Avaliação de Sistemas de Cromatografia Líquida Uni e Bidimensional Acoplados a Espectrometria de Massas na Análise do Proteoma dos Corpos Protéicos do Milho

Rogério de Campos Bicudo

Tese apresentada ao Instituto de Química de São Carlos, da Universidade de São Paulo, para a obtenção de título de Doutor em Ciências (Química Analítica).

Orientador: Prof. Dr. Fernando M. Lanças

São Carlos
2007

DEDICATÓRIA

Dedico esta tese, primeiramente, a Deus. Dedico ainda, a toda minha família, em especial aos meus pais **Elias** e **Eliana** pelo apoio, incentivo, amor e exemplo de vida, a minha irmã **Tatiana** e a você, **Ana Paula**.

AGRADECIMENTOS

- A Deus, pela constante presença em minha vida.
- A minha família, pelo apoio e incentivo.
- Agradeço especialmente a minha irmã, Tatiana, pelo companheirismo e pela ajuda no trabalho.
- A Ana Paula, pela paciência, carinho e amor.
- Ao meu “irmão” Fábio D. Vieira, pela amizade verdadeira.
- Ao Prof. Dr. Fernando M. Lanças, pela orientação, dedicação e confiança.
- Ao Dr. Colnago e a Dr^a Lucimara, pela parceria e pelas sugestões.
- Ao amigo José Carlos, pela inestimável ajuda e companheirismo.
- A todos os amigos, que de alguma forma, direta ou indiretamente, contribuíram para que esta tese pudesse ser concluída.
- A Embrapa Instrumentação Agropecuária e a Embrapa Pecuária Sudeste, pelo fornecimento das amostras.
- Ao Laboratório de Espectrometria de Massas do Laboratório Nacional de Luz Síncrotron (MAS-LNLS), por permitir a realização dos experimentos.
- Ao Laboratório de Cromatografia do IQSC/USP, pelo apoio institucional.
- A FAPESP, pelo apoio financeiro.

SUMÁRIO

<i>Capítulo I</i>	13
<i>Introdução</i>	13
1. Justificativas do Trabalho	14
1.1. Aspectos Gerais do Milho	14
1.2. Aprimoramento Nutricional	15
1.3. Proteoma dos Corpos Protéicos	17
2. Objetivos	19
3. Revisão da Literatura	20
3.1. As Proteínas	20
3.1.1. Por que estudar as proteínas?	20
3.1.2. Composição e estrutura das proteínas	21
3.2. Genoma e Proteoma	25
3.3. Proteômica	27
3.3.1. Eletroforese bidimensional em gel de poliacrilamida (2D PAGE)	29
3.3.2. Cromatografia líquida (LC)	31
3.3.3. Cromatografia líquida de fase reversa (RP LC)	32
3.3.4. Cromatografia de troca iônica	35
3.3.5. Cromatografia líquida bidimensional	37
3.3.6. Espectrometria de massas	40
3.3.7. Ionização por "electrospray"	41
3.3.7. Analisadores de massas	46
3.3.8. Espectrometria de massas em "tandem"	49
3.3.9. Identificação de proteínas por espectrometria de massas	52
<i>Capítulo II</i>	58
<i>Parte Experimental</i>	58
1. LC-MS/MS	59
1.1. Preparação das Amostras	59
1.1.1. Reagentes	59
1.1.2. Obtenção dos corpos protéicos do milho	60
1.1.3. Digestão das proteínas	60
1.2. Equipamentos e Condições Experimentais	61
1.2.1. Cromatografia de fase reversa	61
1.2.2. Espectrometria de massas	63
1.3. Tratamento dos Dados	64
1.3.1. Pesquisa no banco de dados	64
1.3.2. Validação das identificações	65
2. LC/LC-MS/MS	66
2.1. Preparação das Amostras	66
2.2. Equipamentos e Condições Experimentais	67
2.2.1. Cromatografia de troca catiônica forte	67
2.2.2. Cromatografia de fase reversa	70
2.2.3. Espectrometria de massas	70
2.3. Tratamento dos Dados	70
2.3.1. Pesquisa no banco de dados	70
2.3.2. Validação das identificações	71
3. Esquema Geral	72

