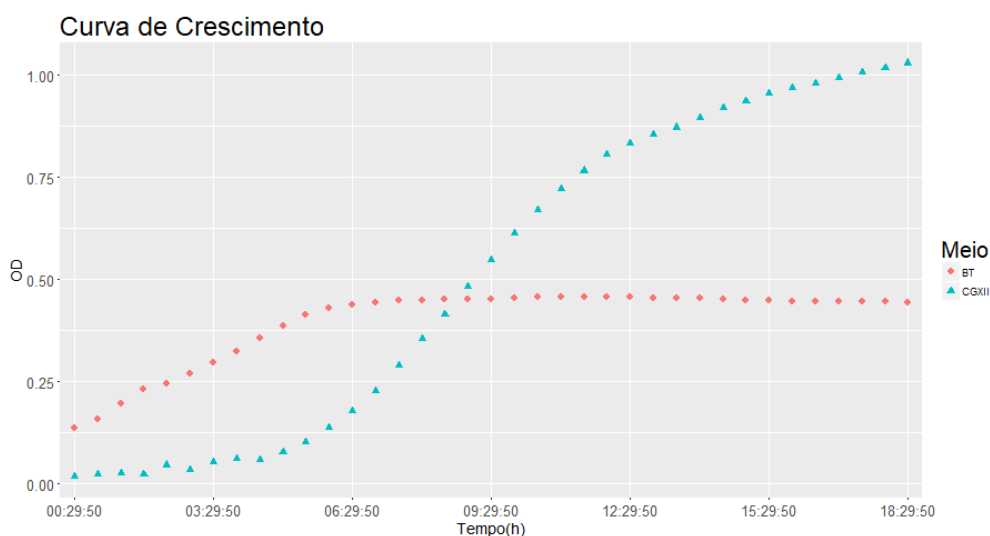
## 5. Resultados e Discussão

### 5.1. Crescimento bacteriano

Analisando as curvas de crescimento foi possível observar maior multiplicação de *C. glutamicum* em meio CGXII, atingido OD próxima a 1.0 após 15 h de cultivo, enquanto em meio BT a OD máxima atingida não passou de 0.5, alcançada com 6 h de cultivo. Isso pode ter ocorrido devido a CGXII ser um meio com mais componentes e pelo fato dessa bactéria ser considerada como auxotrófica para a produção de sideróforos, necessitando

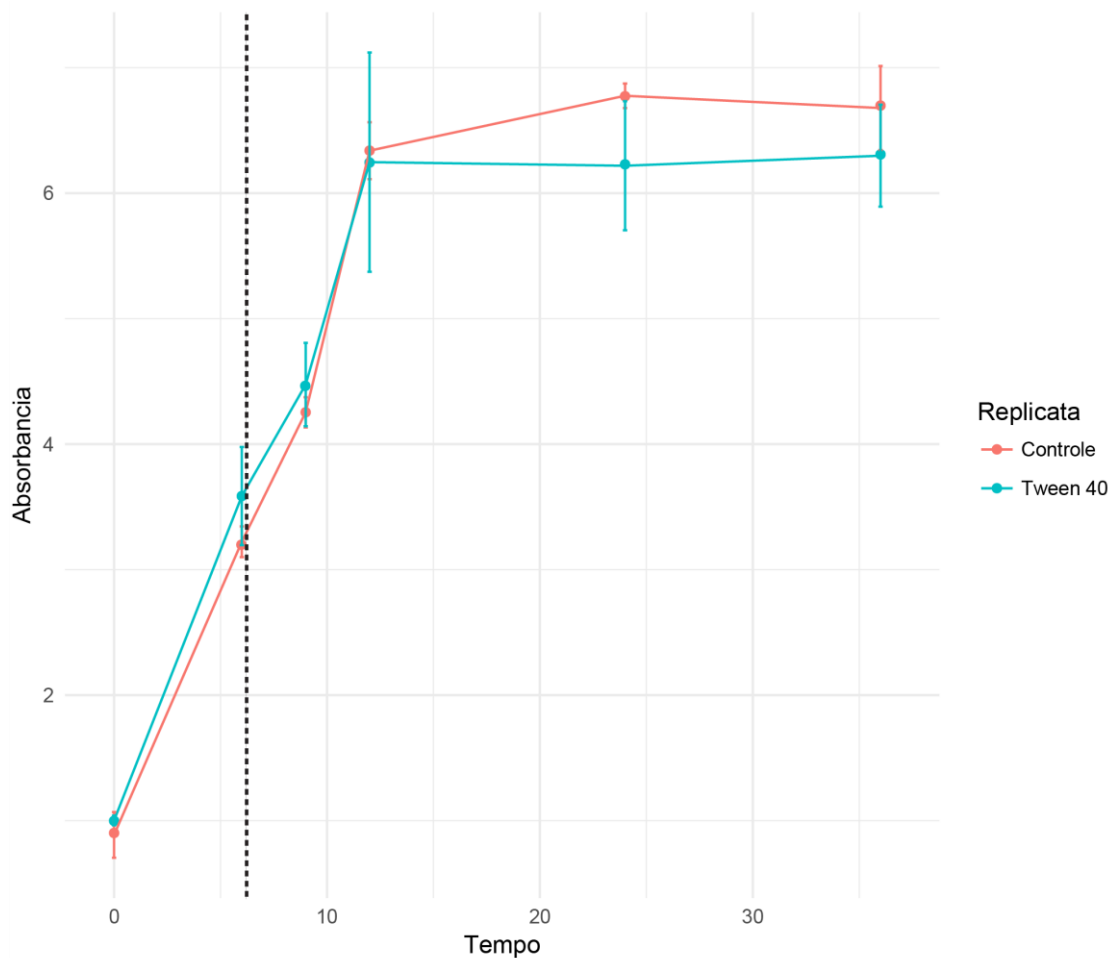
assim de um quelante em seu meio (LIEBL; KLAMER; SCHLEIFER, 1989). No entanto, a fase de latência (*lag*) em CGXII durou mais tempo, o que pode ter sido causado pela etapa de lavagem no processo de inóculo realizado em nesse meio (EGGELING; BOTT, 2005). Essa etapa pode ocasionar estresse do organismo e causar o aumento de fase *lag*.

A partir desses dados, o meio BT foi escolhido para identificação inicial das proteoformas presentes nessa bactéria e para observar qual seria o provável espectro de cobertura da técnica de proteômica *top-down* para identificar proteínas em vias metabólicas, enquanto o meio CGXII foi selecionado para realizar os experimentos em condições de estímulo à produção de aminoácidos.



**Figura 14. Curvas de crescimento da bactéria *C. glutamicum* em meio BT e CGXII.** *C. glutamicum* em meio BT, representado por círculos vermelhos, e em meio CGXII, representado por triângulos azuis.

Para análise da produção de aminoácidos o meio escolhido foi o CGXII por alcançar maior crescimento, quando comparado ao BT (figura 14). A bactéria alcançou a fase estacionária após 12 h de crescimento em ambas condições quando cultivada em meio CGXII (figura 15). Ao fazer análise estatística não foi possível determinar que a diferença de crescimento entre as duas condições nesse meio foi significativa (ANOVA,  $p = 0,94$ ).

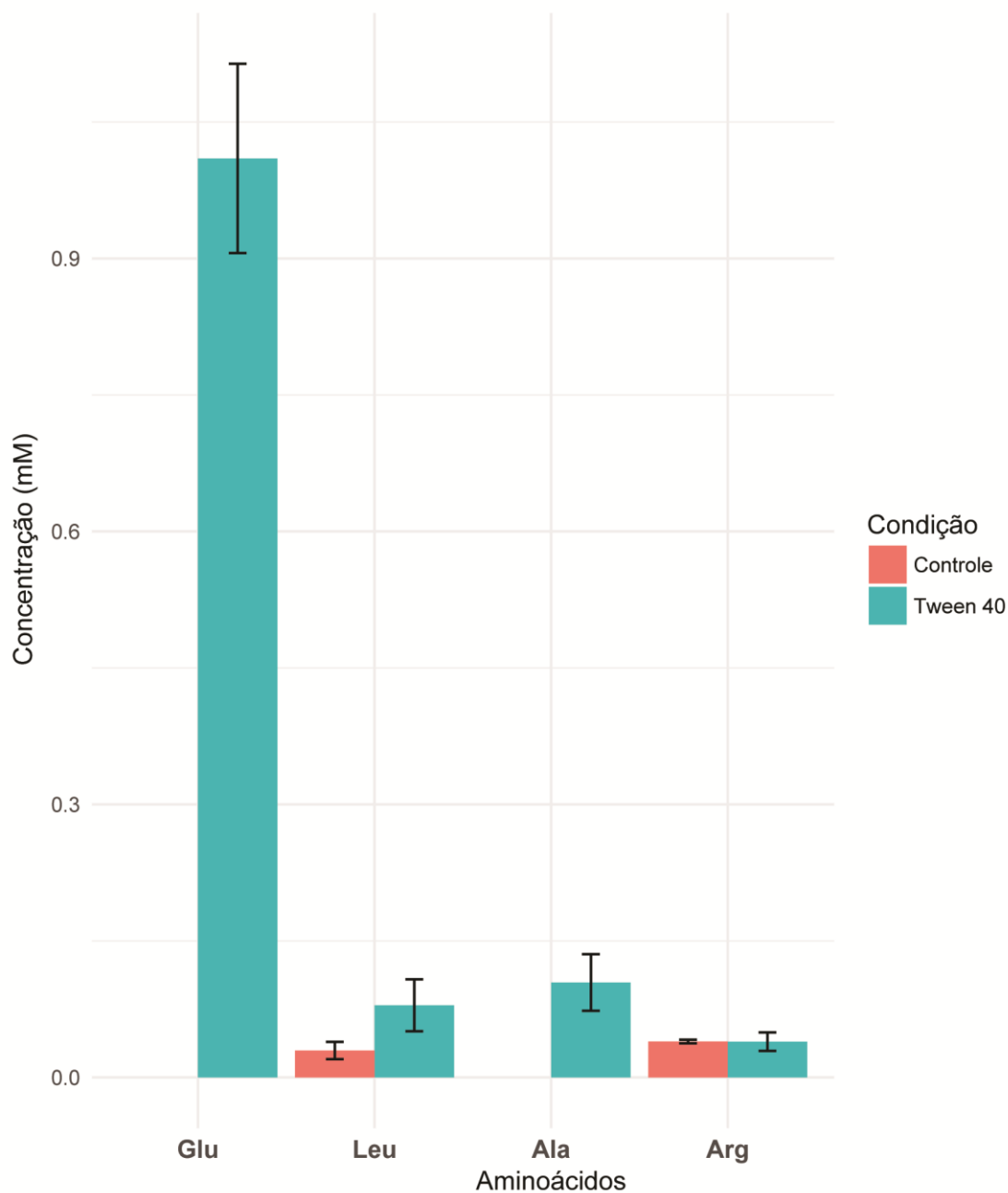


**Figura 15.** Curva de crescimento de *C. glutamicum* em meio CGXII em condição controle e com tween 40. A adição do tween 40 foi feita 6 h após o inóculo, no meio da fase *log*, indicada pela linha tracejada em preto na figura.

## 5.2. Análise de aminoácidos

Não foram detectados aminoácidos em nenhuma das amostras nos tempos 6 h e 12 h. Após 36 h de cultivo foi possível detectar os aminoácidos glutamato, leucina, alanina e arginina. Como esperado, foi produzido mais glutamato sob efeito de tween 40 quando comparada à produção na condição controle. A concentração de glutamato foi de  $1,01 \pm 0,10$  mM na condição tween 40, enquanto não foi detectado na condição controle. Além disso, pode ser observado aumento na produção de alanina em condição com tween 40 ( $0,10 \pm 0,03$ ), quando comparada a condição controle (não detectado). Foi possível observar diferença significativa na concentração de leucina entre as condições com tween 40,  $0,08 \pm 0,03$  mM, e controle,  $0,03 \pm 0,01$  mM (ANOVA,  $p = 0,046$ ). Ademais, arginina

foi detectada nas duas condições, mas sem diferença significativa nas suas concentrações (ANOVA,  $p = 0,98$ ), indicando não ter relação com o estímulo utilizado (figura 16).

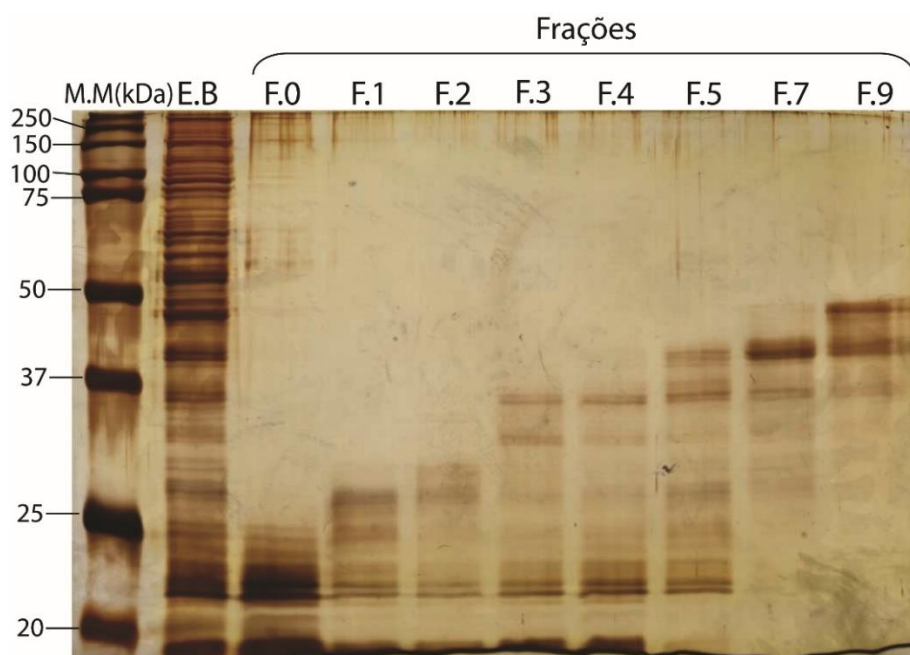


**Figura 16. Concentração de aminoácidos detectados no sobrenadante após 36 h de cultivo em condição controle e com tween 40.** Comparação da quantidade dos aminoácidos detectados entre as condições controle e com tween 40 após 36 h de cultivo. O cálculo da concentração de aminoácido de cada condição foi feito considerando a média das replicatas biológicas. O desvio padrão desses aferimentos está ilustrado pela barra preta.

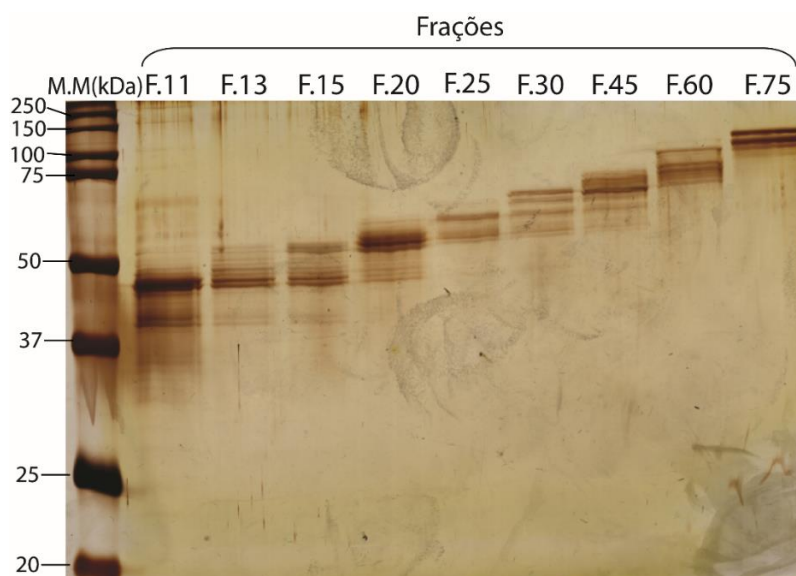
A presença de glutamato somente na condição sob efeito de tween 40 corrobora o funcionamento dessa metodologia para estimular a produção desse aminoácido, como já descrito em outros estudos (EGGELING, 2005; KIM et al., 2009; NIEBISCH et al., 2006). Apesar de ter sido pequena a quantidade de alanina detectada em condição com tween 40, foi o suficiente para diferenciá-la da condição controle. O aumento da produção de alanina já foi causado pelo aumento da expressão de genes de enzimas glicolíticas, como *gapA* (codificando gliceraldeído-3-fosfato desidrogenase), *pyk* (codificando piruvato cinase) e *pfk* (codificando fosfoenolpiruvato cinase). Em condições anaeróbias foi possível produzir 2,430 mM de alanina após 48 h de cultivo (YAMAMOTO et al., 2012). A adição do tween 40 causa aumento da abundância relativa de algumas enzimas das vias glicolíticas e na inibição de parte do ciclo do ácido cítrico (MIZUNO et al., 2016). Devido ao aumento das atividades de enzimas glicolíticas durante estímulo com tween 40, esse estimulante para a produção de glutamato também pode ter ocasionado nessa leve diferença na produção de alanina notada após 36 h de cultivo.

### **5.3. Análise proteômica de *C. glutamicum* em meio BT**

As proteínas intracelulares foram quantificadas por Quibit, apresentando concentração de 1,46 µg/µL. Com base nessa quantificação, 500 µg de proteínas foram submetidos à separação por GELFrEE desnaturante, gerando 17 frações de proteínas. 10 µL dessas frações passaram por SDS-PAGE para confirmar seu perfil de massa (Figuras 17 e 18). Por meio da visualização do perfil proteico obtido foi possível determinar a separação de proteínas por massa entre as frações adquiridas do GELFrEE. Frações coletadas até 15 min após a saída da frente de corrida apresentaram, predominantemente, bandas abaixo da faixa de 50 kDa indicada pelo marcador molecular. Por meio dessa estimativa, as frações foram submetidas à LC-MS/MS por dois tipos de aquisição, LH ou HH, de forma que, até a fração 15 a aquisição foi feita por HH e da fração 20 a 75 foi realizada por LH.



**Figura 17. SDS-PAGE 12% das frações de F.0 a F.9 provenientes de GELFrEE e o extrato bruto.** As frações são representadas por “F.x”, onde x é o tempo em minutos após o início da saída do tampão de amostra do GELFrEE. (MM=Marcador molecular; E.B= Extrato Bruto; F.=Fração).



**Figura 18. SDS-PAGE 12% realizado com as frações de F.11 a F.75 provenientes de GELFrEE.** As frações são representadas por “F.x”, onde x é o tempo em minutos após o início da saída do tampão de amostra do GELFrEE. (MM=Marcador molecular; E.B= Extrato Bruto; F.=Fração).