<i>Capítulo III</i>	74
<i>Resultados e Discussões</i>	74
1. LC-MS/MS	75
2. LC/LC-MS/MS	79
3. Comparação entre os Métodos LC-MS/MS e LC/LC-MS/MS	86
3.1. Número de Proteínas e Peptídeos Identificados	86
3.2. Qualidade do Espectro de MS/MS	88
3.3. Comparação de Parâmetros	90
4. Proteínas Identificadas	93
4.1. Zeínas	93
4.2. Seqüência das Proteínas Identificadas	94
4.3. Proteínas com mais de uma Identificação Potencial	101
<i>Capítulo IV</i>	102
<i>Conclusões</i>	102
<i>Capítulo V</i>	105
<i>Referências Bibliográficas</i>	105

LISTA DE FIGURAS

Figura 1: Estágios do desenvolvimento dos corpos protéicos de milho. Inicialmente, essas organelas são formadas apenas pelas zeínas β e γ as quais permanecem na periferia das mesmas (A). Em seguida, há um acúmulo das zeínas α e δ na região central (B, C e D), provocando um aumento no diâmetro dessas estruturas (1-2 μ m). ⁴	14
Figura 2: Fórmula estrutural geral de um aminoácido.	21
Figura 3: Os 20 aminoácidos mais comuns.	22
Figura 4: Formação da ligação peptídica (em vermelho) através da eliminação de uma molécula de água.	23
Figura 5: A sequência de aminoácidos da cadeia de polipeptídeos de uma proteína é chamada de estrutura primária. Diferentes regiões dessa sequência formam uma estrutura secundária regular local, como a hélice alfa (α) ou folha beta (β). A estrutura terciária é formada pelo empacotamento de tais elementos estruturais em uma ou várias unidades globulares compactas chamadas domínios. A proteína final pode conter várias cadeias de polipeptídeos organizadas em uma estrutura quaternária. ²⁰	24
Figura 6: Formação da proteína. As células são as unidades fundamentais de funcionamento de todo sistema vivo; nelas estão os genes que possuem as instruções para a produção das proteínas. ²³	25
Figura 7: Análise esquemática do proteoma. A entrada do sistema é o proteoma em toda sua complexidade. Proteínas e peptídeos são separados e analisados pelas ferramentas listadas. Os dados brutos, obtidos pela associação MS/MS, são posteriormente processados por um software para produção de informação a respeito da identidade das proteínas detectadas. ³²	28
Figura 8: Metodologia usada para a realização da eletroforese bidimensional em gel de poliacrilamida (2D PAGE). Após a aplicação da amostra, o gel de poliacrilamida é submetido a um campo elétrico para separação bidimensional. Na primeira dimensão, a separação ocorre de acordo com os pontos isoelétricos das proteínas (focalização isoelétrica). Já na segunda dimensão, a separação se dá de acordo com suas massas moleculares. Em seguida, proteínas individuais são recortadas do gel e fragmentadas (com tripsina, por exemplo). Os peptídeos gerados da digestão enzimática são analisados por espectrometria de massas. ³⁴	30
Figura 9: O polipeptídeo entra na coluna com a fase móvel. A parte hidrofóbica do polipeptídeo adsorve na superfície hidrofóbica da fase reversa e, ali permanece, até que a concentração de solvente orgânico atinja o valor crítico e desorva o polipeptídeo.	34
Figura 10: A retenção de moléculas pequenas, como a bifenil, diminui gradualmente a medida que a concentração de solvente orgânico aumenta, pois a bifenil é retida por partição. Já a retenção de polipeptídeos, como a lisozima, muda repentinamente e drasticamente quando a concentração de solvente orgânico atinge o valor crítico necessário para desorver o polipeptídeo. ⁴⁷	35
Figura 11: Cadeias laterais básicas protonadas.	37
Figura 12: Separação de uma mistura de peptídeos por cromatografia de troca catiônica em linha com cromatografia de fase reversa. ⁵³	39
Figura 13: Esquema de uma fonte de “electrospray” e do processo de formação do “spray”. A região I está sob pressão atmosférica, a região II, sob pressão intermediária e a região III, sob alto vácuo. ^{54,58,59}	42
Figura 14: Esquema do processo de geração de micro gotas por “electrospray”. As setas brancas representam a evaporação do solvente, e o tempo indicado sobre elas é aquele necessário para a redução do diâmetro até o valor no qual ocorre a liberação das micro gotas (seta preta). O número, dentro das gotas maiores, indica a quantidade de cargas em excesso. Já o valor numérico ao lado das gotículas representa o número de cargas de cada uma delas. Portanto, o resultado da fissão é sempre uma gota principal e várias outras de menor tamanho. O esquema A, em amarelo, mostra que as gotas menores também sofrem o processo de fissão. Nesse esquema, há a geração de gotículas com diâmetro em torno de 6nm e com duas cargas em excesso. Tomando-se, como exemplo, o diâmetro de um íon sulfato hidratado que é da ordem de 0,4nm, pode-se perceber que as gotas já atingem o nível molecular, após poucas fissões. No esquema B, também em amarelo, é apresentado o processo de deformação e fissão de uma gota, com excesso de carga, numa região de alto campo elétrico. Em I, têm-se uma gota, com excesso de carga, que acabou de ser inserida em uma região com um campo elétrico. Já em II, a gota começa a se deformar e o excesso de carga, antes distribuído uniformemente sobre sua superfície, começa a se concentrar na direção estabelecida pelo campo. Em III, observa-se uma significativa deformação da região da gota de maior densidade de carga. Finalmente, em IV, há o aumento da deformação e a conseqüente liberação das gotículas, que possuem maior densidade de carga que a gota original. O tempo que decorre até a produção dos íons, na fase gasosa, é da ordem de microssegundos. ^{54,59,60}	44

Figura 15: Esquema de um filtro de massas quadrupolar. Os íons entram e viajam na direção z enquanto oscilam no plano x-y. Tal oscilação é controlada pelos potenciais DC (U) e RF (V) aplicados em cada par de barras. Somente atravessarão o quadrupolo e serão detectados, os íons que tiverem trajetórias estáveis sob os valores de U e V aplicados.⁶³ 48

Figura 16: Diagrama esquemático de um analisador de massas do tipo TOF.³⁶ 49

Figura 17: Diagrama esquemático do funcionamento básico de um triplo quadrupolo. Inicialmente, com o uso do primeiro quadrupolo (Q1), faz-se uma varredura de todos os íons existentes na amostra. Em seguida, um íon da primeira varredura é selecionado. Após isso, o primeiro quadrupolo permite que apenas o íon escolhido entre na cela de colisão (q2). Na cela, tal íon é fragmentado mediante colisão com gás (em geral, argônio ou até mesmo ar). Finalmente, o segundo quadrupolo (Q3) faz a varredura dos íons (chamados de íons filhos) gerados pela colisão com o gás. 50

Figura 18: Esquema do espectrômetro de massas comercial Q-TOF, da Waters/Micromass. Seu funcionamento é similar ao funcionamento básico do triplo quadrupolo, descrito na Figura 17. A diferença reside no fato de que, no Q-TOF, é no analisador do tipo tempo de voo que ocorre a separação dos íons, gerados pela colisão com o gás. 51

Figura 19: Foto (A) e esquemas representativos (B e C) do sistema cromatográfico “capLC”, acoplado ao espectrômetro de massas, da Waters/Micromass.⁶⁷ (B) A válvula se encontra na posição de lavagem da amostra, a qual está retida na pré-coluna. (C) A válvula se encontra na posição de análise. A fase móvel empregada nas análises cromatográficas foi: A – Água/0,1%Ácido Fórmico, B – Acetonitrila/0,1%Ácido Fórmico e C – Água. O volume de injeção foi de 6,4µl e o fluxo na coluna analítica foi de 200nl/min. As colunas cromatográficas, de modelo Waters Symmetry 300, tiveram as seguintes especificações: pré-coluna de fase reversa (C₁₈) - 5mmx0,350mm e d_p (diâmetro das partículas) de 5µm; coluna analítica de fase reversa (C₁₈) - 150mmx0,075mm e d_p (diâmetro das partículas) de 3µm. Utilizou-se o seguinte gradiente: 10min – 10%B, 70min – 45%B, 80min – 60%B, 85min – 10%B; a etapa de lavagem da amostra com o solvente C, a 10µl/min, durou 10 minutos. 62