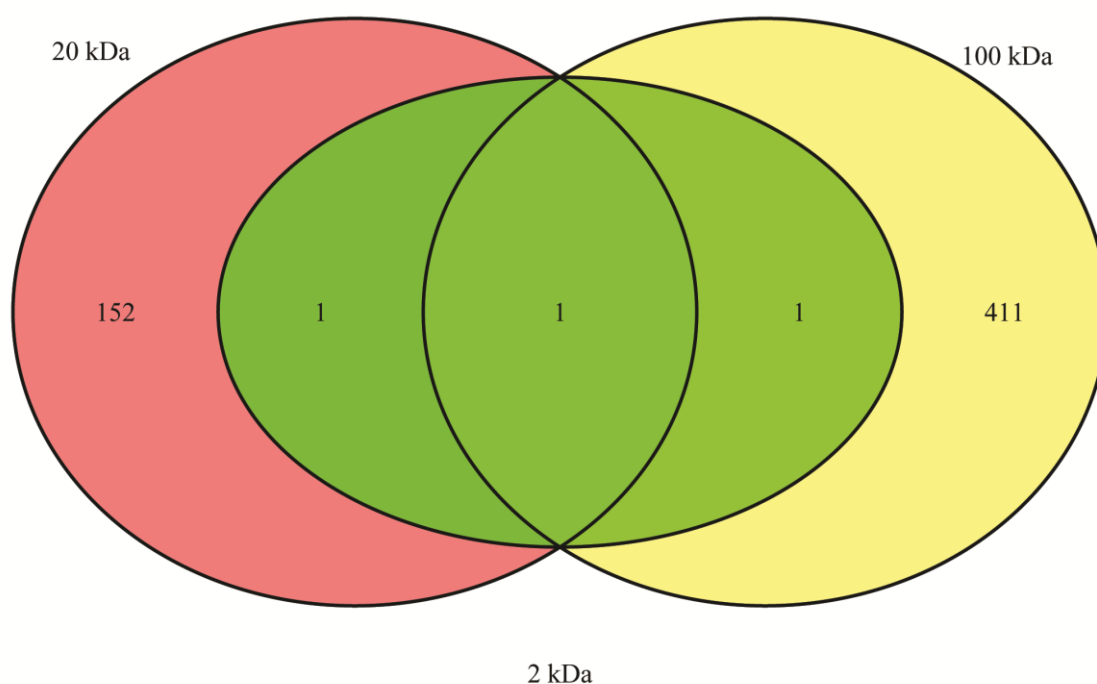
Os resultados gerados pelo espectrômetro de massa foram analisados pelo software ProSight PC 4.0 e os íons com sua relação  $m/z$  foram identificados como proteínas e fragmentos de proteínas.

A formilação foi selecionada como possível PTM devido à maquinaria de tradução clássica de procariotos a qual promove normalmente essa modificação, e em pesquisas recentes que indicam a função dessa modificação como sinal de degradação de proteínas (PIATKOV et al., 2015). Apesar da acetilação no N-terminal (N<sup>α</sup>-acetilação) ser mais comum em eucariotos, devido a sua maquinaria de tradução, ela também está presente em procariotos (CAIN; SOLIS; CORDWELL, 2014). Além disso, recentemente foram descobertas evidências do papel da acetilação não enzimática no metabolismo de *E. coli* (VERDIN; OTT, 2013). Tendo em vista essas características, a N<sup>α</sup>-acetilação também foi considerada na construção do nosso banco de dados.

Em relação às diferentes tolerâncias de precursor utilizadas para HH foi observada grande colaboração nas buscas com alta tolerância (20 kDa e 100 kDa) para identificação de acessos de proteínas não redundantes (figura 19). Foi observado que 85 acessos das aquisições HH só foram identificados em buscas com tolerâncias alta, tal evento pode ocorrer devido a presença de várias PTMs, conseqüentemente proteoformas que geram diferença entre a massa teórica e a massa de precursor observada, afetada por essas possíveis PTMs. Por exemplo, podem ser identificadas proteoformas com glicosilações, capazes de causar grandes diferenças entre a massa do precursor teórico e a massa do precursor observado.







**Figura 20. Digrama de Venn representando os acessos únicos identificados e suas relações com as tolerâncias de precursor utilizada nas buscas por AM adquiridas por LH.** As identificações foram feitas a partir de proteínas provenientes de cultivo em meio BT. As circunferências representam as diferentes tolerâncias de identificação do precursor utilizadas na ferramenta de busca AM, sendo a sobreposição acessos que foram identificados utilizando mais de uma tolerância e as partes sem sobreposição os acessos que foram identificados somente em uma das tolerâncias. As tolerâncias de precursor utilizadas foram: 2 kDa (verde), 20 kDa (vermelho) e 100 kDa (amarelo).

A diferença na quantidade de acessos identificados por cada diferente nível de tolerância de identificação do precursor é relevante, considerando que ela indica o estado de massa da proteoforma identificada. Por exemplo, se ela está com alguma modificação que adiciona massa a sua sequência, essa proteoforma terá a massa do precursor observado maior do que a do precursor teórico, sendo identificadas nas buscas com tolerâncias maiores do que 2,3 Da. Outra possibilidade é a proteoforma ter perdido um fragmento de sua sequência, possuindo assim um precursor observado com massa menor do que o teórico, tal evento pode ser identificado tanto pela busca por BM, quanto por

buscas em AM com tolerâncias maiores do que 2,3 Da, nesse caso sendo identificadas proteoformas com clivagens não preditas pelo banco de dados.

Apesar da grande sobreposição de identificações em vários níveis de tolerâncias, o que foi levado em consideração para construir esses diagramas de Venn foram somente os dados de acessos identificados. Logo um mesmo acesso pode corresponder a diferentes proteoformas. Isso pode ocorrer principalmente no caso de acesso identificado em vários níveis de tolerâncias, por exemplo: um acesso é identificado a partir busca AM em duas detecções em MS, uma foi identificada com tolerância de 2,3 Da e outra com tolerância de 2 kDa, ao ser observada a massa do precursor desses dois casos o primeiro vai ser a proteoforma intacta desse acesso, enquanto o segundo provavelmente será uma proteoforma com modificação de até 2 kDa, como uma clivagem ou adição de grupamentos.

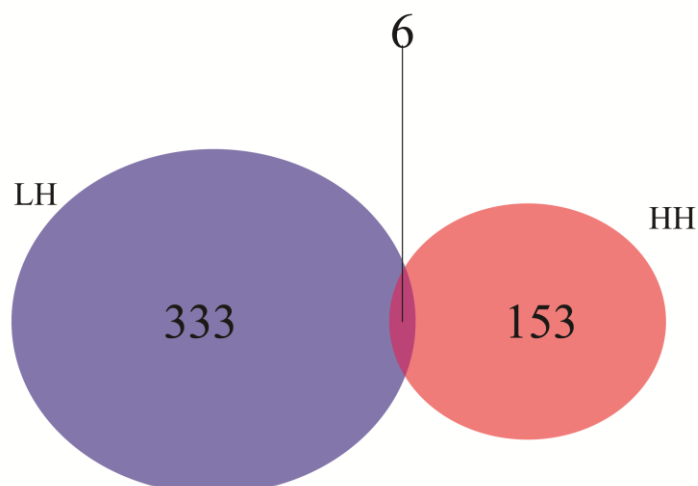
Para utilização dos dados de todos os tipos de busca ao mesmo tempo foi necessário remover a redundância de eventos entre elas, por exemplo: várias proteínas ou proteoformas sendo identificadas a partir de um mesmo evento de detecção. Essa redundância ocorreu justamente por serem feitas várias buscas (material suplementar: CG\_13032\_BTmediu\_DB\_M+F+A+allPTMs, GFC\_36.Rmd e GFT\_36.Rmd).

Além da remoção de eventos de detecção redundantes, as identificações passaram por filtros estatísticos, gerando assim 1873 PrSMs, total de espectros identificados com validação estatística. Com base nesses espectros foram contabilizados os acessos únicos identificados por cada tipo de busca (AM ou BM) e proteoformas únicas identificadas automaticamente (tabela 4). Para contabilizar as proteoformas foram considerados os dados de PTMs, acesso e tipo de busca anotados automaticamente pelo ProSightPC 4.0.

**Tabela 4. Quantidade de PrSMs detectados, proteínas não redundantes e proteoformas não redundantes identificadas.** Identificações feitas a partir de detecções de proteínas provenientes de cultivo em meio BT. Contagem feita de acordo com o tipo de busca (AM ou BM) para PrSMs e proteínas não redundantes. Para contabilizar proteoformas não foi possível separar as buscas, pois o tipo de busca BM é utilizado para categorizar proteoformas truncadas.

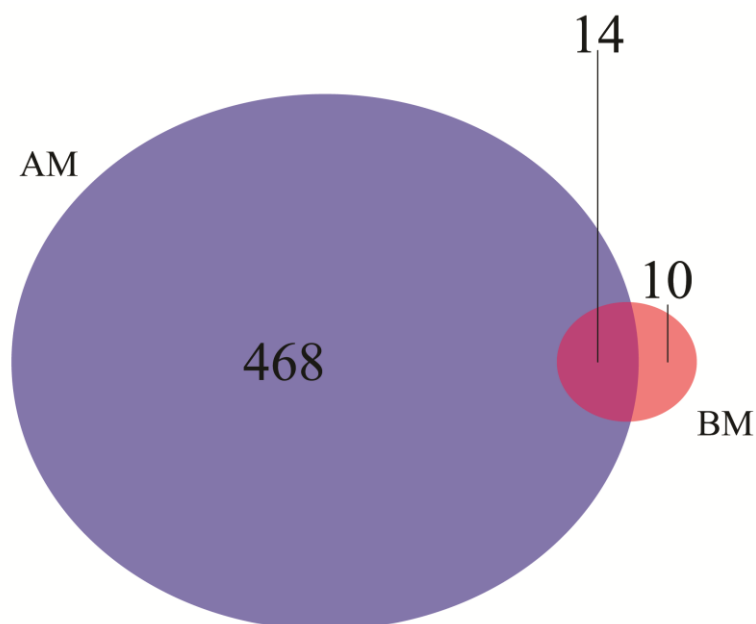
Busca	PrSMs	Proteínas	Proteoformas
AM	1826	482	X
BM	47	24	X
Total	1873	492	681

Também foi analisada a contribuição de cada tipo de aquisição para as identificações de acessos proteicos não redundantes (figura 21). Foi possível observar a grande diferença nas identificações entre os modos, como era esperado devido à característica do fracionamento das amostras, que separa as proteínas de acordo com a massa. Consequentemente, as amostras submetidas à LH devem possuir proteínas diferentes das submetidas à HH. A quantidade de proteínas não redundantes identificadas somente por LH demonstrou grande potencial desse tipo de aquisição para caracterização de *C. glutamicum*.



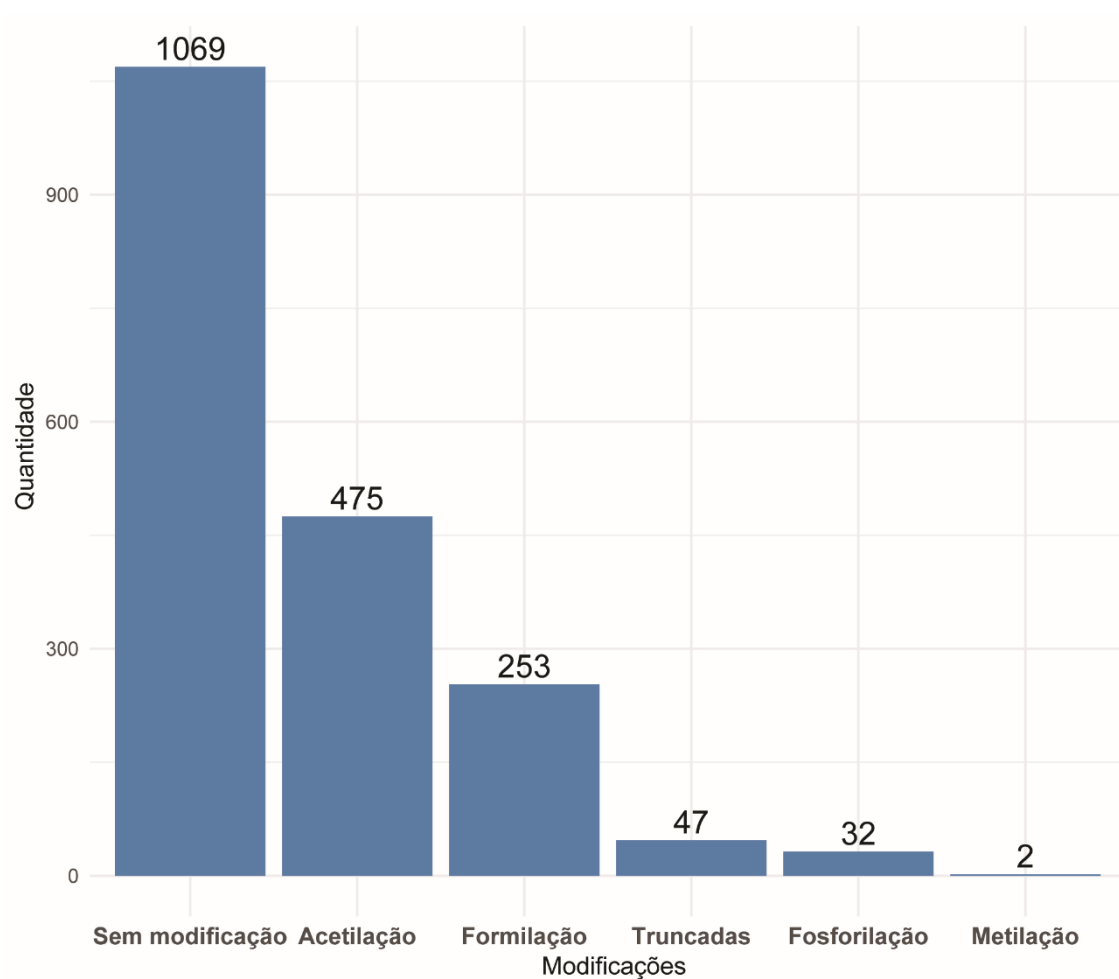
**Figura 21. Quantidade de acessos proteicos não redundantes identificados em cada tipo de aquisição.**  
*low-high = LH (azul) e high-high = HH (vermelho).*

Para fazer a análise de cada tipo de busca (AM e BM), as identificações de todas as frações foram separadas de acordo com a ferramenta de busca utilizada em cada identificação e cada grupo teve suas proteínas redundantes removidas, separadamente. Sendo assim, foram identificados 482 acessos não redundantes por AM e 24 por BM (tabela 4). Os dados das diferentes buscas podem parecer redundantes devido à grande sobreposição observada no gráfico de Venn, 14 dos 24 acessos de BM (figura 22), no entanto, para critério de proteoformas, as identificações por BM são complementares às de AM, já que cada uma delas apresenta uma proteoforma clivada para as proteínas identificadas por AM. Ambos métodos de buscas foram utilizados com o intuito de aumentar a quantidade de proteoformas identificadas, considerando a complementariedade de seus mecanismos.



**Figura 22. Diagrama de Venn com a quantidade de acessos não redundantes, para cada tipo de busca.** Os acessos identificados por cada tipo de busca são representados por círculos de diferentes cores, sendo: AM = *Absolute mass* (azul), e BM= *Biomarker* (vermelho). As áreas formadas por sobreposição de circunferências representam acessos identificados pelos dois métodos.

Os 1873 PrSMs detectados foram analisados quanto aos tipos de PTMs presentes, e quantidade de detecções de cada tipo. Os espectros foram quantificados de acordo com tipos de modificação detectadas, obteve-se proteoformas referentes a sequências sem modificação, sequências com acetilação, formilação, fosforilação e metilação. As proteoformas com maior quantidade de PrSMs foram as sequências sem modificações, 1069 PrSMs, a acetilação foi encontrada como a modificação mais detectada nessas proteoformas, com 475 PrSMs detectados, seguida da formilação, com 253 PrSMs. Modificações difíceis de serem encontradas em estratégia *bottom-up* como as proteínas truncadas também foram detectadas, 47 PrSMs (figura 23).

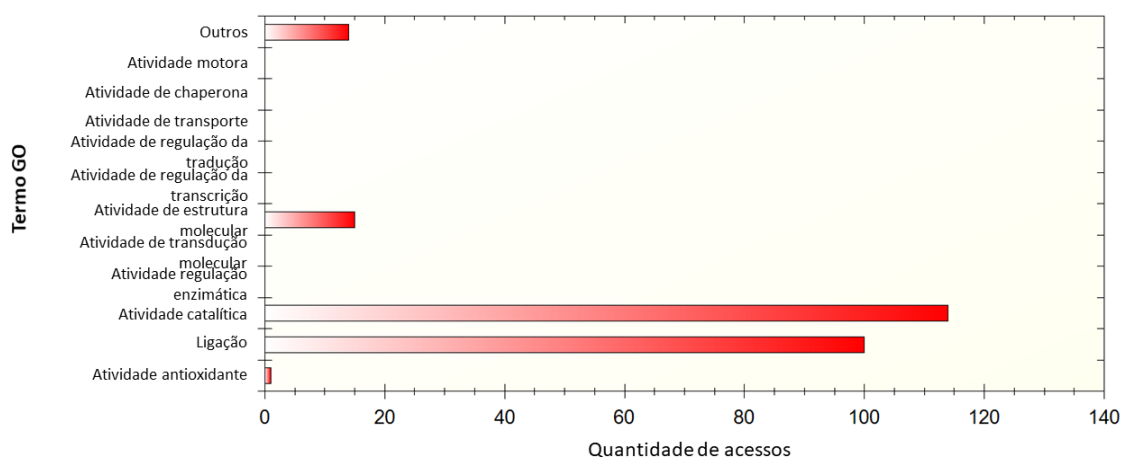


**Figura 23. Tipos e quantidade de modificações identificadas em todos os PrSMs provenientes de cultivo em meio BT.** Só foram consideradas as modificações anotadas automaticamente pelo software ProSightPC 4.0 (Thermo Scientific™).

A acetilação recentemente foi relacionada como possível mecanismo para regulação da produção de aminoácidos sob estimulação por tween 40 (MIZUNO et al., 2016). Mizuno et al. 2016 relacionou a redução de acetilações e aumento de succinilações em resíduos de lisina com a adição de tween 40 no meio de cultura e consequentemente aumento na produção de glutamato. As acetilações detectadas no presente estudo não são resíduo específicas por serem delimitadas apenas ao N-terminal (N<sup>α</sup>-acetilação).

A N<sup>α</sup>-acetilação foi descrita como pouco presente em procariotos, sendo identificada em proteínas ribossomais, fatores de virulências e associados a algumas lipoproteínas de bactérias gram-positivas (CAIN; SOLIS; CORDWELL, 2014). Em contraste com a literatura apresentada, foram identificadas 223 proteínas diferentes acetiladas, a maioria delas sendo relacionadas a atividades catalíticas e de ligação (figura

24). Já foram apresentados vários estudos relacionando acetilações de lisina a regulações metabólicas em procariotos (CAIN; SOLIS; CORDWELL, 2014), no entanto até então a N<sup>α</sup>-acetilação é pouco registrada nesses organismos, mesmo para essa função (WEINERT et al., 2013). Uma hipótese para essa subrepresentação é a predominância de abordagem proteômica *bottom-up* na análise de acetilações, o que dificulta a identificação de N-terminais proteicos.



**Figura 24. Quantidade de acessos anotados para *gene ontologies* (GO) de função molecular das proteínas identificadas com N<sup>α</sup>-acetilação.** Dados obtidos a partir de proteínas de *C. glutamicum* cultivado em meio BT e análise de GO realizada em STRAP.

Foram detectados 253 PrSMs formilados, correspondendo a 134 acessos únicos. Essa modificação foi apresentada recentemente como possível mecanismo de degradação de proteínas em procariotos (PIATKOV et al., 2015). A grande quantidade de identificações com formilações e acetilações demonstra o potencial da abordagem proteômica *top-down* para estudo da função dessas PTMs no metabolismo de *C. glutamicum*.