Figura 20: Esquema representativo do sistema “on-line” de LC-MS/MS.⁶⁷ O sistema capLC (A), com a válvula na posição de análise, está diretamente acoplado ao espectrômetro de massas do tipo ESI-Q/TOF (B). As análises por espectrometria de massas foram feitas no modo positivo usando os seguintes parâmetros: voltagem do cone 35V, voltagem do capilar 3500V, temperatura da fonte 100°C, varredura (modo MS) 200-2500 (m/z) e varredura (modo MS/MS) 50-4000 (m/z). O espectrômetro foi operado no modo DDA (“Data Direct Analysis”) segundo o qual os peptídeos foram selecionados automaticamente para o modo MS/MS baseado na carga (+2, +3 e +4) e na intensidade (os três íons mais intensos detectados no modo MS). Empregou-se argônio como gás de colisão. 63

Figura 21: Esquema geral resumido das análises realizadas. 73

Figura 22: Cromatogramas (BPI) dos peptídeos selecionados para o modo MS/MS; as funções 2, 3 e 4 são as responsáveis pela seleção automática dos peptídeos, para o modo MS/MS, com base na carga (+2, +3 e +4) e na intensidade (os três íons mais intensos detectados no modo MS - função 1). Os peptídeos são derivados das proteínas, digeridas com tripsina, dos corpos protéicos do milho BR 473. 76

Figura 23: Cromatograma de SCX com indicação das sete frações produzidas da amostra de peptídeos dos corpos protéicos digeridos do milho BR 473. Para cada fração são mostrados os cromatogramas (BPI) dos peptídeos selecionados para o modo MS/MS do espectrômetro de massas. 80

Figura 24: Número de proteínas identificadas em cada um dos dois métodos empregados: LC-MS/MS e LC/LC-MS/MS. No método LC-MS/MS, foram identificadas 6 proteínas no total sendo que 5 delas também foram identificadas no outro método. Já no método LC/LC-MS/MS, foram identificadas 16 proteínas no total. A área de interseção dos círculos indica o número de proteínas identificadas comuns nos dois métodos. No total, foram identificadas 17 proteínas únicas considerando-se ambos os métodos. 86

Figura 25: Número de peptídeos identificados em cada um dos dois métodos empregados: LC-MS/MS e LC/LC-MS/MS. No método LC-MS/MS, foram identificados 19 peptídeos no total sendo que 17 deles também foram identificados no outro método. Já no método LC/LC-MS/MS, foram identificados 55 peptídeos no total. A área de interseção dos círculos indica o número de peptídeos identificados comuns nos dois métodos. No total, foram identificados 57 peptídeos únicos considerando-se ambos os métodos. 87

Figura 26: Comparação entre os espectros de MS/MS, do peptídeo “K.DVYDNVAHAFVK.I” de m/z 689,27 (2+), gerados pelos métodos de LC-MS/MS e LC/LC-MS/MS. 89

Figura 27: Comparação entre os métodos LC-MS/MS e LC/LC-MS/MS, considerando-se os valores normalizados, para todas as proteínas identificadas, dos seguintes parâmetros: o “score”, a porcentagem de cobertura da proteína (%CP) e o número de íons “b” e “y” (b+y). 91

Figura 28: Contribuição de cada proteína única identificada, pelos métodos de LC-MS/MS e LC/LC-MS/MS, para o alto conteúdo de lisina e triptofano da variedade de milho BR 473. 100