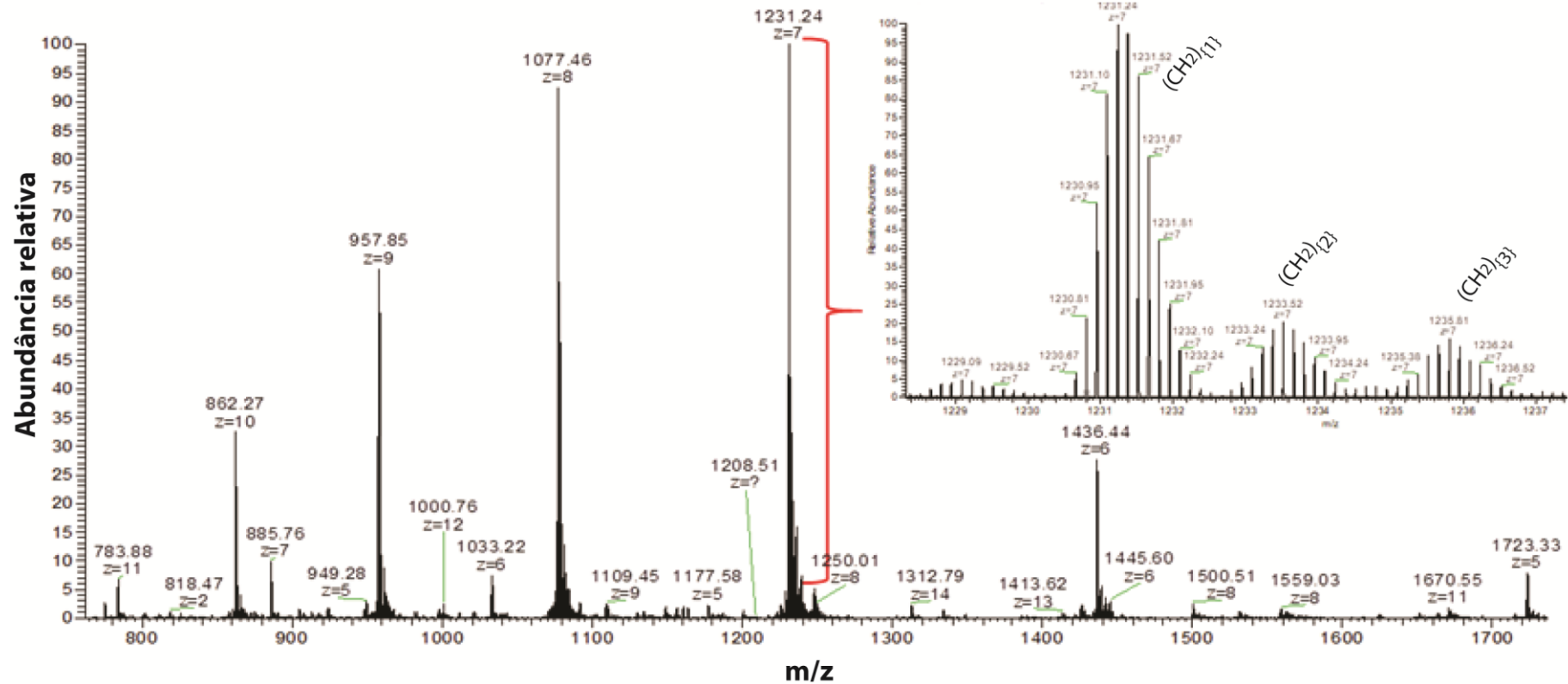
Em um primeiro momento de observação dos padrões de variação de massa diferença entre a massa do precursor teórico e o precursor observado (P.T.-P.O.), duas proteínas foram detectadas com variações significantes, a Subunidade PurS da Fosforibosil formilglicinamidina sintase e uma chaperonina de 10 kDa (tabela 5).



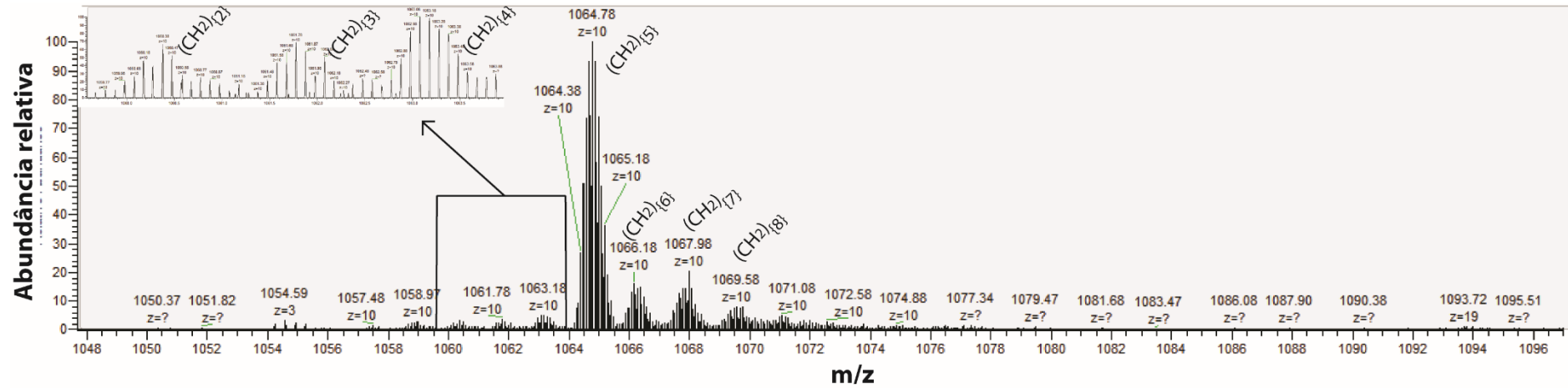
**Tabela 5. PrSMs de Q8NMI3 e Q8NSS1 identificados com variação de massa maior ou igual a 14 Da.** P.T.-P.O. = diferença entre a massa do precursor teórico e o precursor observado.

Acesso	Descrição	P.T.-P.O.	E-value
Q8NMI3	Subunidade PurS da Fosforibosil formilglicinamidina sintase	20,98	2,00E-25
Q8NMI3	Subunidade PurS da Fosforibosil formilglicinamidina sintase	23,95	9,53E-17
Q8NSS1	10 kDa chaperonina	-5501,02	6,13E-06
Q8NSS1	10 kDa chaperonina	14,00	2,77E-28
Q8NSS1	10 kDa chaperonina	15,00	2,92E-26
Q8NSS1	10 kDa chaperonina	30,99	1,34E-27
Q8NSS1	10 kDa chaperonina	49,00	2,38E-37

Com base nessas variações de massas detectadas os espectros referentes a essas proteínas foram analisados e foi possível detectar envelopes isotópicos de carga semelhante a observada nos precursores responsáveis pela identificação dessas proteínas e massa entre esses envelopes foi calculada, correspondendo a 14 Da, sugerindo presença de vários níveis de metilação em ambas proteínas (figuras 25 e 26). Por meio da análise dos espectros é possível identificar 4 possíveis proteoformas com metilações para a proteína Q8NMI3, duas fragmentadas e identificadas (figura 25). Enquanto para a proteína Q8NSS1 foi possível detectar 9 possíveis proteoformas com diferentes níveis de metilação (figura 26).



**Figura 25. Espectro de massa MS1 apresentando dois precursores responsáveis pela identificação de Q8NMI3 e supostas metilações.** Envelope da proteína pode ser observado em espectro maior e a distribuição de envelopes isotópicos pode ser observada pela aproximação indicada pela chave vermelha. Os níveis das metilações sugeridas estão indicados entre chaves ao lado dos envelopes isotópicos. Os envelopes que geram as identificações de Q8NMI3 são os com 1 e 3 metilações.



**Figura 26. Espectro de massa MS1 apresentando dois precursores responsáveis pela identificação de Q8NSS1 e supostas metilações.** A distribuição de envelopes isotópicos pode ser observada em todos os envelopes com cargas iguais. Os níveis das metilações sugeridas estão indicados entre chaves ao lado dos envelopes isotópicos. Os envelopes que geraram as identificações de Q8NSS1 são os apresentados com 2 e 8 metilações.

As sequências das 492 proteínas não redundantes foram obtidas pelo UniprotKB (<https://www.uniprot.org/uniprot/>) e analisadas com o software BlastKOALA (<http://www.kegg.jp/blastkoala/>). 308 proteínas encontraram relação com algum *kegg orthology* (KO) e foram mapeadas de acordo com sua função ou inseridas diretamente em vias metabólicas.

As vias centrais de metabolismo dos carbonos, pentose fosfato (PPP) e ciclo do ácido cítrico são descritas como parte relevante da produção de aminoácidos em *C. glutamicum* (GEORGI; RITTMANN; WENDISCH, 2005; MIZUNO et al., 2016), por esse motivo, as enzimas encontradas nessas vias foram escolhidas para serem analisadas e anotadas de forma manual em relação às suas proteoformas, juntamente às proteínas mapeadas na biossíntese de aminoácidos, formando assim o grupo que será chamado daqui em diante de proteínas relacionadas ao metabolismo de aminoácidos. Nesse grupo foram identificados 23 acessos de proteínas para a biossíntese de aminoácidos, 10 na glicólise/gliconeogênese, 9 no ciclo do ácido cítrico e 4 na PPP. A partir desses mapeamentos em vias metabólicas, foram escolhidas proteínas para serem anotadas e os PrSMs detectados para essas proteínas foram observados. Dessa forma foram encontradas no total 34 proteínas não redundantes (tabela 6) e alguns PrSMs detectados para essas proteínas foram observados quanto à variação na massa entre precursor teórico e observado. Além disso, os espectros dessas proteínas foram analisados em busca de proteoformas que tenham sido detectadas, mas não identificadas pelo ProsightPC 4.0.

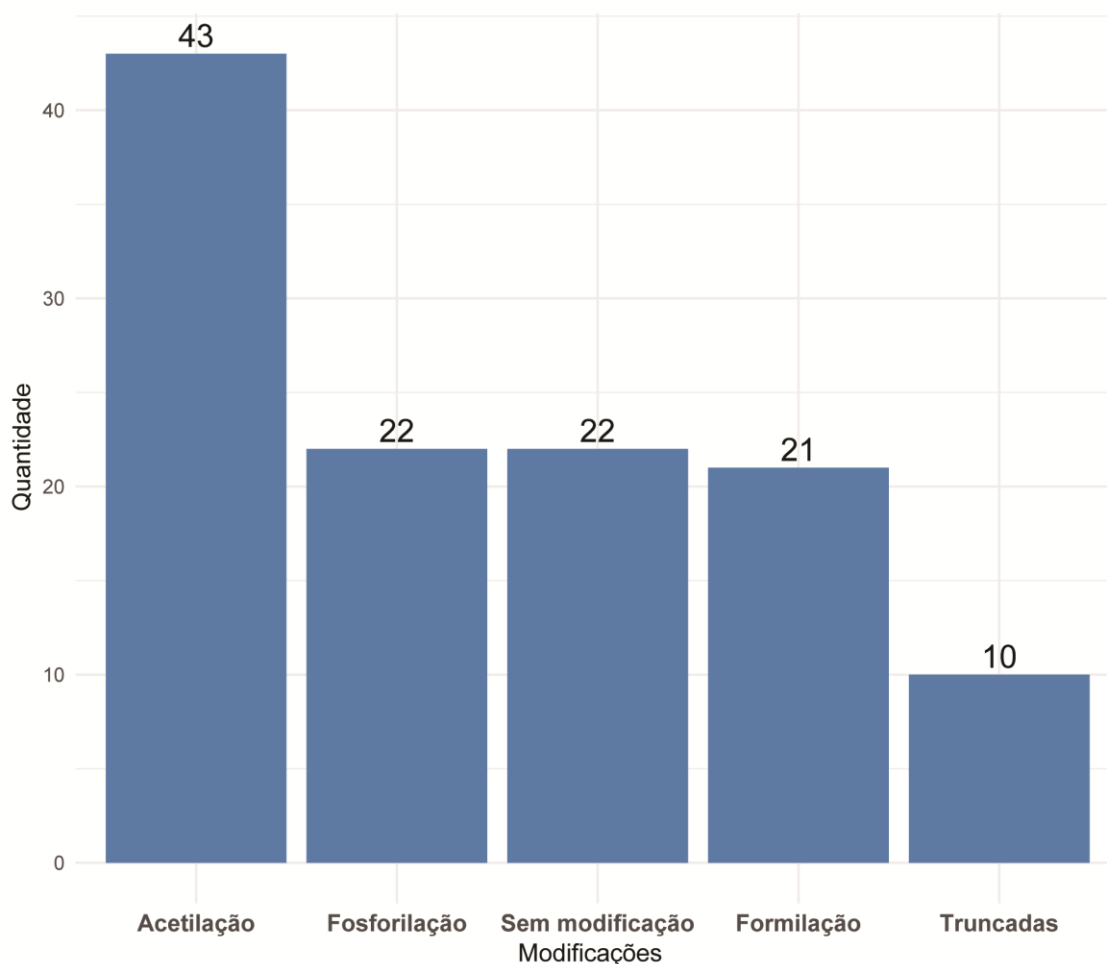
**Tabela 6. Códigos de acessos detectados a partir de cultivo em meio BT com função anotadas dentro das vias de biossíntese de aminoácidos (AA), ciclo do ácido cítrico (TCA), glicólise/gliconeogênese (GLI) e via das pentoses fosfatos (PPP). São apresentadas as vias em que cada acesso fora anotado, a descrição dessas proteínas de acordo com o ProSightPC 4.0 e o tipo de busca utilizado para identificar as mesmas.**

Tipo de busca	Código de acesso	Descrição da proteína	E-value	Vias
AM	O31139	Subunidade HisF da imidazol glicerol fosfato sintase	5,50E-18	AA
AM	O69282	Malato:quinona oxidoreductase	2,53E-05	TCA
AM	P07128	Homoserina cinase	1,15E-07	AA
AM	POC1E4	Pirrolina-5-carboxilato reductase	3,40E-07	AA
AM	P19537	Frutose-bifosfato aldolase	3,31E-19	GLI, AA, PPP
AM	P23669	Treonina sintase	1,38E-06	AA
AM	P40110	4-hydroxi-tetrahidrodipicolinate reductase	1,08E-06	AA
AM	P42457	Citrato sintase	2,46E-20	TCA, AA
AM	Q01651	Gliceraldeído-3-fosfato desidrogenase	1,29E-11	GLI, AA
AM	Q01655	Fosfoglicerato cinase	1,29E-09	GLI, AA
BM	Q46078	Piruvato cinase	3,82E-10	GLI, AA
AM	Q79VE3	Glutamina sintase	1,77E-06	AA
AM	Q8NLD4	Proteína não caracterizada	1,66E-05	PPP
AM	Q8NLN0	L-lactato desidrogenase	2,95E-20	GLI
AM	Q8NMK4	Proteína não caracterizada	1,95E-07	TCA
AM	Q8NMK7	Subunidade beta da Succinato--CoA ligase [formando ADP]	4,21E-09	TCA
AM	Q8NML1	Cisteina sintase	3,02E-16	AA
AM	Q8NNF6	Componente E1 da piruvato desidrogenase	2,02E-05	GLI, TCA
AM	Q8NNJ2	Componente Dihidrolipoilisina-resíduo acetiltransferase do complexo da piruvato desidrogenase	2,88E-05	TCA
AM	Q8NKK1	Proteína não caracterizada	4,05E-08	AA
AM	Q8NNT5	Histidinol desidrogenase	6,53E-07	AA
BM	Q8NPA4	Proteína não caracterizada	4,44E-12	GLI
AM	Q8NQ49	Ribulose-fosfato 3-epimerase	1,75E-16	AA, PPP

Continuação tabela 6.

Tipo de busca	Código de acesso	Descrição da proteína	E-value	Vias
AM	Q8NQ98	Aconitato hidratase A	2,84E-07	GLI, TCA
AM	Q8NQD1	Metionina sintase	5,22E-07	AA
AM	Q8NRB3	5-metiltetrahydropteroltriglutamato--homocisteína metiltransferase	1,15E-07	AA
AM	Q8NRS1	Enolase	2,38E-29	GLI, AA
AM	Q8NS31	Glicose-6-fosfato isomerase	1,90E-07	GLI, PPP
AM	Q8NTA5	2, 3-bifosfoglicerato-dependente fosfoglicerato mutase	3,21E-16	GLI, AA
AM	Q8NTD6	Proteína não caracterizada	6,96E-08	TCA
AM	Q8NTD7	Proteína não caracterizada	2,92E-06	TCA
AM	Q8NTS6	Proteína não caracterizada	1,96E-10	AA
AM	Q8NTW7	Proteína não caracterizada	1,70E-05	AA
AM	Q9X5D2	3-dehidroquinato sintase	2,53E-05	AA

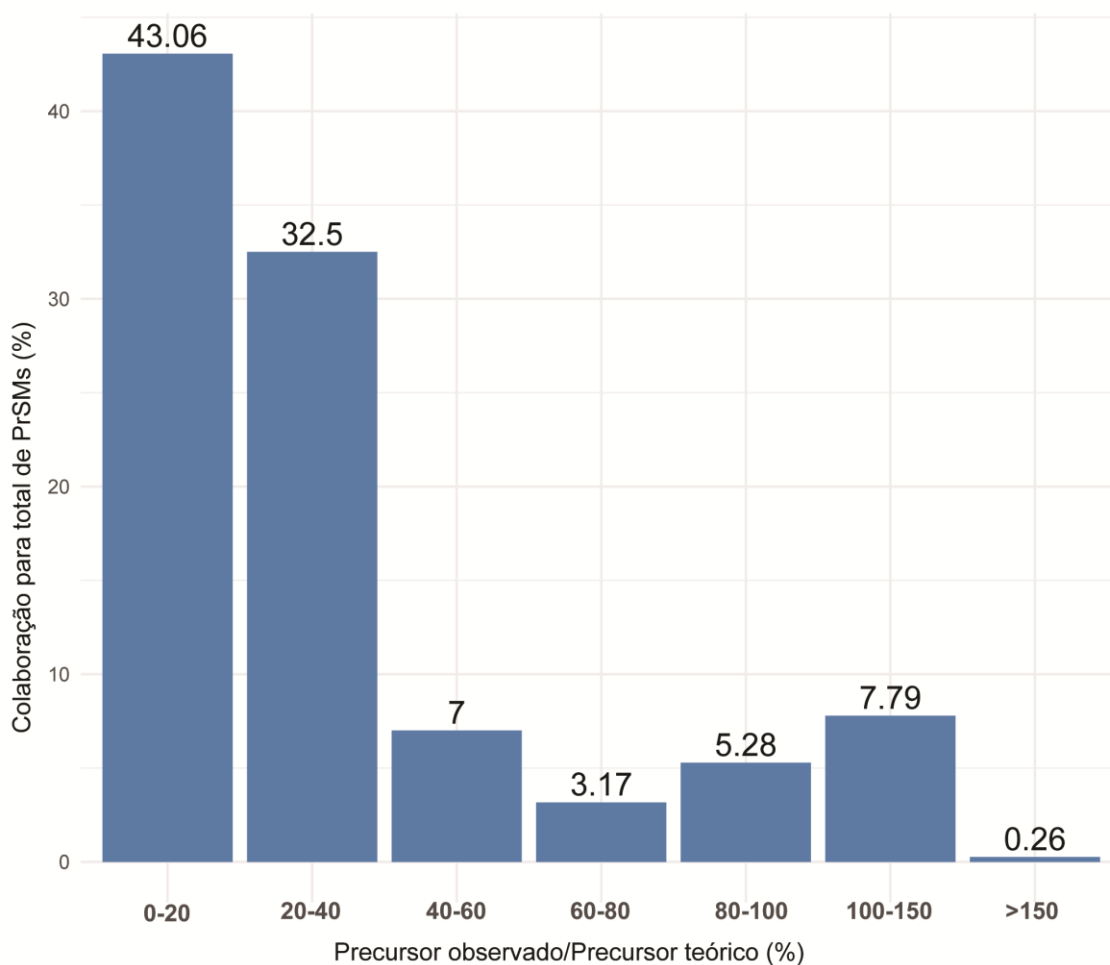
Foram detectados 154 PrSMs referentes às proteínas classificadas como relacionadas ao metabolismo de aminoácidos, como descrito acima. Foram identificadas proteoformas referentes a acetilações, formilações e grande parte das fosforilações presentes no total de proteínas, indicando alguma função específica para esse grupo de proteínas (figura 27). Além disso, as proteoformas detectadas com acetilações, 43, superaram as sem modificação, 22, (figura 27), sugerindo enriquecimento dessas PTM nesse grupo.