LISTA DE TABELAS

<i>Tabela 1: Superioridade nutricional da variedade de milho BR 473 em relação ao milho comum¹⁵</i>	17
<i>Tabela 2: Peptídeos com a mesma razão massa/carga (para carga igual a um), mas com seqüência de aminoácidos e fonte de origem diferentes⁵³</i>	55
<i>Tabela 3: Algumas ferramentas (softwares), que usam dados de espectrometria de massas, para identificação de proteínas via pesquisa em banco de dados⁶⁴</i>	57
<i>Tabela 4: Reagentes utilizados na parte experimental</i>	59
<i>Tabela 5: Parâmetros empregados para a pesquisa no banco de dados do NCBI, com o uso do “software” Mascot, utilizando os dados resultantes das análises por LC-MS/MS</i>	64
<i>Tabela 6: Especificações do cromatógrafo líquido “2796 Alliance Bio”, da empresa Waters⁶⁷, usado para as análises de cromatografia de troca catiônica forte</i>	67
<i>Tabela 7: Gradiente usado para a análise, por SCX, do digerido protéico da variedade de milho BR 473</i>	68
<i>Tabela 8: Reagrupamento das frações geradas pela cromatografia de troca catiônica forte e respectivos intervalos de tempo</i>	69
<i>Tabela 9: Parâmetros empregados para a pesquisa no banco de dados do NCBI, com o uso do “software” Mascot, utilizando os dados resultantes das análises por LC/LC-MS/MS</i>	71
<i>Tabela 10: Informações sobre as proteínas identificadas, da variedade de milho BR 473, usando o método de LC-MS/MS</i>	77
<i>Tabela 11: Informações sobre as proteínas identificadas, da variedade de milho BR 473, usando o método de LC/LC-MS/MS</i>	81
<i>Tabela 12: Aminoácidos que compõem as 17 proteínas únicas identificadas pelos métodos de LC-MS/MS e LC/LC-MS/MS. Nos casos em que houve mais de uma identificação potencial (adenine nucleotide translocator, eEF1A e GAPC3/GAPC4), segundo a Tabela 11, foi considerada apenas uma (gil22162, gil1321656 e gil6166167, respectivamente), devido a grande semelhança de seqüência entre elas. Abreviatura: N°AA – número de aminoácidos</i>	95

LISTA DE ABREVIATURAS

eEF1A: "alpha subunit of translation elongation factor 1"

ESI: "electrospray ionization"

ESI-Q/TOF: "electrospray ionization - quadrupole/time of flight"
(espectrometria de massas seqüencial com ionização do tipo electrospray e analisadores de massas do tipo quadrupolo e tempo de vôo)

LC: "liquid chromatography"

LC/LC: "liquid chromatography/liquid chromatography" (cromatografia líquida bidimensional)

LC/LC-MS/MS: "liquid chromatography/liquid chromatography - mass spectrometry/mass spectrometry" (acoplamento entre cromatografia líquida bidimensional e espectrometria de massas seqüencial)

LC-MS/MS: "liquid chromatography - mass spectrometry/mass spectrometry" (acoplamento entre cromatografia líquida unidimensional e espectrometria de massas seqüencial)

MS: "mass spectrometry"

MS/MS: "mass spectrometry/mass spectrometry" (espectrometria de massas seqüencial)

NCBI: "national center for biotechnology information"

Q: "quadrupole"

QPM: "quality protein maize"

Q/TOF: "quadrupole/time of flight" (acoplamento entre os analisadores de massas quadrupolo e tempo de vôo)

RP LC: "reversed phase liquid chromatography"

SCX: "strong cation exchange"

TOF: "time of flight"

Resumo

O valor nutricional do milho é limitado devido ao seu alto conteúdo de zeínas (proteínas de reserva), as quais são deficientes em aminoácidos essenciais como lisina e triptofano. Por esse motivo, a Embrapa (Empresa Brasileira de Pesquisa Agropecuária) desenvolveu a variedade de milho BR 473, a qual tem um excelente valor energético e um maior conteúdo protéico quando comparado ao milho comum. Tal variedade possui 0,9 e 4,0 g/kg de grão de triptofano e lisina, respectivamente, contra 0,6 e 2,6 do milho comum. O alto conteúdo de lisina em algumas variedades de milho como a opaco-2 (o2), por exemplo, é relacionado à presença de proteínas não-zeínas. Para identificar essas proteínas, o que poderia explicar o alto conteúdo protéico da variedade BR 473, foi analisado o proteoma dos corpos protéicos dos grãos desse milho. Para tal, foram usadas e avaliadas as cromatografias uni e bidimensional, acopladas a espectrometria de massas (LC-MS/MS e LC/LC-MS/MS), uma vez que tais técnicas se tornaram ferramentas poderosas para seqüenciamento e identificação de proteínas. Neste trabalho, foram identificadas seis proteínas por LC-MS/MS e dezesseis proteínas por LC/LC-MS/MS.

Abstract

The nutritional value of maize seed is limited due to its high content of zeins (storage proteins), which are deficient in essential amino acids as lysine and tryptophan. Because of this, Embrapa (Brazilian Agricultural Research Corporation) developed the BR 473 maize variety, which has excellent energetic value and higher proteic content than the normal maize. BR 473 maize variety has 0.9 and 4.0 g/kg of grain of tryptophan and lysine, respectively, against the 0.6 and 2.6 from normal maize. The high lysine content in some maize varieties as *opaque-2* (*o2*), for example, has been related to the presence of non-zein proteins. In order to identify them, which could explain the high proteic content of BR 473 maize variety, the protein bodies proteome of these grains were analyzed. For this purpose, one and two-dimensional liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS and LC/LC-MS/MS) were used and evaluated, since these techniques have become a powerful tool for protein sequencing and identification. We have identified six proteins by LC-MS/MS and sixteen proteins by LC/LC-MS/MS.

Capítulo I

Introdução

1. Justificativas do Trabalho

1.1. Aspectos Gerais do Milho

O milho (*Zea mays*) é uma planta da família das gramíneas (*Gramineae*), cultivada em diversas regiões do mundo. Ele é um cereal altamente energético empregado na alimentação humana e animal.

Em geral, o grão de milho é constituído de 70-80% de amido, 8-12% de proteínas, 3-6% de óleo e 1-4% de açúcares solúveis. As proteínas podem ser classificadas como não-zeínas (20-50%) e zeínas (50-80%). De acordo com a solubilidade e homologia de sequência, as zeínas são classificadas em α , β , γ e δ . Após sua síntese, as zeínas são depositadas, em diferentes padrões temporais e espaciais, em organelas esféricas denominadas corpos protéicos, através de um mecanismo ainda pouco compreendido¹⁻³ (Figura 1)⁴.

Figura 1: Estágios do desenvolvimento dos corpos protéicos de milho. Inicialmente, essas organelas são formadas apenas pelas zeínas β e γ , as quais permanecem na periferia das mesmas (A). Em seguida, há um acúmulo das zeínas α e δ na região central (B, C e D), provocando um aumento no diâmetro dessas estruturas (1-2 μ m).⁴

1.2. Aprimoramento Nutricional

O valor nutricional do milho é determinado, principalmente, pela composição de aminoácidos das suas proteínas de reserva (zeínas), cuja função biológica principal é servir como fonte de nitrogênio para o embrião durante a germinação das sementes.¹ As zeínas contêm grande quantidade de ácido glutâmico e prolina, mas são deficientes em aminoácidos essenciais, principalmente lisina e triptofano, sendo a lisina o mais limitante.^{1,5-6} Essa deficiência aliada à alta concentração das zeínas nos grãos de milho, faz com que este cereal possua baixa qualidade nutricional.

Diante desse problema, novas técnicas de aprimoramento nutricional foram desenvolvidas como, por exemplo, a técnica que consiste na redução da síntese de zeínas nos grãos promovida por genes mutantes. Como exemplo dessas mutações, está o popularmente conhecido opaco-2 (o2), o qual apresenta um aumento no conteúdo de lisina devido à redução da síntese de zeínas. As sementes do opaco-2 contêm cerca de 50% menos zeínas do que os milhos normais.⁷⁻⁸