**Figura 27. Quantidade de PrSMs em relação aos tipos de modificações detectadas e classificadas como referentes ao metabolismo de aminoácidos.**

A observação dos espectros das PrSMs relacionadas ao metabolismo de aminoácidos apresentou vários precursores observados com tamanhos muito menores do

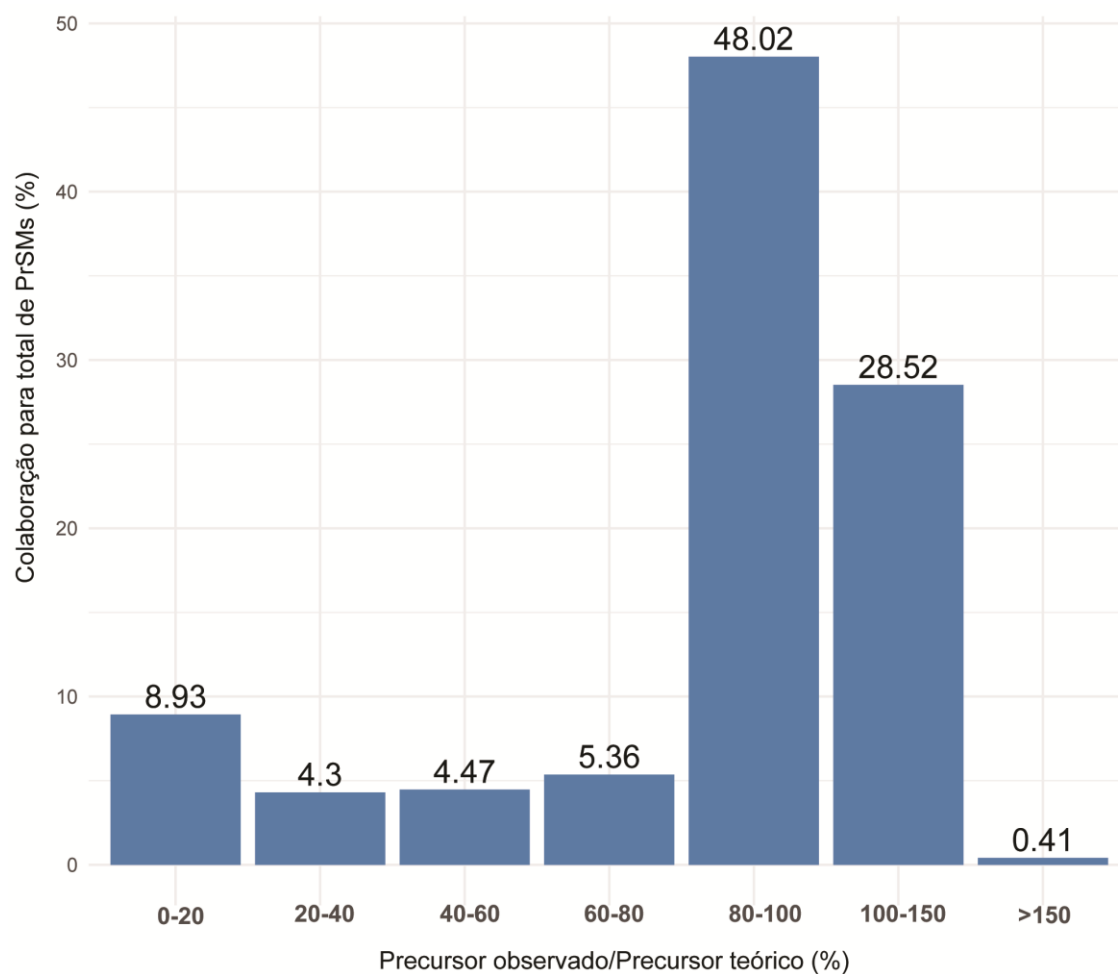
que os teóricos, evento esse melhor observado em um gráfico de barras com a relação tamanho do precursor observado em relação ao teórico e a colaboração desses eventos no total de PrSMs (figura 28). A grande representação de PrSMs com precursores observados correspondendo a 0%-20% do tamanho total do precursor teórico indica a presença de algum tipo de degradação dessas proteínas, onde na verdade estão sendo identificados fragmentos dessas ainda no MS1.



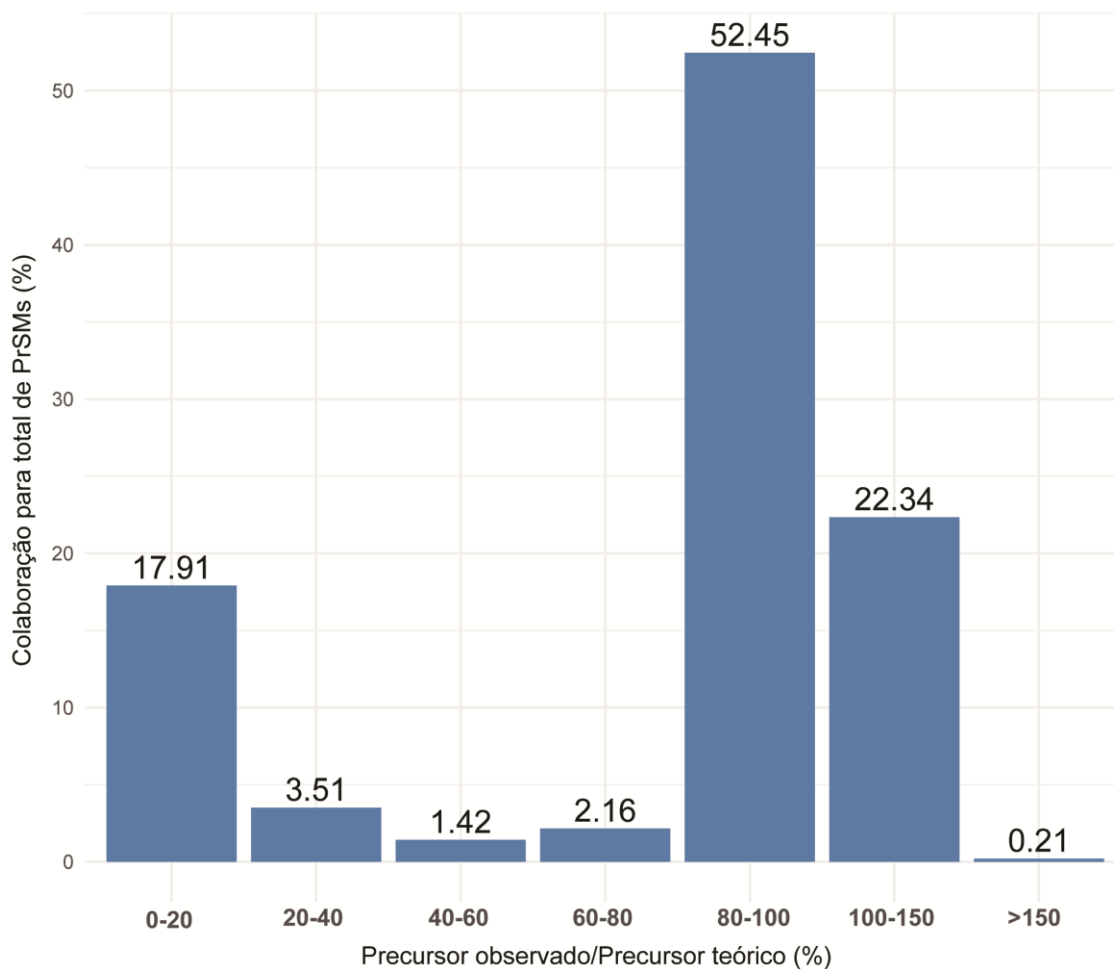
**Figura 28. Relação entre a massa dos precursores observados e dos teóricos de acordo com a quantidade de PrSMs (%) de detecções válidas em meio BT.**

A degradação das proteínas presente em meio BT pode ter sido estimulada por causas naturais, mas o mais provável é que tenha ocorrido no preparo da amostra, já que o padrão de presença de proteínas com relação entre precursor teórico e observado de 0% a 20% não se repetiu nos experimentos que ocorreram em meio CGXII (figuras 29 e 30).





**Figura 29. Relação entre a massa dos precursores observados e dos teóricos de acordo com a quantidade de PrSMs válidas (%) da condição controle cultivada em meio CGXII.**

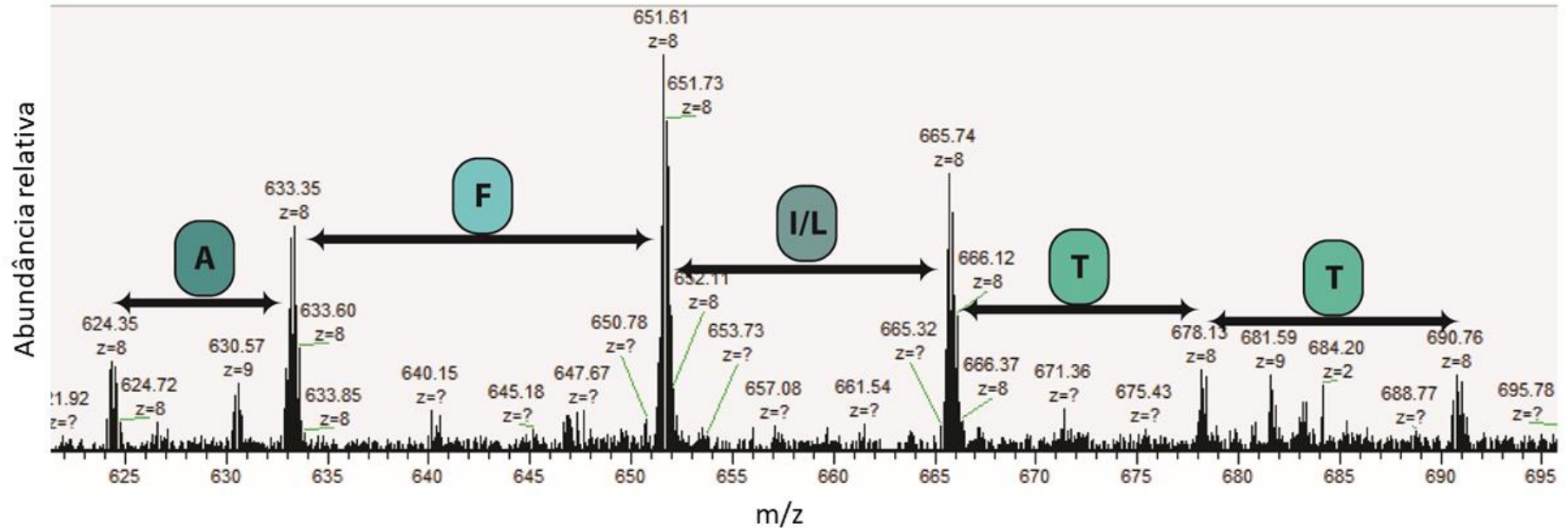


**Figura 30. Relação entre a massa dos precursores observados e dos teóricos de acordo com a quantidade de PrSMs válidas (%) da condição com tween 40 cultivada em meio CGXII.**

Uma possibilidade para a diferença nos padrões de quantidade de PrSMs referentes à relação entre massa do precursor teórico e precursor observado é a presença de SDS em tampão de lise utilizado para romper as células provenientes de meio CGXII. Detergente altamente desnaturante, esse evita a degradação de proteínas por proteases por afetar a estrutura tridimensional de proteínas e conseqüentemente suas atividades (KACHUK; DOUCETTE, 2018).

Além da presença de muitas proteínas com massa do precursor observada muito menor do que a teórica, também foram encontradas nas amostras provenientes de meio BT evidências que sugerem degradação das proteínas a partir da observação de espectros. A proteína de código P42457, citrato sintase, foi identificada em várias frações (9, 11, 13, 15), possuindo identificações com tamanhos observados pequenos em relação aos teóricos, com massas entre 4.027 Da e 6.123 Da. No entanto, a massa de seu precursor

teórico é 48.809 Da. Observando os espectros das PrSMs dessa proteína foi notada a presença de vários envelopes isotópicos com cargas iguais, apresentando diferenças de massas semelhantes à massas de resíduos de aminoácidos (Ala, Phe, Isso/Leu e Thr) (figura 31), sugerindo outro tipo de fragmentação dessa proteína.



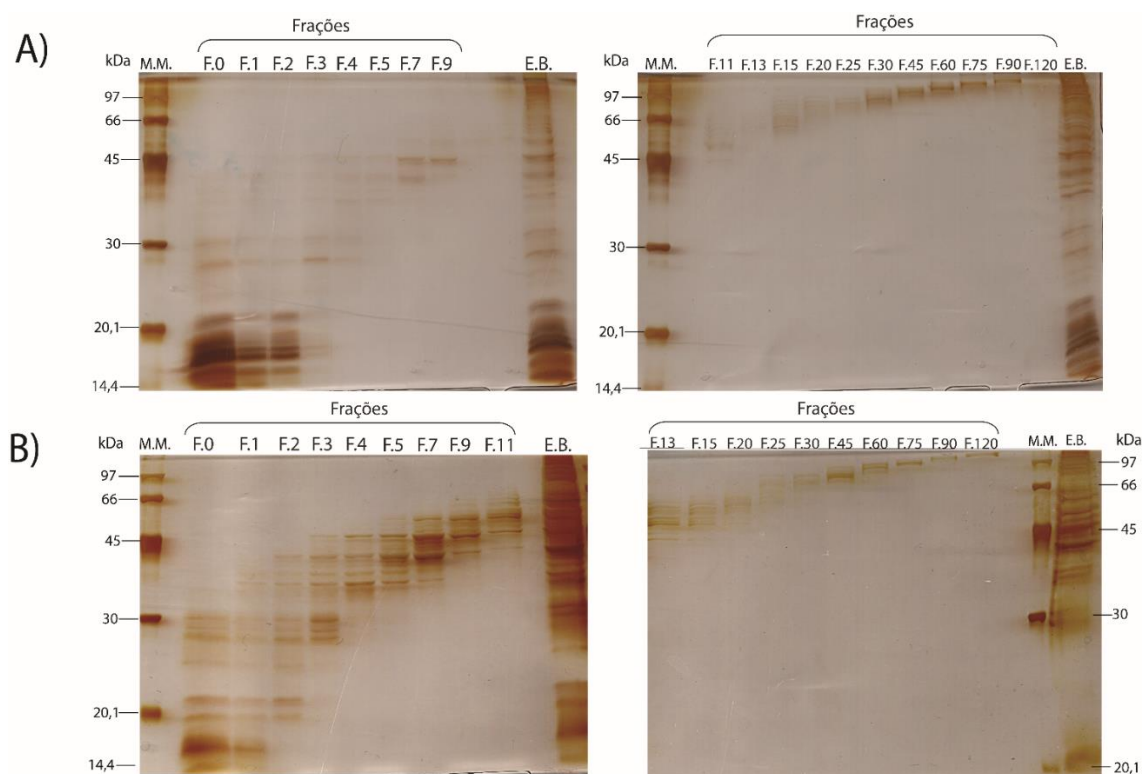
**Figura 31.** Scans referentes à identificação de uma das PrSMs da proteína P42457. É possível observar o m/z e a abundância relativa de vários envelopes isotópicos com cargas iguais ( $z=8$ ). A diferença de massa entre os picos monoisotópicos de cada envelope é representada pelas setas pretas e a essas diferenças foram atribuídos possíveis resíduos de aminoácidos, indicados na caixa logo acima de cada seta.

A metodologia adotada aqui por meio de proteômica *top-down* e contagem de espectros levando em consideração a massa de seu precursor teórico, apresenta potencial para avaliar a degradação de proteínas em amostras, sejam elas por causas biológicas ou durante o preparo para espectrometria de massa. No entanto, testes com outras abordagens experimentais devem ser realizados para confirmar essa hipótese.

Devido as evidências apresentadas de degradação das proteínas presentes na amostra proveniente do cultivo em meio BT, a anotação manual desse experimento foi abandonada para dar lugar ao experimento em meio CGXII.

#### **5.4. Análise proteômica de *C. glutamicum* em meio CGXII**

As replicatas biológicas, considerando as duas condições (controle ou com tween 40), foram normalizadas de acordo com a concentração de proteínas e unidas, formando um *pool* para cada condição. Os dois *pools* foram submetidos ao fracionamento por GELFrEE. Por meio de SDS-PAGE das frações obtidas foi possível determinar que as frações de 0 a 11 possuem padrão de massa predominantemente abaixo de 50 kDa, essas foram então adquiridas por HH e as de 13 a 120 por LH. Apesar de ser feita a aquisição dos dados de LH, esses não foram considerados para identificação.



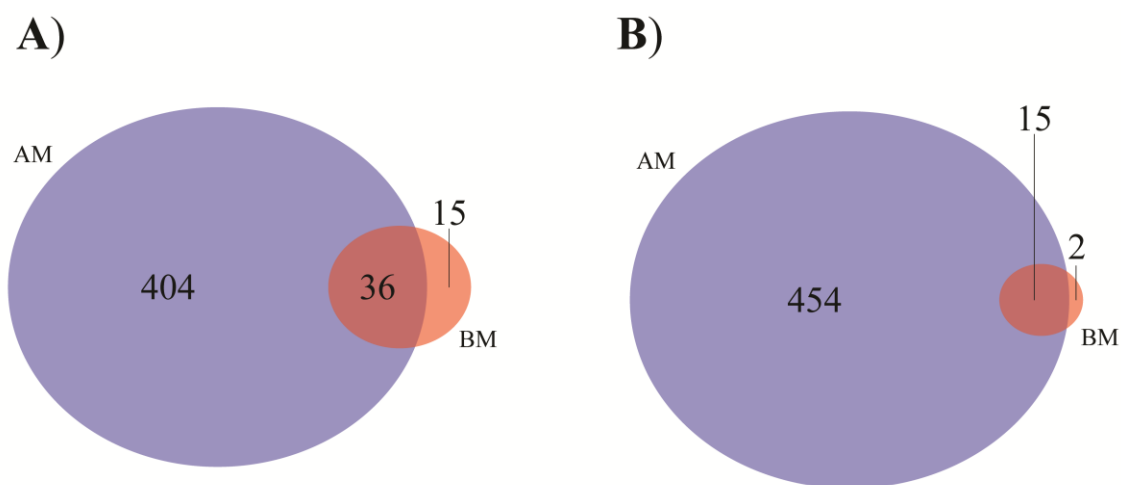
**Figura 32. SDS-PAGE 12% realizado com 10  $\mu$ L das frações coletadas a partir dos fracionamentos em GELFrEE das amostras provenientes de cultivo CGXII.** As condições são indicadas pelas letras, sendo a controle (A) e com tween 40 (B), onde as frações são representadas por “F.x”, sendo x o tempo em minutos após a saída da frente de corrida do GELFrEE. M.M. representa o marcador molecular.

Na condição controle foram detectados 5.396 PrSMs, possuindo 455 acessos de proteínas não redundantes, correspondentes à 732 proteoformas identificadas automaticamente pelo ProSightPC 4.0. Em condição com estímulo à produção de glutamato, adição de tween 40, foram detectados 2.840 PrSMs, sendo desses 471 acessos não redundantes e 643 proteoformas identificadas automaticamente. Combinando-se as duas condições foram obtidas 8.216 PrSMs, 702 proteínas não redundantes e 1.135 proteoformas diferentes (tabela 7). A diferença na quantidade de PrSMs entre as amostras provenientes de meio BT e CGXII provavelmente ocorreu devido a quantidade de injeções utilizadas em cada análise, três para frações obtidas a partir de CGXII e uma para as obtidas de meio BT.

**Tabela 7. Quantidade de PrSMs, proteínas não redundantes e proteoformas diferentes identificadas nas duas condições de cultivo em meio CGXII.** A quantificação do total foi feita a partir da junção dos dados de PrSMs das condições e fazendo a remoção das redundâncias, como descrito anteriormente.

Condição	PrSMs	Proteínas	Proteoformas
Controle	5396	455	732
Tween 40	2820	471	643
<b>Total</b>	<b>8216</b>	<b>702</b>	<b>1135</b>

Em relação ao tipo de busca e suas respectivas identificações, foi notado padrão semelhante entre as condições, com maioria dos acessos não redundantes sendo identificados em ferramenta AM com grande sobreposição entre AM e BM (figura 33).

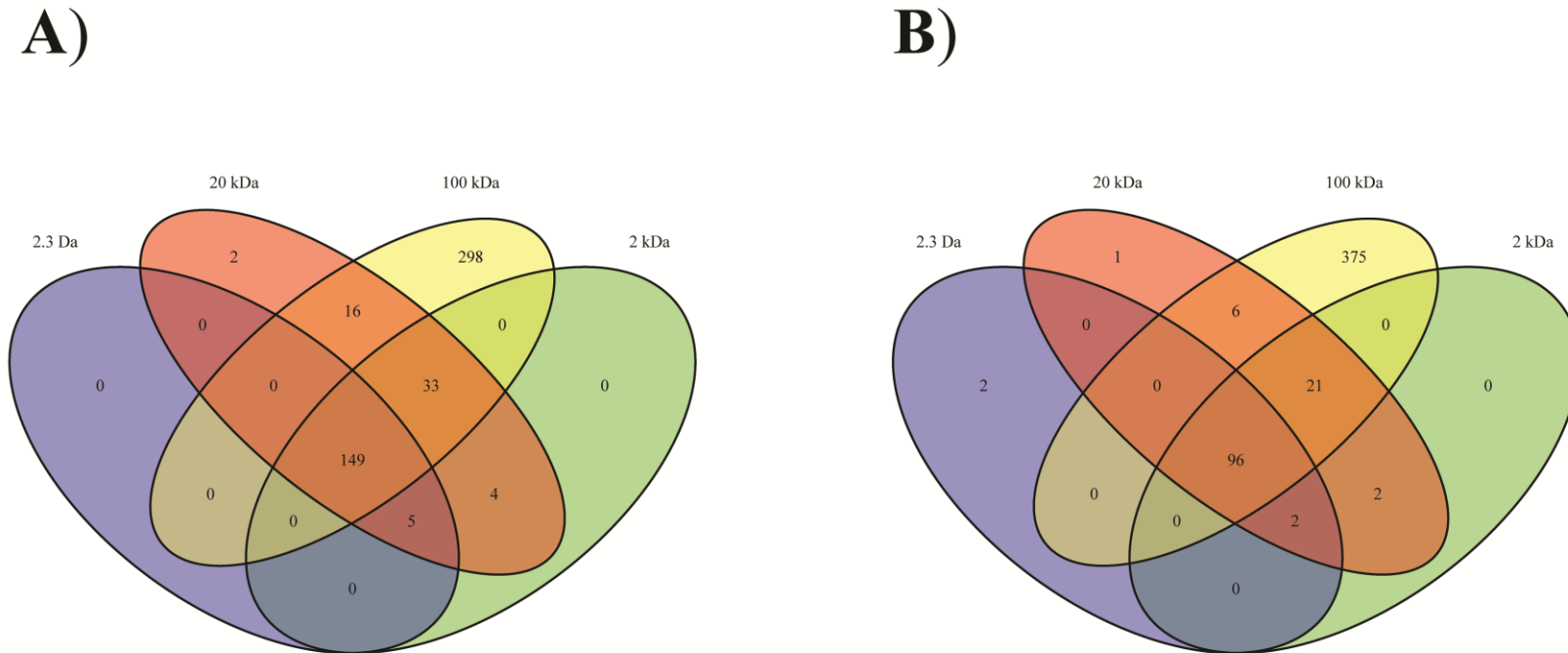


**Figura 33. Relação entre a quantidade de acessos únicos identificados em cada ferramenta de busca das proteínas provenientes de meio CGXII.** *Absolute mass* =AM e *Biomarker* = BM. Essa análise foi feita separadamente para os acessos únicos identificados nas condições controle (A) e com tween 40 (B).

As identificações por AM foram feitas com vários níveis de tolerância do precursor, como descrito anteriormente. Novamente as buscas com tolerância alta (20 kDa e 100 kDa) contribuíram muito nas identificações de acessos únicos (figura 34). Esses experimentos demonstram a importância de utilizar vários níveis de tolerância e

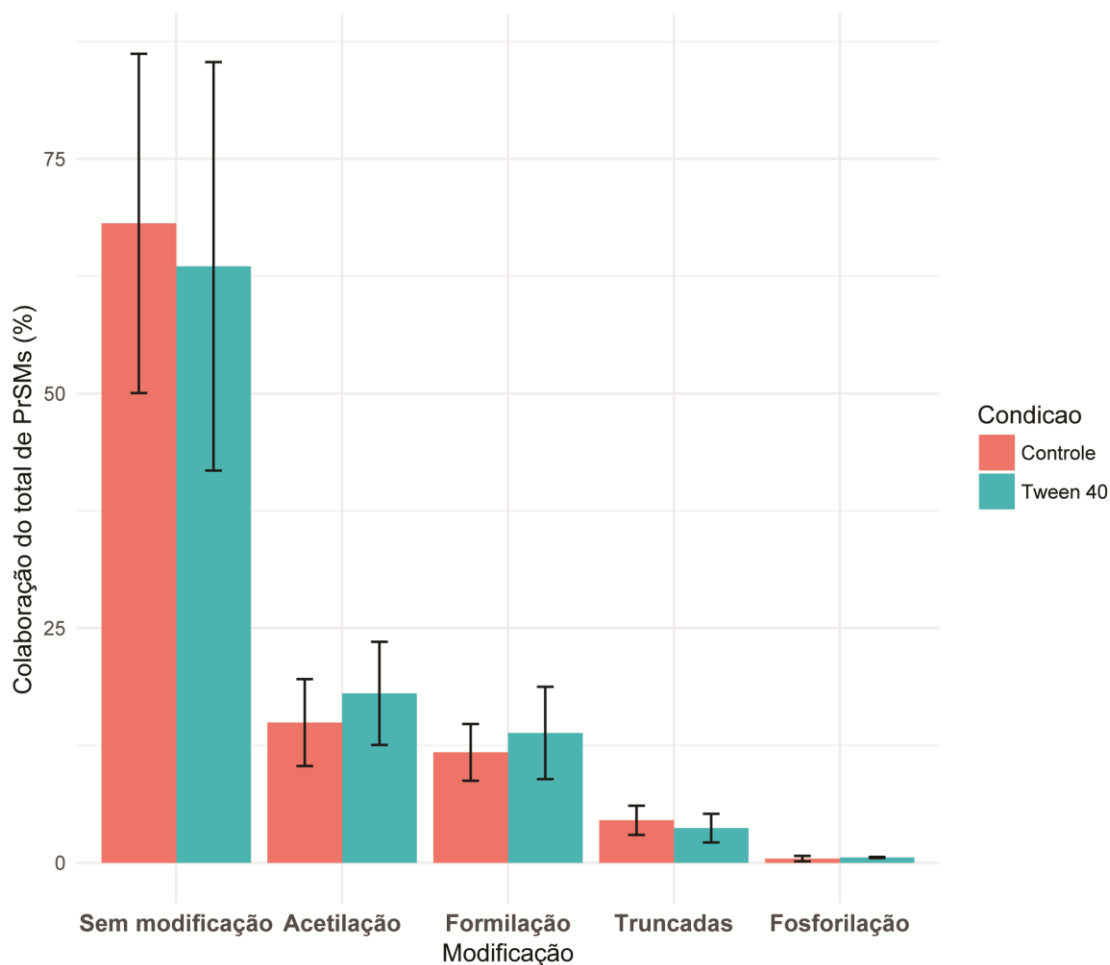
indicam a presença de proteoformas com massas diferentes das previstas por bancos de dados, além de sugerir a presença de mais de uma proteoforma no caso de acessos identificados em mais de um nível de tolerância.





**Figura 34. Relação de acessos únicos identificados a partir de proteínas de cultivos em meio CGXII em cada condição.** A análise foi feita com as identificações obtidas por AM, sendo A a condição controle e B com tween 40. Os diferentes níveis de tolerância para identificação do precursor são representados pelas circunferências: 2,3 Da (azul); 2 kDa (verde); 20 kDa (Vermelho) e 100 kDa (amarelo). A intersecção entre os círculos demonstra acessos identificados por mais de um nível de tolerância e os espaços sem intersecção demonstram acessos identificados apenas em um nível de tolerância.

Foram identificadas em ambas condições proteoformas referentes a acetilação, formilação, fosforilação, truncadas e sem modificações, sendo as acetilações as mais presentes em ambas condições. A acetilação de resíduos de lisina foi relacionada com o estímulo da produção de glutamato por *C. glutamicum*. Foi demonstrado que a partir da adição de tween 40 ocorre redução de acetilações e aumento de succinilações em proteínas de *C. glutamicum* (MIZUNO et al., 2016). Baseado nisso, foi feita quantificação de PrSMs de acordo com tipo de modificação tentando observar diferenças nesses padrões. Brevemente, as replicatas técnicas foram consideradas para observar a variação do equipamento e a partir da média delas foram feitos gráficos de barra com desvio padrão de cada condição. As quantidades de cada tipo de modificação das duas condições foram comparadas em relação a quantidade total de espectros identificados para cada condição, sendo feita a normalização pelo número de PrSMs em cada condição. Não foi notada nenhuma mudança significativa nos tipos de modificações entre as condições com tween 40 e controle (figura 35), demonstrando que apenas a contagem de espectros não é o suficiente para diferenciá-las.

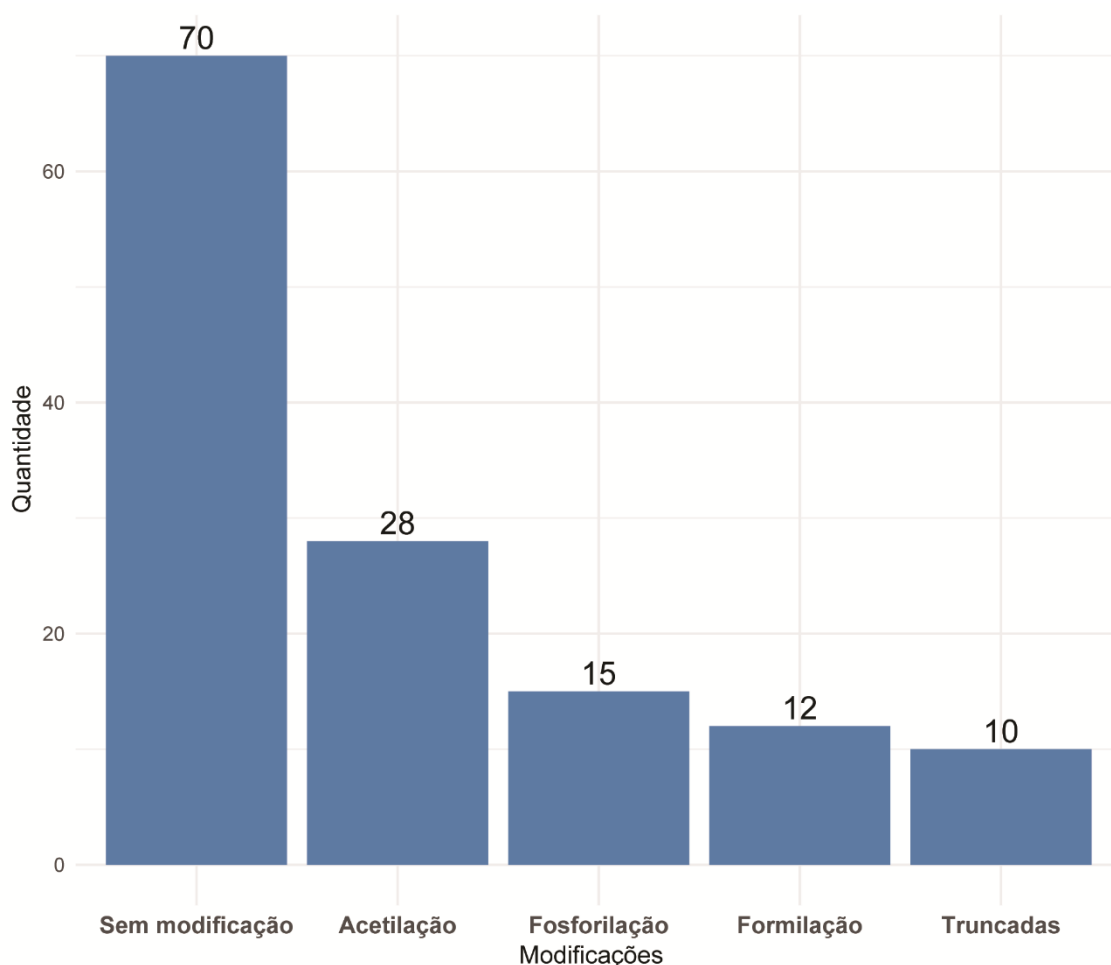


**Figura 35. Relação da quantidade de PrSMs detectados para cada tipo de modificação nas condições controle e com adição de Tween 40.**

#### 5.4.1. Anotação de proteínas em condição controle

Inicialmente as sequências correspondentes aos acessos únicos identificados em condição controle foram submetidas a BlastKOALA. As 455 proteínas não redundantes da condição controle foram mapeadas e 272 foram anotadas de acordo com suas KOs. Sobre as vias relacionadas à produção de aminoácidos, 16 acessos foram anotados como constituintes da via de biossíntese de aminoácidos, 10 à via da glicólise/gliconeogênese, 5 ao ciclo do ácido cítrico e 3 à PPP. A diferença na quantidade de acessos obtidos para essas vias entre o cultivo em meio BT e meio CGXII provavelmente foi ocasionada pela ausência das identificações adquiridas por LH em meio CGXII, ocasionando na redução das identificações de proteínas maiores que 50 kDa.

Igualmente ao realizado com as proteínas de meio BT, todos os acessos identificados nas vias supracitadas (24 acessos não redundantes), foram considerados como relacionados à produção de aminoácidos. As PrSMs (132) referentes a essas proteínas foram anotadas manualmente quanto à presença de proteoformas observando o espectro e a diferença de massas dos precursores, esperadas e teóricas. A anotação automática do ProSight PC referente a essas PrSMs detectou todos os tipos de modificações observados na amostra total (figura 36).



**Figura 36. Quantidade de PrSMs detectados em proteínas relevantes na produção de aminoácidos de acordo com a modificação identificada.**

Apenas uma proteína não redundante foi identificada com fosforilação, a lactato desidrogenase (LDH), representada pelo código Q8NLN0, anotada na via de glicólise/gliconeogênese. Essa enzima auxilia na conversão de piruvato em L-lactato, podendo assim influenciar o fluxo de carbono para o ciclo do ácido cítrico e metabolismo

de vários aminoácidos, como glutamato e lisina. A influência de LDH na produção de aminoácidos por *C. glutamicum* foi demonstrada recentemente por Zhang, et al. (2014). Por meio da deleção do gene *ldhA*, codificador de LDH, ocorreu aumento de 10,1% na produção de glutamato ao reduzir a formação de subprodutos, no entanto isso só foi alcançado em condições micro-aeróbicas (ZHANG et al., 2014).

Além de fosforilação, essa enzima foi encontrada com proteoformas truncadas, formiladas e acetiladas no N-terminal e uma forma com formilação em resíduo de lisina. Dessas proteoformas, as maiores identificadas foram com 24,37 kDa e 23,73 kDa, com e sem fosforilação, respectivamente. Todas as formas possuindo formilações ou acetilações foram detectadas com massa próximas à 6,90 kDa, sugerindo degradação dessas moléculas. A fosforilação de LDH está descrita em eucariotos como mecanismo de regulação da atividade dessa enzima, onde estados mais fosforilados são relacionados à redução de afinidade pelo seu substrato e aumento de inibição por piruvato, servindo como regulação perante a condição de oxigênio presente nas células (STOREY, 2016). A detecção de LDH fosforilada faz sentido considerando que as condições utilizadas nesse experimento são aeróbicas. Levando em conta a identificação dessa proteoforma fosforilada e a relevância de LDH na produção de aminoácidos, abre-se a perspectiva de um estudo mais detalhado dos mecanismos de fosforilação de LDH e como esses mecanismos podem afetar a produção de aminoácidos por *C. glutamicum*.

Algumas proteínas que possuem relação com o metabolismo de aminoácidos foram identificadas com o precursor observado com massa consideravelmente menor do que a do precursor teórico (3% a 50%) (tabela 8). Apesar disso, a maioria dos PrSMs foram identificados com massa dos precursores observados mais próximas ou maiores à do precursor teórico (figura 29), indicando presença de proteoformas.

**Tabela 8. Proteínas relacionadas ao metabolismo de aminoácidos possuindo relação P.O./P.T. ≤ 50 %.** P.O./P.T. = relação entre massas de precursores observados e teóricos.

Acesso	Descrição	P.O./P.T.(%)	E-value
Q8NQ98	Aconitato hidratase A	3,93	2,94E-06
Q8NNF6	Componente E1 da Piruvato desidrogenase	3,98	4,48E-09
Q8NTD7	Proteína não caracterizada	4,28	1,41E-09
P42455	2-isopropilmalato sintase	7,64	1,64E-05
Q8NN78	Glicose-6-fosfato 1-epimerase putativa	8,24	3,30E-11
P19537	Frutose-bifosfato aldolase	9,88	1,32E-07

Continuação tabela 8.