A descoberta do mutante opaco-2 no milho⁹ e de outros genes mutantes para proteínas de alta qualidade nutricional, despertou inicialmente um acentuado entusiasmo entre pesquisadores no mundo todo. No entanto, esse entusiasmo foi reduzido à medida em que o milho opaco-2 não conseguiu aceitação no mercado; isso ocorreu devido a sua inferioridade agrônômica, a qual estava relacionada à baixa produtividade

de grãos, ao fenótipo do grão (farináceo e de baixa densidade), à vulnerabilidade a insetos e organismos causadores de podridões das espigas, a menor velocidade de secagem e maior incidência de danos mecânicos aos grãos, por ocasião da colheita.^{7,10-12}

Embora o interesse mundial pela pesquisa com os milhos opaco-2 tivesse declinado, o Centro Internacional de Melhoramento de Milho e Trigo (CIMMYT - Centro Internacional de Mejoramiento de Maíz y Trigo), do México, manteve esse interesse e conseguiu superar os principais problemas agrônômicos com a introdução de genes modificadores de endosperma e seleção contínua para grãos vítreos e de maior densidade. Surgiram, então, as variedades de milho denominadas "Quality Protein Maize" ou simplesmente QPM, as quais reúnem as boas qualidades do milho normal, ou seja, grãos vítreos e de maior densidade, e com qualidade protéica similar ao milho opaco-2.¹²⁻¹⁴

No Brasil, os estudos sobre melhoramento genético de milho são relativamente recentes. Merece destaque o trabalho desenvolvido pela Embrapa Milho e Sorgo, a qual melhorou a qualidade protéica do milho e desenvolveu a variedade QPM conhecida como BR 473; ela é produtiva, de ciclo precoce, com grãos semiduros de cor amarelo-alaranjada, com excelente valor energético e, ao mesmo tempo, maior valor protéico (Tabela 1). Esses grãos têm a aparência e o sabor do milho comum e fornecem alimentos e rações mais nutritivas.¹⁵

Tabela 1: Superioridade nutricional da variedade de milho BR 473 em relação ao milho comum¹⁵

Aminoácidos Essenciais	Triptofano (g/kg de grão)	Lisina (g/kg de grão)
BR 473	0,9	4,0
Milho Comum	0,6	2,6

1.3. Proteoma dos Corpos Protéicos

Sabe-se que a proteína “alpha subunit of translation elongation factor 1 (eEF1A)” é rica em lisina (~10%) e que seu nível é duplicado nos mutantes. Além disso, há uma forte correlação entre sua concentração e o conteúdo total de lisina no grão. Deste modo, o mecanismo pelo qual as mutações aumentam o conteúdo de lisina é função do aumento dos níveis da eEF1A e, possivelmente, de outras proteínas não-zeínas relacionadas, além da redução da síntese de zeínas.

Além disso, espécies com altos níveis de eEF1A ou de lisina possuem um número maior de corpos protéicos, e estes são menores do que em espécies com baixos níveis de eEF1A ou de lisina. Como consequência, esse corpos protéicos possuem uma área superficial maior, o que aumenta a quantidade de proteínas não-zeínas, possivelmente contendo lisina, que podem estar agregadas as suas superfícies.

Esse fato sugere que o aprimoramento nutricional do milho deve se concentrar em espécies com altos níveis da proteína eEF1A ou de lisina e com o maior número de pequenos corpos protéicos, pois isso levará ao mais alto conteúdo de lisina possível.¹⁶

Assim, a continuidade da melhoria da qualidade nutricional do milho deve prosseguir com o estudo do proteoma dos corpos protéicos extraídos de espécies com as características descritas anteriormente. O objetivo é conseguir identificar proteínas não-zeínas, possivelmente agregadas aos corpos protéicos da espécie estudada.

As proteínas identificadas poderão, ou não, ser relacionadas ao melhor conteúdo protéico das espécies QPM, além de poderem ajudar na caracterização do mecanismo através do qual as mutações aumentam a síntese de proteínas não-zeínas contendo lisina.

Portanto, considerando-se a grande importância da continuidade do aprimoramento nutricional do milho, neste trabalho foi estudado o proteoma dos corpos protéicos da variedade BR 473, escolhida por apresentar altos níveis de lisina. Para esse estudo, foram empregadas técnicas de alta confiabilidade na identificação de proteínas, como as de LC-MS/MS e LC/LC-MS/MS.