Acesso	Descrição	P.O./P.T.(%)	E-value
<b>Q8NMN0</b>	Proteína não caracterizada	10,95	1,88E-05
<b>Q59284</b>	Succinil-diaminopemelate desuccinilase	13,87	4,92E-05
<b>Q8NLN0</b>	L-lactato desidrogenase	15,33	5,81E-05
<b>Q8NQRO</b>	Ribocinase	16,80	8,25E-10
<b>Q8NN33</b>	Malato desidrogenase	17,83	6,54E-05
<b>P06557</b>	Componente 1 da antranilato sintase	20,24	1,80E-05
<b>P42457</b>	Citrato sintase	22,31	1,34E-05
<b>Q57179</b>	cetol-ácido redutoisomerase ((NADP(+))	23,58	5,89E-07
<b>Q01651</b>	Gliceraldeído-3-fosfato desidrogenase	30,47	9,04E-06
<b>Q8NLX9</b>	Proteína não caracterizada	31,62	2,95E-05
<b>Q8NTA5</b>	2,3-bifosfoglicerato mutase dependente de fosfoglicerato	31,66	2,62E-05
<b>Q8NLZ0</b>	Proteína não caracterizada	39,37	4,08E-05

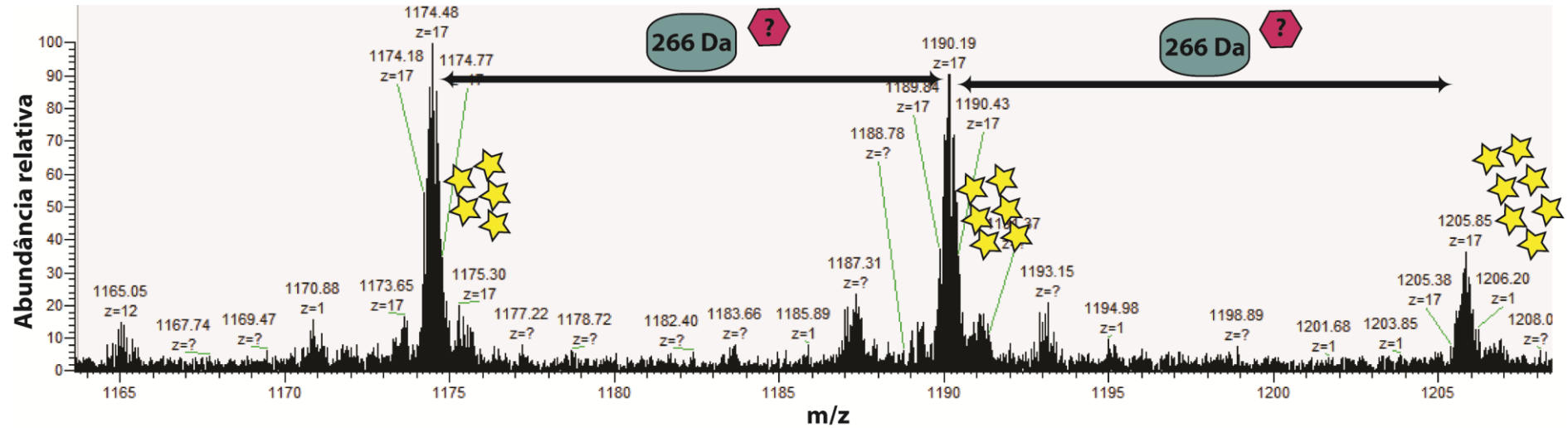
Foram detectados 19 PrSMs com relação entre precursor observado e teórico maior do que 101%, correspondendo a seis proteínas não redundantes (tabela 9).

**Tabela 9. PrSMs detectados com relação de diferença entre precursor observado e teórico (P.O./P.T.) maior do que 101%.**

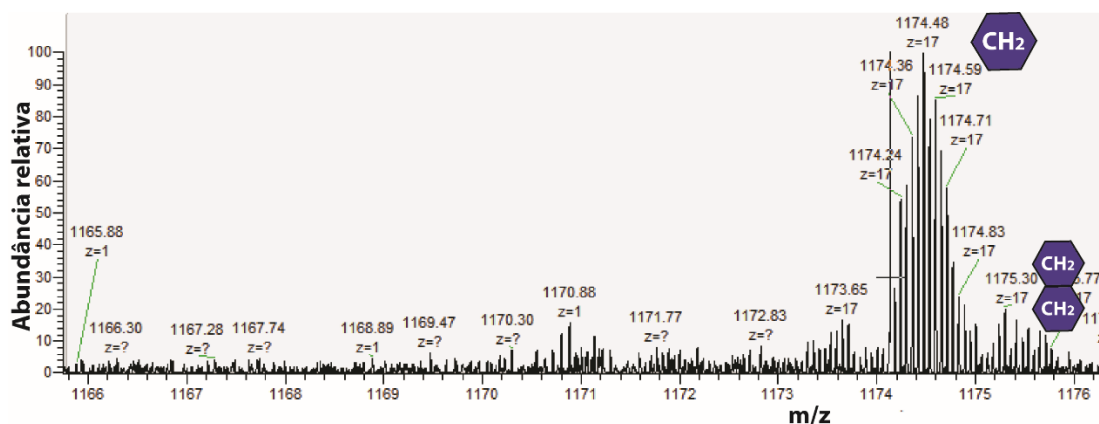
Acesso	Descrição	P.O./P.T.	E-value	Diferença de massa (Da)
P19583	Triosefosfato isomerase	104,7491133	5,64E-10	1.284,87
Q79VG4	Proteína não caracterizada	108,58292	9,00E-07	1.596,95
Q79VG4	Proteína não caracterizada	108,5827378	2,78E-10	1.596,92
Q79VG4	Proteína não caracterizada	108,5827103	1,49E-18	1.596,92
Q79VG4	Proteína não caracterizada	108,5827017	4,51E-07	1.596,91
Q79VG4	Proteína não caracterizada	108,5826276	1,40E-28	1.596,90
Q79VG4	Proteína não caracterizada	107,1523372	2,36E-10	1.330,78
Q79VG4	Proteína não caracterizada	107,1523173	3,90E-09	1.330,77
Q79VG4	Proteína não caracterizada	107,1523007	9,97E-10	1.330,77
Q79VG4	Proteína não caracterizada	107,1522985	1,74E-24	1.330,77
Q79VG4	Proteína não caracterizada	102,8661211	5,44E-09	533,28
Q79VG4	Proteína não caracterizada	102,8607406	9,65E-07	532,27
Q79VG4	Proteína não caracterizada	101,4248806	1,15E-09	265,12
Q79VG4	Proteína não caracterizada	101,424844	7,72E-33	265,11
Q8NQV7	Subunidade pequena da 3-isopropilmalato desidratase	102,3991572	1,30E-07	532,26
Q8NQV7	Subunidade pequena da 3-isopropilmalato desidratase	101,1994852	2,66E-08	266,11
Q8NQV7	Subunidade pequena da 3-isopropilmalato desidratase	101,1948109	1,25E-11	265,07
Q8NS29	Proteína não caracterizada	120,7091181	1,12E-05	2.395,41
Q9Z471	Fosforibosil-ATP pirofosfatase	108,134229	6,65E-09	798,46

Uma distribuição de massas semelhante foi observada em duas proteínas da biossíntese de valina e leucina, Q79VG4 e Q8NQV7. Essa variação de massa corresponde a um padrão de diferentes níveis de adições de grupamentos com 266 Da. Foram identificadas proteoformas de Q79VG4 com massas referentes a 1, 2, 4 e 5 adições de grupamentos com 266 Da e formas com 1 e 2 adições em Q8NQV7 (tabela 9 e figura 38). Essa massa já foi relacionada tanto à adição de grupamentos estearoil (SENYILMAZ et al., 2015), no entanto também já foi identificada como decorrente de adutos de SDS (FRIDRIKSSON; BAIRD; MCLAFFERTY, 1999). Recentemente a modificação de proteínas por estearoilação foi relacionada à ativação de via de sinalização em mitocôndrias e várias proteínas foram identificadas com essa modificação em células de mamíferos (SENYILMAZ et al., 2015). Ademais, ao observar alguns dos íons isolados responsáveis pela identificação da Q79VG4 foi possível observar dois envelopes de cargas próximos a eles com carga igual à do envelope do íon fragmentado e identificado, a diferença de massa entre os íons monoisotópicos desses envelopes foi calculada como 14 Da, sugerindo presença de metilação (figura 38).



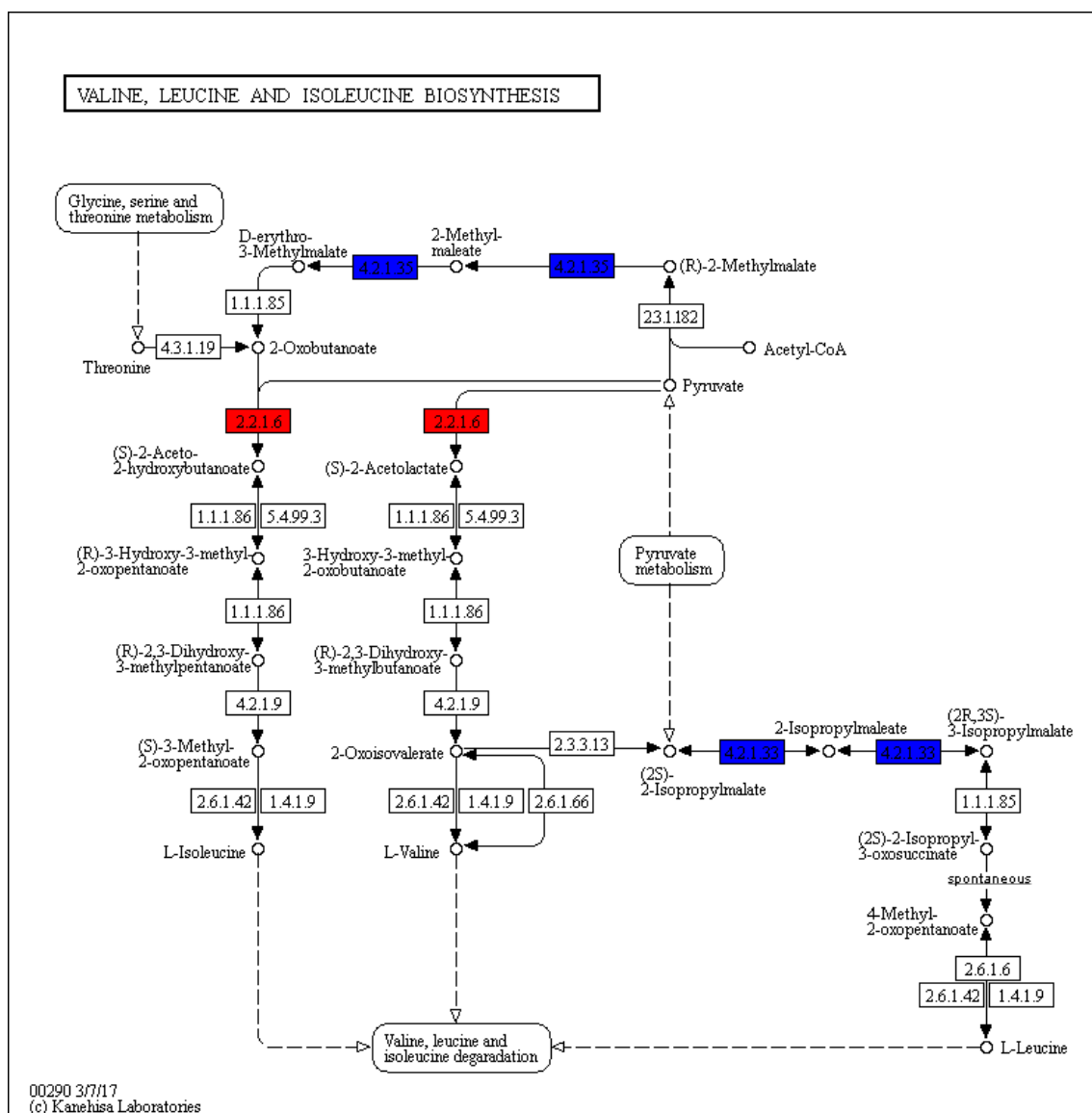


**Figura 37.** Espectro de massa com a relação massa carga ( $m/z$ ) e abundância relativa dos envelopes isotópicos responsáveis pela identificação de duas formas da proteína Q79VG4. Os íons com  $m/z$  1174,48 e 1190,19 foram fragmentados e identificados como a proteína Q79VG4, com 5 e 6 adições de grupamentos com 266 Da, respectivamente. O envelope com íon de  $m/z = 1205,85$  não foi identificado, mas apresentou variação de massa semelhante a predita nos anteriores.



**Figura 38.** Espectro de massa com m/z e abundância relativa de um dos íons identificadores da proteína Q79VG4 e envelopes de cargas demonstrando possíveis metilações. As metilações são identificadas pelos hexágonos azuis, sendo que a forma identificada possui supostamente uma metilação não identificada automaticamente pelo ProSightPC 4.0.

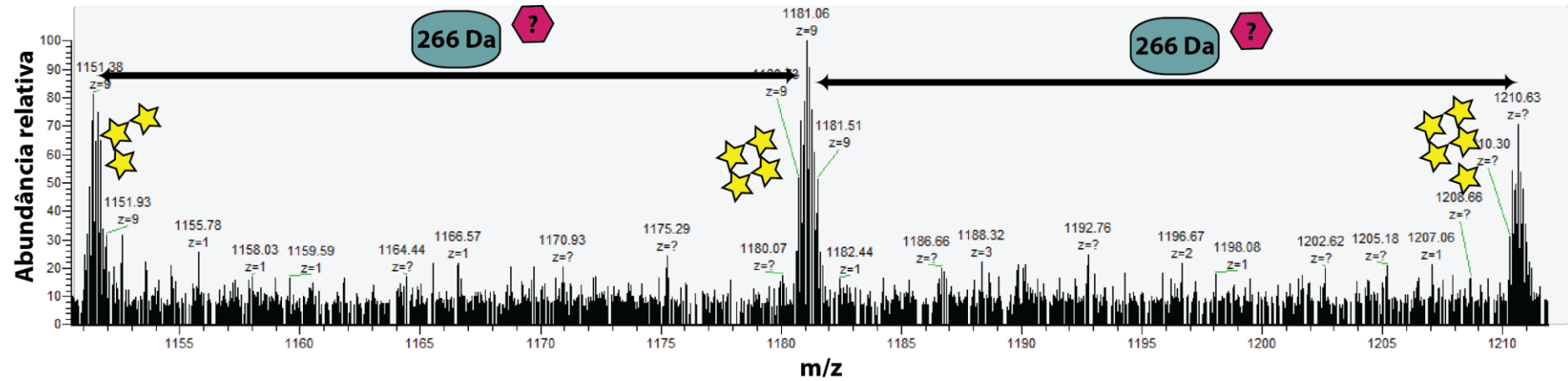
O acesso Q79VG4 é correspondente à proteína não caracterizada de *C. glutamicum*, no entanto ela foi identificada por similaridade pelo BlastKOALA como uma subunidade da acetolactato sintase, enzima mapeada na via de biossíntese de aminoácidos pelo KEGG, participando da conversão de piruvato em (S)-2-acetolactato, substrato para produção de 2-oxoisovalerato, que por sua vez é substrato para a produção de valina. O 2-oxoisovalerato está situado em ponto de ramificação para a produção de valina e leucina, ele é precursor de 4-metil-2-oxopentanoato, que por sua vez é precursor da leucina (figura 39), a proteína responsável pelo primeiro passo de conversão de 4-metil-2-oxopentanoato em L-leucina é a Q8NQV7, subunidade pequena da 3-isopropilmalato desidratase.



**Figura 39.** Mapa da biossíntese de valina, isoleucina e leucina gerado pelo KEGG. Os metabólitos são representados por circunferências e enzimas em retângulos com seus respectivos ECs. A proteína Q79VG4, identificada por similaridade como subunidade da acetolactato sintase, está destacada em vermelho e a proteína Q8NQV7, subunidade pequena da 3-isopropilmalato desidratase. (ECs: 1.1.1.85 = 3-isopropilmalato desidrogenase; 1.1.1.8 = cetol-ácido redutoisomerase ((NADP(+))); 1.4.1.9 = leucina desidrogenase; 2.2.1.6 = acetolactato sintase; 2.3.1.182 = (R)-citramalato sintase; 2.3.3.13 = 2-isopropilmalato sintase; 2.6.1.42 = transaminase de aminoácidos de cadeias ramificadas; 2.6.1.6 = leucina transaminase; 2.6.1.66 = Valina--piruvato transaminase; 4.2.1.33 = 3-isopropilmalato desidratase; 4.2.1.35 = (R)-2-metilmalato desidratase; 4.2.1.9 = Dihidroxi-ácido desidratase; 4.3.1.19 = Treonina amônia-liase; 5.4.99.3 = 2-acetolactato mutase).

Além das variações de massa detectadas nos acessos Q79VG4 e Q8NQV7, as proteínas Q8NS29 e Q9Z471 foram identificadas com diferença de massas grandes, 2395 e 798, respectivamente, indicando a presença de glicosilações ou várias PTMs

combinadas. A Q8NS29 é uma proteína não caracterizada, no entanto foi identificada por homologia como corismato mutase, enzima presente na via de biossíntese de fenilalanina, tirosina e triptofano. A proteína identificada como Q9Z471 é uma fosforibosil-ATP pirofosfatase, pertencente ao metabolismo de histidina. Ao observar o espectro responsável pela identificação da fosforibosil-ATP pirofosfatase foram notados envelopes com cargas semelhantes à do íon precursor da identificação dessa proteína. Ao calcular a diferença de massa entre eles, foi detectada mais uma vez a variação de 266 Da entre os íons monoisotópicos, sugerindo novamente presença de adutos de SDS ou modificação por estearoilação, sendo o íon precursor da identificação identificado com três de alguma dessas modificações (figura 40).

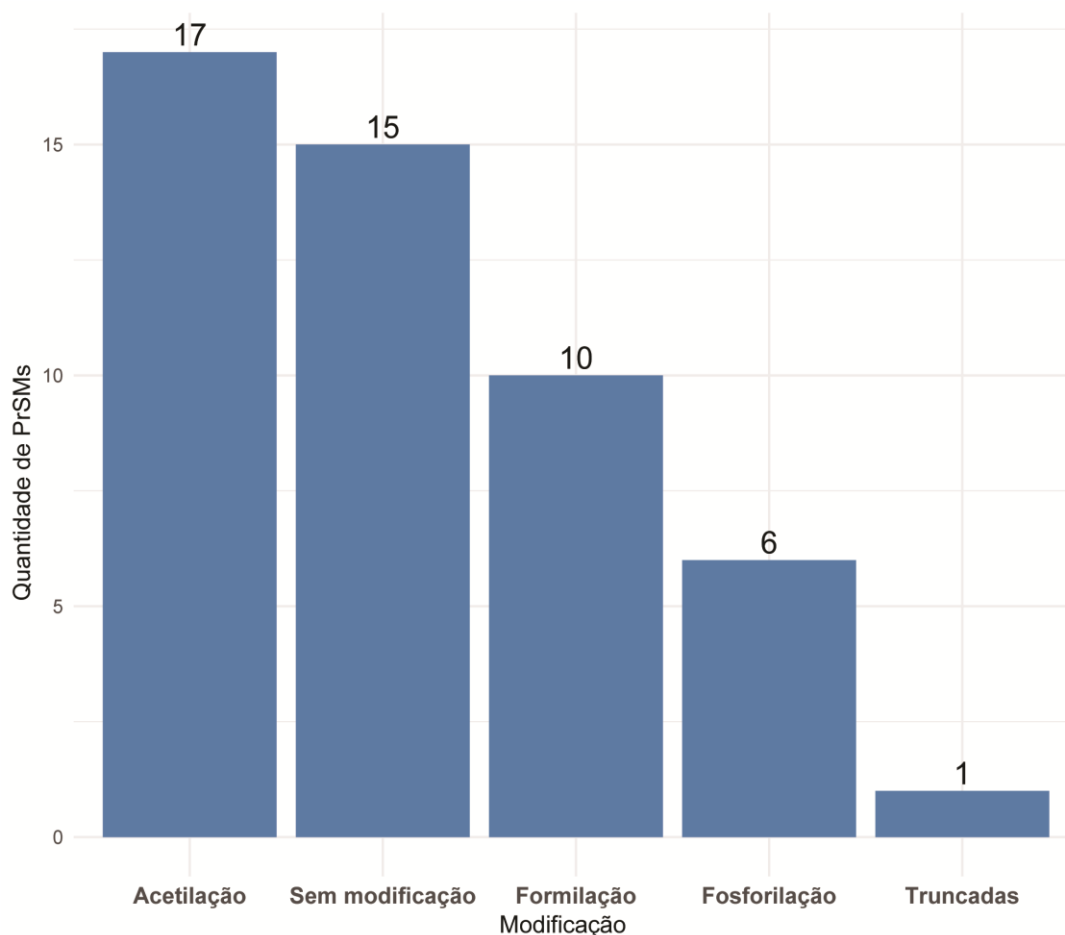


**Figura 40. Espectro de massa do íon precursor da identificação da proteína fosforibosil-ATP pirofosfatase.** O íon responsável pela identificação do acesso Q9Z471 (fosforibosil-ATP pirofosfatase) possui  $m/z = 1181,06$ , nenhum íon dos envelopes de cargas próximos foi identificado como Q9Z471, no entanto coeluíram próximos a essa proteína e apresentam carga igual, indicação de serem a mesma proteína com modificações. A diferença de massa calculada para esses é de 266 Da, presumindo-se grupamentos SDS ou estearoils, representados na imagem pelas estrelas amarelas.

A adição de grupamentos a proteínas responsáveis pela biossíntese de aminoácidos é esperada, considerando a característica regulatória das PTMs e a necessidade de regulação altamente controlada na produção dessas biomoléculas. No entanto, as modificações sugeridas sem identificação necessitam de experimentos mais aprofundados para serem validadas.

#### **5.4.2. Anotação de proteínas em condição com tween 40**

As 471 proteínas identificadas em condição com tween 40 foram mapeadas por blastKOALA e 274 dessas foram anotadas de acordo com seus KOs. Na biossíntese de aminoácidos foram encontradas 15 proteínas, 3 componentes da glicólise/gliconeogênese, 6 do ciclo do ácido cítrico e 2 da via das pentoses fosfato. Todas essas proteínas correspondem ao total de 20 acessos não redundantes, com 27 proteoformas anotadas automaticamente em 48 PrSMs (figura 41). Como observado nas outras condições, os PrSMs de acetilação superaram, mesmo que pouco, os de sem modificação, indicando função para essa modificação nessas proteínas.



**Figura 41. Quantidade de PrSMs detectados para cada tipo de modificação presente nas proteínas relacionadas ao metabolismo de aminoácidos da condição com tween 40.**

Os PrSMs relacionados ao metabolismo de aminoácidos foram analisados quanto a relação tamanho do precursor observado e teórico para verificar a presença de degradação ou proteoformas. Nenhum precursor observado foi detectado como maior do que o teórico e 37 precursores observados, referentes a 18 das 20 proteínas não redundantes identificadas, possuíam massa menor do que 50 % da massa do precursor teórico (tabela 10). Essas identificações de pedaços proteicos indicam processamento proteolítico nesse grupo de proteínas.

**Tabela 10.** Acessos não redundantes detectados com maior similaridade (mais próximo de 100%) na relação de diferença entre precursor observado e teórico (P.O./P.T.).

Acesso	Descrição	P.O./P.T.	E.Value	Diferença de massa (Da)
Q59280	Cadeia alfa da proteína bifuncional ArgJ	99,99	8,87E-18	-2,06
Q8NS29	Proteína não caracterizada	98,31	1,52E-07	-193,07
Q8NLN0	L-lactato desidrogenase	25,26	2,15E-20	-25.757,29
Q8NRS1	Enolase	25,04	3,05E-05	-33.605,55
P42457	Citrato sintase	22,88	8,20E-05	-37.611,55
Q8NQ25	Proteína não caracterizada	22,85	4,91E-05	-22.617,28
P35170	Fosfo-2-dehidro-3-deoxiheptanato aldolase	18,77	6,71E-07	-31.800,04
Q8NMK7	Subunidade beta da succinato--CoA ligase [formando ADP]	17,68	2,53E-06	-34.284,58
Q8NQ98	Aconitato hidratase A	14,33	3,40E-05	-87.120,32
Q8NNF6	Componente E1 da piruvado desidrogenase	12,82	6,17E-05	-89.468,86
Q8NTD6	Succinato desidrogenase/fumarato redutase, subunidade de flavoproteína	12,25	1,87E-05	-58.372,83
P94631	3-isopropilmalato desidrogenase	11,05	2,41E-06	-32.053,30
Q04513	L-treonina desidratase biosintetica IlvA	10,48	2,23E-06	-41.897,09
Q8NS65	Proteína não caracterizada	9,42	9,11E-08	-40.122,42
Q9X5C9	quinato/shiquimato desidrogenase (NAD(+))	9,36	4,01E-06	-26.924,43
Q8NLD4	Proteína não caracterizada	8,11	1,82E-07	-49.280,58
P42463	Subunidade grande da acetolactato sintase	6,70	2,73E-07	-62.243,23
P50216	Isocitrato desidrogenase [NADP]	5,02	6,40E-09	-76.017,34
Q8NQ64	Transaldolase	4,65	3,10E-06	-36.400,49
Q59279	N-acetil-gama-glutamil-fosfato redutase	3,63	2,90E-05	-34.435,58



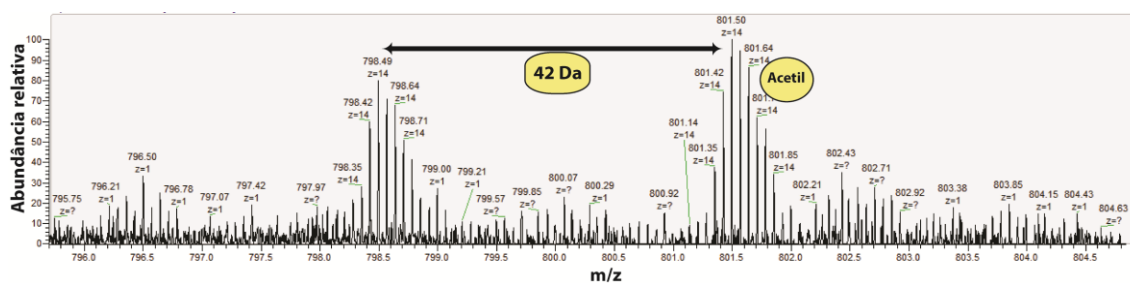
Foram detectados seis PrSMs com fosforilações no grupo com tween 40, referentes à duas proteínas não redundantes, no entanto eles apresentaram tamanho do precursor observado pequeno em relação ao teórico, representando 10%, na proteína Q04513, e entre 22% e 25% para a proteína Q8NLN0.

Como esperado, a proteína LDH (Q8NLN0) foi identificada fosforilada nas duas condições, já que não houve alteração em condições de oxigenação. A proteína Q04513 é descrita como L-treonina desidratase biosintética IlvA e participa do metabolismo de L-treonina e L-isoleucina, auxiliando na conversão de L-treonina em 2-oxobutanoato, primeiro precursor da biossíntese de isoleucina. A via de biossíntese de isoleucina é uma via metabólica altamente regulada, a inibição de L-treonina desidratase já fora descrita por ser dependente da concentração de L-isoleucina em *C. glutamicum* (MORBACH; SAHM; EGGELING, 1995). Baseado nisso Guillouet et al. (1999) expressou em *C. glutamicum* o gene *tdcB*, também uma L-treonina desidratase, no entanto, essa é resistente ao efeito de inibição por L-isoleucina. Por meio dessa expressão heteróloga a bactéria alcançou aumento de 50 vezes na produção de L-isoleucina. A L-treonina desidratase utilizada nesse estudo é produzida em *E. coli* durante condição de crescimento anaeróbico (GUILLOUET et al., 1999). Considerando essa característica de expressão e a capacidade de regulação da fosforilação, inclusive sua característica como reguladora da enzima LDH reagindo a condições de oxigênio, estudos mais aprofundados dessa modificação na enzima L-treonina desidratase em *C. glutamicum* apresentam potencial para descobertas de mecanismos de regulação para a biossíntese de L-isoleucina.

A proteína de acesso Q8NS29 não está caracterizada para *C. glutamicum*, no entanto foi anotada por similaridade em blast à corismato mutase. Diferença de massa de 42 Da entre os PrSMs dessa proteína foi observada (tabela 11), essa diferença de massa normalmente é correlacionada a adição de um grupamento acetil. Ademais, duas PrSMs correspondentes à Q8NS29 foram detectadas coeluidando e a diferença entre os seus picos monoisotópicos foi calculada, como esperado essa diferença correspondeu a 42 Da, reforçando a hipótese de eventos acetilação nessa proteína. A partir de seus espectros, detalhe quantitativo pode ser notado entre essas duas formas, ambas foram encontradas com abundâncias relativas semelhantes, com pequeno aumento para a forma com uma suposta acetilação a mais (figura 42).

**Tabela 11. Tabela de PrSMs do acesso Q8NS29.** Diferença entre massa do precursor observada e precursor teórico (P.O. - P.T.) e diferença de massa representa a diferença de massa entre os PrSMs detectados, no caso, entre o indicado e o da linha acima.

Acesso	E. Value	P.O.-P.T. (Da)	Diferença de massa (Da)
Q8NS29	1,93E-05	-393,12	
Q8NS29	4,66E-10	-351,10	42,01
Q8NS29	4,60E-05	-351,10	0,00
Q8NS29	7,18E-09	-351,10	0,00
Q8NS29	7,23E-07	-276,09	75,01
Q8NS29	9,61E-05	-235,09	41,00
Q8NS29	1,52E-07	-193,07	42,02



**Figura 42. Espectro de massa de duas PrSMs identificadas como Q8NS29, apresentando diferença de 42 Da entre si.** A diferença entre os íons monoisotópicos de cada um está representada pela seta de duas pontas com a massa indicada na forma abaixo, 42 Da, correspondendo a uma forma com uma provável adição de grupamento acetil, representado pela circunferência amarela.

#### 5.4.3. Anotação de proteínas presentes na condição controle e tween 40

Cinco proteínas relacionadas a produção de aminoácidos foram identificadas nas duas condições (P42457, Q59280, Q8NLN0, Q8NNF6 e Q8NS29). Foram detectados 46 PrSMs na condição controle e 25 na condição com tween 40. Apenas estados fragmentados da citrato sintase foram encontrados nas duas condições, foram identificadas automaticamente proteoformas digeridas, com acetilação no N-terminal, com fosforilação e sem modificação. A cadeia alfa da proteína bifuncional ArgJ (Q59280) foi detectada por apenas um PrSM na condição controle e por quatro PrSMs na condição com tween 40, em ambas condições foi encontrada com massa intacta e sem

modificações. A L-lactato desidrogenase (Q8NLN0) foi detectada 25 vezes na condição controle e 6 vezes na condição com tween 40, na condição controle ela foi identificada 7 vezes pela ferramenta BM e as outras 18 foram pela ferramenta AM com alta tolerância (20 kDa ou 100 kDa), nenhum PrSM foi identificado por BM na condição com tween 40, no entanto todos eles foram identificados por AM com tolerância de 100 kDa. Essas identificações novamente reforçam a hipótese de que essa proteína esteja sofrendo proteólise em ambas condições. Apesar desse suposto processamento, foram identificados precursores com até 24,5 kDa de L lactato desidrogenase na condição controle, enquanto na condição com tween o maior precursor observado foi de 8,71 kDa. Além disso foram anotadas proteoformas com fosforilação de tirosina nas duas condições.

O componente E1 da piruvato desidrogenase só foi identificado por pedaços, representando até 16 % da massa teórica, nas duas condições. Por essa ser uma proteína de 102,76 kDa e não ter sido feito a aquisição por LH esse é um dado esperado, já que ela não deve estar presente nas primeiras frações do GELFrEE, devido a sua massa teórica ser maior que 50 kDa.

A proteína que foi detectada por um precursor com massa mais próxima do precursor teórico nas duas condições foi a Q8NS29, proteína não caracterizada de *C. glutamicum*, foi identificada por similaridade por blastKOALA como sendo uma corismato mutase, responsável pelo rearranjo de corismato a pterinato, molécula precursora de tirosina e fenilalanina (LI; LIU; LIU, 2009). Burschowsky et al. (2018) demonstrou que a corismato mutase de *C. glutamicum* formando um homotetrâmero é complexada com o homotetrâmero da 3-Deoxi-D-arabinoheptulosonato 7-fosfato (DAHP) sintase, proteína que catalisa o primeiro passo da via do chiquimato. Além disso, nesse estudo foi comprovado que a formação desse complexo aumenta a atividade da corismato mutase de *C. glutamicum* e que esse complexo sofre regulação negativa de tirosina e fenilalanina. Tal inibição ocorre por meio de ligação alostérica com o complexo da DAHP sintase, o que desmonta o supercomplexo de DAHP sintase e corismato mutase, reduzindo a sua atividade (BURSCHOWSKY et al., 2018).

A corismato mutase foi identificada nas duas condições possuindo pequenas clivagens na sua sequência, correspondendo a fragmentos de 393 Da a 193 Da em condição com tween 40, enquanto na condição controle foram detectados fragmentos com perdas entre 192 Da e 350 Da, indicando processamentos proteolíticos semelhantes. Em ambas condições foram detectadas proteoformas com acetilação ou formilação no N-terminal. A presença dessas clivagens na sequência da corismato mutase nas duas

condições sugerem relevância no funcionamento dessa enzima. Essa evidência de processamento proteolítico junto a característica funcional peculiar dessa enzima em formação de supercomplexo demonstra potencial no estudo dessas clivagens e seus efeitos na atividade da proteína corismato mutase e de seu supercomplexo com DAHP sintase, como meio para regular a biossíntese de aminoácidos aromáticos.

## 6. Conclusões

A análise proteômica *top-down* de *C. glutamicum* demonstrou grande capacidade de caracterização de proteoformas processadas proteoliticamente. Fator interessante observado em todas as condições foi a grande quantidade de N<sup>α</sup>-acetilação como modificação, apesar dessa ser descrita como pouco usual em procariotos (CAIN; SOLIS; CORDWELL, 2014). Tal fator pode ser ocasionado pela predominância de abordagem *bottom-up* para análise proteômica, o que dificulta a identificação de N-terminais proteicos. Além disso foi observado grande detecção de N-formilação em todas as condições, demonstrando o potencial da abordagem *top-down* em estudos de degradação de proteínas procarióticas, considerando recente mecanismo de sinalização para proteólise proposto baseado nessa modificação (PIATKOV et al., 2015).

Apesar de ser esperada diferença no padrão de acetilações entre as condições controle e com tween 40 (MIZUNO et al., 2016), tal evento não foi observado nesse experimento. Isso pode ser decorrente da abordagem de contagem de espectros (PrSMs) não ter sido o suficiente para diferenciar essas condições em relação ao tipo de modificação apresentada.

Além das proteoformas anotadas automaticamente, foi possível identificar diferentes padrões de modificações não previstas no banco de dados. Uma delas foi proposta como diferentes níveis de estearoilações, no entanto estudos mais aprofundados são necessários para confirmar a identificação dessas possíveis PTMs.

A utilização de buscas com alta tolerância (20 kDa e 100 kDa) também possibilitou a identificação de vários fragmentos de proteínas. A quantidade desses foi comparada entre as condições possibilitando avaliar a degradação proteica em cada uma,

demonstrando maior conservação em amostras tratadas com SDS no tampão de lise quando comparada à utilização de Triton-X 100. Essa avaliação dificilmente seria possível em estratégia *bottom-up*.

A utilização de abordagem proteômica *top-down* em *C. glutamicum* permitiu a identificação de proteoformas que sugerem regulação metabólica por PTMs nos metabolismos de isoleucina, leucina, treonina, fenilalanina, tirosina, triptofano e glutamato. Foram detectadas nessas vias proteínas com proteoformas fosforiladas, truncadas e metiladas.

## 7. Perspectivas

Foram alcançados os objetivos de apresentar proteínas susceptíveis à PTMs no metabolismo de aminoácidos de *C. glutamicum*, possibilitando direcionamento de estudos para regulação da produção de aminoácidos. No entanto, tais conquistas trouxeram consigo necessidades de validações por metodologia complementares a espectrometria de massa e o desenvolvimento de novas hipóteses, com grande potencial para descobertas de mecanismos capazes de aperfeiçoar a produção de aminoácidos em microrganismos. Baseado nisso apresento como perspectivas:

- 1) Estudo dos mecanismos de regulação das enzimas lactato desidrogenase e treonina desidratase por fosforilação.
- 2) Validação de acetilações propostas em corismato mutase e sua função na regulação dessa enzima.
- 3) Estudo da influência da clivagem da corismato mutase na formação do supercomplexo DAHP sintase + corismato mutase.
- 4) Validação de proteoformas sugeridas como estearoilação e investigação de possíveis mecanismos de regulação causados por essa modificação.

Além das perspectivas geradas diretamente pelos resultados da análise proteômica *top-down* realizada em *C. glutamicum*, também foi proposto projeto com intuito de caracterizar de forma mais holística o fenótipo apresentado por essa bactéria. Esse foi submetido e aprovado, como projeto de doutorado, prevendo a análise de biologia de

sistemas dessa bactéria, por meio de análises transcriptômicas, proteômicas, metabolômicas e fluxômicas.

## 8. Referências bibliográficas

AEBERSOLD, R.; MANN, M. Mass spectrometry-based proteomics. **Nature**, v. 422, n. 6928, p. 198–207, 2003.

BECKER, J.; WITTMANN, C. Bio-based production of chemicals, materials and fuels--Corynebacterium glutamicum as versatile cell factory. **Curr Opin Biotechnol**, v. 23, 2012.

BECKER, M. et al. Glutamate efflux mediated by Corynebacterium glutamicum MscCG, Escherichia coli MscS, and their derivatives. **Biochimica et Biophysica Acta - Biomembranes**, v. 1828, n. 4, p. 1230–1240, 2013.

BELTRAO, P. et al. Evolution and functional cross-talk of protein post-translational modifications. **Molecular Systems Biology**, v. 9, p. 714, 2013.

BENDT, A. K. et al. Towards a phosphoproteome map of Corynebacterium glutamicum. **Proteomics**, v. 3, n. 8, p. 1637–1646, 2003.

BEYNON, R. J. et al. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. **Nature Methods**, v. 2, n. 8, p. 587–589, 2005.

BLUM, H.; HILDBURG, B.; GROSS, H. J. Improved silver staining of plant proteins , RNA and DNA in polyacrylamide gels. p. 93–99, 1987.

BOTT, M.; BROCKER, M. Two-component signal transduction in Corynebacterium glutamicum and other corynebacteria: On the way towards stimuli and targets. **Applied Microbiology and Biotechnology**, v. 94, n. 5, p. 1131–1150, 2012.

BOULAHYA, K. A. et al. OdhI dephosphorylation kinetics during different glutamate production processes involving Corynebacterium glutamicum. **Applied Microbiology and Biotechnology**, v. 87, n. 5, p. 1867–1874, 2010.

BURSCHOWSKY, D. et al. Inter-Enzyme Allosteric Regulation of Chorismate Mutase in Corynebacterium glutamicum: Structural Basis of Feedback Activation by Trp. **Biochemistry**, v. 57, n. 5, p. 557–573, 2018.

BUSCHKE, N. et al. Metabolic engineering of industrial platform microorganisms for biorefinery applications - Optimization of substrate spectrum and process robustness by rational and evolutive strategies. **Bioresource Technology**, v. 135, p. 544–554, 2013.

CAIN, J. A.; SOLIS, N.; CORDWELL, S. J. Beyond gene expression: The impact of protein post-translational modifications in bacteria. **Journal of Proteomics**, v. 97, p. 265–286, 2014.

CATHERMAN, A. D. et al. Top down proteomics of human membrane proteins from

enriched mitochondrial fractions. **Analytical Chemistry**, v. 85, n. 3, p. 1880–1888, 2013a.

CATHERMAN, A. D. et al. Large-scale Top-down Proteomics of the Human Proteome: Membrane Proteins, Mitochondria, and Senescence. **Molecular & Cellular Proteomics**, v. 12, n. 12, p. 3465–3473, 2013b.

CATHERMAN, A. D.; SKINNER, O. S.; KELLEHER, N. L. Top Down proteomics: Facts and perspectives. **Biochemical and Biophysical Research Communications**, v. 445, n. 4, p. 683–693, 2014.

CHAIT, B. T. Mass spectrometry: Bottom-up or top-down? **Science**, v. 314, n. 5796, p. 65–66, 2006.

CHAN, C. H. et al. In *Salmonella enterica*, the sirtuin-dependent protein acylation/deacylation system (SDPADS) maintains energy homeostasis during growth on low concentrations of acetate. **Molecular Microbiology**, v. 80, n. 1, p. 168–183, 2011.

EGGELING, L. **Handbook of Corynebacterium glutamicum**. [s.l: s.n.].

EGGELING, L.; BOTT, M. **Handbook of Corynebacterium glutamicum**. Cambridge: [s.n.]. v. 1

ELSHOLZ, A. K. W. et al. Global impact of protein arginine phosphorylation on the physiology of *Bacillus subtilis*. **Proceedings of the National Academy of Sciences**, v. 109, n. 19, p. 7451–7456, 2012.

FANOUS, A. et al. A proteome analysis of the cadmium and mercury response in *Corynebacterium glutamicum*. **Proteomics**, v. 8, n. 23–24, p. 4976–4986, 2008.

FANOUS, A. et al. *Corynebacterium glutamicum* as an indicator for environmental cobalt and silver stress—a proteome analysis. **Journal of environmental science and health. Part. B, Pesticides, food contaminants, and agricultural wastes**, v. 45, n. 7, p. 666–675, 2010.

FRIDRIKSSON, E. K.; BAIRD, B.; MCLAFFERTY, F. W. Electrospray mass spectra from protein electroeluted from sodium dodecylsulfate polyacrylamide gel electrophoresis gels. **Journal of the American Society for Mass Spectrometry**, v. 10, n. 5, p. 453–455, 1999.

FUHRMANN, J. et al. McsB Is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator ctsr. **Science**, v. 324, n. 5932, p. 1323–1327, 2009.

GEORGI, T.; RITTMANN, D.; WENDISCH, V. F. Lysine and glutamate production by *Corynebacterium glutamicum* on glucose, fructose and sucrose: Roles of malic enzyme and fructose-1,6-bisphosphatase. **Metabolic Engineering**, v. 7, n. 4, p. 291–301, 2005.

GRANGEASSE, C.; STÜLKE, J.; MIJAKOVIC, I. Regulatory potential of post-translational modifications in bacteria. **Frontiers in Microbiology**, v. 6, n. MAY, p. 2014–2015, 2015.

GUILLOUET, S. et al. Expression of the *Escherichia coli* catabolic threonine dehydratase in *Corynebacterium glutamicum* and its effect on isoleucine production. **Applied and Environmental Microbiology**, v. 65, n. 7, p. 3100–3107, 1999.

HENTCHEL, K. L.; ESCALANTE-SEMERENA, J. C. Acylation of Biomolecules in Prokaryotes: a Widespread Strategy for the Control of Biological Function and Metabolic Stress. **Microbiology and Molecular Biology Reviews**, v. 79, n. 3, p. 321–346, 2015.

HERMANN, T. Industrial production of amino acids by coryneform bacteria. **Journal of Biotechnology**, v. 104, p. 155–172, 2003.

HIRASAWA, T.; SHIMIZU, H. Recent advances in amino acid production by microbial cells. **Current Opinion in Biotechnology**, v. 42, p. 133–146, 2016.

HÜSER, A. T. et al. 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**, v. 71, n. 6, p. 3255–3268, 2005.

IKEDA, M.; NAKAGAWA, S. The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. **Appl Microbiol Biotechnol**, v. 62, 2003.

KACHUK, C.; DOUCETTE, A. A. The benefits (and misfortunes) of SDS in top-down proteomics. **Journal of Proteomics**, v. 175, p. 75–86, 2018.

KALINOWSKI, J. et al. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. **J Biotechnol**, v. 104, 2003a.

KALINOWSKI, J. et al. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. **Journal of Biotechnology**, v. 104, n. 1–3, p. 5–25, 2003b.

KANEHISA, M.; SATO, Y.; MORISHIMA, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. **Journal of Molecular Biology**, v. 428, n. 4, p. 726–731, 2016.

KATAOKA, M. et al. Gene expression of *Corynebacterium glutamicum* in response to the conditions inducing glutamate overproduction. **Letters in Applied Microbiology**, v. 42, n. 5, p. 471–476, 2006.

KAWAGUCHI, H. et al. Engineering of a xylose metabolic pathway in *Corynebacterium glutamicum*. **Appl Environ Microbiol**, v. 72, 2006.

KEILHAUER, C.; EGGELING, L.; SAHM, H. Isoleucine synthesis in *Corynebacterium glutamicum*: {molecular analysis of the *ilvB-ilvN-ilvC* operon.}. **J. Bacteriol.**, v. 175, n. 17, p. 5595–5603, 1993.

KELLIE, J. F. et al. Robust analysis of the yeast proteome under 50 kDa by molecular-mass-based fractionation and top-down mass spectrometry. **Analytical Chemistry**, v. 84, n. 1, p. 209–215, 2012.

KIM, J. et al. Effect of *odhA* overexpression and *odhA* antisense RNA expression on Tween-40-triggered glutamate production by *Corynebacterium glutamicum*. **Applied Microbiology and Biotechnology**, v. 81, n. 6, p. 1097–1106, 2009.

KIM, J. et al. Investigation of phosphorylation status of OdhI protein during penicillin-



and Tween 40-triggered glutamate overproduction by *Corynebacterium glutamicum*. **Applied Microbiology and Biotechnology**, v. 91, n. 1, p. 143–151, 2011.

KOBIR, A. et al. Protein phosphorylation in bacterial signal transduction. **Biochimica et Biophysica Acta - General Subjects**, v. 1810, n. 10, p. 989–994, 2011.

LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680–685, 1970.

LANOUILLE, S. et al. The functional diversity of protein lysine methylation. **Molecular Systems Biology**, v. 10, n. 4, p. 1–26, 2014.

LEDUC, R. D. et al. ProSight PTM: An integrated environment for protein identification and characterization by top-down mass spectrometry. **Nucleic Acids Research**, v. 32, n. WEB SERVER ISS., p. 340–345, 2004.

LEE, J. Y. et al. The actinobacterium *Corynebacterium glutamicum*, an industrial workhorse. **J Microbiol Biotechnol**, v. 26, n. 5, p. 807–822, 2016.

LEUCHTENBERGER, W.; HUTHMACHER, K.; DRAUZ, K. Biotechnological production of amino acids and derivatives: Current status and prospects. **Applied Microbiology and Biotechnology**, v. 69, n. 1, p. 1–8, 2005.

LI, P. P.; LIU, Y. J.; LIU, S. J. Genetic and biochemical identification of the chorismate mutase from *Corynebacterium glutamicum*. **Microbiology**, v. 155, n. 10, p. 3382–3391, 2009.

LIEBL, W.; KLAMER, R.; SCHLEIFER, K. H. Requirement of chelating compounds for the growth of *Corynebacterium glutamicum* in synthetic media. **Applied Microbiology and Biotechnology**, v. 32, n. 2, p. 205–210, 1989.

LITSANOV, B.; BROCKER, M.; BOTT, M. Toward homosuccinate fermentation: Metabolic engineering of *Corynebacterium glutamicum* for anaerobic production of succinate from glucose and formate. **Applied and Environmental Microbiology**, v. 78, n. 9, p. 3325–3337, 2012.

LIU, Q. et al. The impact of PHB accumulation on l-glutamate production by recombinant *Corynebacterium glutamicum*. **Journal of Biotechnology**, v. 132, n. 3, p. 273–279, 2007.

LUBITZ, D.; WENDISCH, V. F. Ciprofloxacin triggered glutamate production by *Corynebacterium glutamicum*. **BMC Microbiology**, v. 16, n. 1, p. 1–12, 2016.

LÜDKE, A. et al. A proteomic study of *Corynebacterium glutamicum* AAA+ protease FtsH. **BMC microbiology**, v. 7, p. 6, 2007.

MA, Q.; WOOD, T. K. Protein acetylation in prokaryotes increases stress resistance. **Biochemical and Biophysical Research Communications**, v. 410, n. 4, p. 846–851, 2011.

MAKAROV, A.; DENISOV, E. Dynamics of Ions of Intact Proteins in the Orbitrap Mass Analyzer. **Journal of The American Society for Mass Spectrometry**, 2009.

MATSUMOTO, K. et al. Production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) in recombinant *Corynebacterium glutamicum* using propionate as a precursor. **Journal**

of **Biotechnology**, v. 152, n. 4, p. 144–146, 2011.

MIZUNO, Y. et al. Altered acetylation and succinylation profiles in *Corynebacterium glutamicum* in response to conditions inducing glutamate overproduction. **MicrobiologyOpen**, v. 5, n. 1, p. 152–173, 2016.

MORBACH, S.; SAHM, H.; EGGELING, L. Use of feedback-resistant threonine dehydratases of *Corynebacterium glutamicum* to increase carbon flux towards L-isoleucine. **Applied and Environmental Microbiology**, v. 61, n. 12, p. 4315–4320, 1995.

MOSTOVENKO, E. et al. Comparison of peptide and protein fractionation methods in proteomics. **EuPA Open Proteomics**, v. 1, p. 30–37, 2013.

NESVIZHSKII, A. I.; AEBERSOLD, R. Interpretation of Shotgun Proteomic Data. **Molecular & Cellular Proteomics**, v. 4, n. 10, p. 1419–1440, 2005.

NEUMEYER, A. et al. Monitoring of population dynamics of *Corynebacterium glutamicum* by multiparameter flow cytometry. **Microbial Biotechnology**, v. 6, n. 2, p. 157–167, 2013.

NIEBISCH, A. et al. Corynebacterial protein kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the OdhI protein. **Journal of Biological Chemistry**, v. 281, n. 18, p. 12300–12307, 2006.

NTAI, I. et al. Applying label-free quantitation to top down proteomics. **Analytical Chemistry**, v. 86, n. 10, p. 4961–4968, 2014.

OKINO, S.; INUI, M.; YUKAWA, H. Production of organic acids by *Corynebacterium glutamicum* under oxygen deprivation. **Applied Microbiology and Biotechnology**, v. 68, n. 4, p. 475–480, 2005.

PARK, H.-S. et al. Transcriptomic analysis of *Corynebacterium glutamicum* in the response to the toxicity of furfural present in lignocellulosic hydrolysates. **Proc Biochem**, v. 50, 2015.

PEKAR, T.; ZABROUSKOV, V.; MAKAROV, A. Utilizing a Hybrid Mass Spectrometer to Enable Fundamental Protein Characterization : Intact Mass Analysis and Top-Down Fragmentation with the LTQ Orbitrap MS. 2010.

PIATKOV, K. et al. Formyl-methionine as a degradation signal at the N-termini of bacterial proteins. **Microbial Cell**, v. 2, n. 10, p. 376–393, 2015.

PISITHKUL, T.; PATEL, N. M.; AMADOR-NOGUEZ, D. Post-translational modifications as key regulators of bacterial metabolic fluxes. **Current Opinion in Microbiology**, v. 24, p. 29–37, 2015.

POLEN, T. et al. Characterization of citrate utilization in *Corynebacterium glutamicum* by transcriptome and proteome analysis. **FEMS Microbiology Letters**, v. 273, n. 1, p. 109–119, 2007.

POLEVODA, B.; SHERMAN, F. N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. **Journal of Molecular Biology**, v. 325, n. 4, p. 595–622, 2003.

RICE, M. S.; DAHLQUIST, F. W. Sites of deamidation and methylation in Tsr, a bacterial chemotaxis sensory transducer. **Journal of Biological Chemistry**, v. 266, n. 15, p. 9746–9753, 1991.

ROGERS, L. D.; OVERALL, C. M. Proteolytic Post-translational Modification of Proteins: Proteomic Tools and Methodology. **Molecular & Cellular Proteomics**, v. 12, n. 12, p. 3532–3542, 2013.

SARAVIA, N. G. et al. Posttranslational Modification of Pili upon Cell Contact Triggers N. meningitidis Dissemination. **Science**, n. February, p. 778–782, 2011.

SCHUBERT, O. T. et al. Quantitative proteomics: Challenges and opportunities in basic and applied research. **Nature Protocols**, v. 12, n. 7, p. 1289–1294, 2017.

SENYILMAZ, D. et al. Regulation of mitochondrial morphology and function by stearoylation of TFR1. **Nature**, v. 525, n. 7567, p. 124–128, 2015.

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 and Biosystems Engineering**, v. 25, n. 5, p. 291–298, 2003.

SKINNER, O. S. et al. Native GELFrEE: A new separation technique for biomolecular assemblies. **Analytical Chemistry**, v. 87, n. 5, p. 3032–3038, 2015.

SMITH, KELLEHER, AND T. C. FOR T. D. P. Proteoform: a single term describing protein complexity. **Nature Methods**, v. 10, n. 3, p. 186–187, 2014.

SMITH, R. D. et al. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. **Proteomics**, v. 2, n. 5, p. 513–523, 2002.

STOREY, K. B. Comparative enzymology—new insights from studies of an “old” enzyme, lactate dehydrogenase. **Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular Biology**, v. 199, p. 13–20, 2016.

TAUCH, A.; BURKOVSKI, A. Molecular armory or niche factors: Virulence determinants of corynebacterium species. **FEMS Microbiology Letters**, v. 362, n. 23, p. 1–6, 2015.

TRAN, J. C.; DOUCETTE, A. A. Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: A proteome prefractionation tool. **Journal of Proteome Research**, v. 7, n. 4, p. 1761–1766, 2008.

TRYFONA, T.; BUSTARD, M. T. Fermentative production of lysine by *Corynebacterium glutamicum*: Transmembrane transport and metabolic flux analysis. **Process Biochemistry**, v. 40, n. 2, p. 499–508, 2005.

VALBUENA, N. et al. Morphological changes and proteome response of *Corynebacterium glutamicum* to a partial depletion of FtsI. **Microbiology**, v. 152, n. 8, p. 2491–2503, 2006.

VERDIN, E.; OTT, M. Acetylphosphate: A novel link between Lysine Acetylation and intermediary metabolism in bacteria. **Molecular Cell**, v. 51, n. 2, p. 132–134, 2013.

VOGES, R. et al. Absolute quantification of *Corynebacterium glutamicum* glycolytic and anaplerotic enzymes by QconCAT. **Journal of Proteomics**, v. 113, p. 366–377, 2015a.

VOGES, R. et al. Absolute quantification of *Corynebacterium glutamicum* glycolytic and anaplerotic enzymes by QconCAT. **Journal of Proteomics**, v. 113, p. 366–377, 2015b.

VOGES, R.; NOACK, S. Quantification of proteome dynamics in *Corynebacterium glutamicum* by <sup>15</sup>N-labeling and selected reaction monitoring. **Journal of Proteomics**, v. 75, n. 9, p. 2660–2669, 2012.

WEINERT, B. T. et al. Acetyl-Phosphate is a critical determinant of Lysine Acetylation in *E.coli*. **Molecular Cell**, v. 51, n. 2, p. 265–272, 2013.

WENDISCH, V. F. et al. Emerging *Corynebacterium glutamicum* systems biology. **Journal of Biotechnology**, v. 124, n. 1, p. 74–92, 2006.

WESSEL, D.; FLÜGGE, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. **Analytical Biochemistry**, v. 138, n. 1, p. 141–143, 1984.

WIŚNIEWSKI, J. R. et al. Universal sample preparation method for proteome analysis. **Nature Methods**, v. 6, n. 5, p. 359–362, 2009.

XIE, F. et al. Liquid chromatography-mass spectrometry-based quantitative proteomics. **Journal of Biological Chemistry**, v. 286, n. 29, p. 25443–25449, 2011.

YAMAMOTO, S. et al. Overexpression of genes encoding glycolytic enzymes in *Corynebacterium glutamicum* enhances glucose metabolism and alanine production under oxygen deprivation conditions. **Applied and Environmental Microbiology**, v. 78, n. 12, p. 4447–4457, 2012.

YAMAMOTO, S. et al. Strain optimization for efficient isobutanol production using *Corynebacterium glutamicum* under oxygen deprivation. **Biotechnology and Bioengineering**, v. 110, n. 11, p. 2938–2948, 2013.

YANG, J.; YANG, S. Comparative analysis of *Corynebacterium glutamicum* genomes: a new perspective for the industrial production of amino acids. **BMC Genomics**, v. 18, n. S1, p. 940, 2017.

YATES, J. R.; KELLEHER, N. L. Top down proteomics. **Analytical Chemistry**, v. 85, n. 13, p. 6151, 2013.

ZHANG, D. et al. Reducing lactate secretion by *ldhA* Deletion in L-glutamate-producing strain *Corynebacterium glutamicum* GDK-9. **Brazilian Journal of Microbiology**, v. 45, n. 4, p. 1477–1483, 2014.

ZHANG, H. et al. Understanding the high l-valine production in *Corynebacterium glutamicum* VWB-1 using transcriptomics and proteomics. **Scientific Reports**, v. 8, n. 1, p. 3632, 2018a.

ZHANG, M. et al. Systematic Proteomic Analysis of Protein Methylation in Prokaryotes and Eukaryotes Revealed Distinct Substrate Specificity. **Proteomics**, v. 18, n. 1, p. 1–13, 2018b.

ZIMMER, J. S. D. et al. Advances in proteomics data analysis and display using an accurate mass and time tag approach. **Mass Spectrometry Reviews**, v. 25, n. 3, p. 450–482, 2006.