2. Objetivos

O presente trabalho teve como objetivo principal o estudo do proteoma dos corpos protéicos da variedade de milho QPM ("Quality Protein Maize") conhecida como BR 473, usando duas técnicas de alta confiabilidade para análises proteômicas, LC-MS/MS e LC/LC-MS/MS.

Os objetivos específicos desse trabalho foram os seguintes:

- Desenvolver métodos para identificação de proteínas não-zeínas, dos corpos protéicos purificados da variedade de milho QPM ("Quality Protein Maize") conhecida como BR 473, usando a técnica de LC-MS/MS.
- Desenvolver métodos para identificação das mesmas proteínas descritas acima, usando a técnica de LC/LC-MS/MS.
- Comparar as duas técnicas de identificação de proteínas utilizadas.
- Relacionar as proteínas identificadas com o alto conteúdo protéico da variedade de milho QPM estudada.

3. Revisão da Literatura

3.1. As Proteínas

3.1.1. Por que estudar as proteínas?

O termo proteína vem do grego *proteios* e significa “da primeira classe”, ou seja, “a mais importante”. Tal palavra foi cunhada por Jons J. Berzelius em 1838 para salientar a importância dessa classe de moléculas.¹⁷

Proteínas são polímeros (não ramificados) de aminoácidos resultantes da tradução das informações genéticas contidas no DNA das células. Elas são as biomoléculas mais abundantes.¹⁸⁻²⁰

As proteínas exercem papéis cruciais (daí o porquê do seu estudo) em virtualmente todos os processos biológicos, tais como:

catálise enzimática: quase todas as enzimas conhecidas são proteínas (portanto, as proteínas são essenciais para a determinação do padrão de transformações químicas em sistemas biológicos)

transporte e armazenamento

proteção imunológica: os anticorpos são proteínas altamente específicas que reconhecem e se combinam com substâncias estranhas tais como vírus, bactérias e células de outros organismos

controle do crescimento e da diferenciação: em organismos superiores o crescimento e a diferenciação são controlados por fatores protéicos de crescimento.¹⁷

3.1.2. Composição e estrutura das proteínas

As proteínas são compostas de unidades estruturais básicas denominadas aminoácidos. Um aminoácido é constituído de um grupo α -amino (α -NH₂), um grupo α -carboxílico (α -COOH), um átomo de hidrogênio e um grupamento R, todos ligados a um carbono assimétrico, conhecido como carbono- α (C _{α}).¹⁷⁻¹⁹ A Figura 2 apresenta uma fórmula estrutural geral para um aminoácido (o grupamento R é a cadeia lateral que diferencia a série de vinte aminoácidos). Os 20 aminoácidos mais comuns estão listados na Figura 3.

Figura 2: Fórmula estrutural geral de um aminoácido.

Figura 3: Os 20 aminoácidos mais comuns.

Os aminoácidos ligam-se entre si através dos grupos α -amino e α -carboxílico, formando uma ligação peptídica pela eliminação de moléculas de água, como ilustra a Figura 4. Os grupos amínicos e carboxílicos “livres” dos peptídeos definem as extremidades N-terminal e C-terminal, respectivamente.¹⁹

Figura 4: Formação da ligação peptídica (em vermelho) através da eliminação de uma molécula de água.

As proteínas são sintetizadas nos ribossomos (organelas celulares), a partir de informações codificadas na molécula de ácido ribonucléico mensageiro (RNAm), produzida por transcrição (cópia) de um trecho do DNA (DNA → RNA → proteína). Para cada três nucleotídeos (as unidades das moléculas genéticas) presentes no RNAm, o ribossomo adiciona um dos 20 diferentes aminoácidos, formando uma cadeia linear polipeptídica (estrutura primária). Essas cadeias de aminoácidos passam por diferentes graus de agregação, dobramento e modificação química, até formar uma estrutura tridimensional (terciária e quaternária) - a forma ativa da proteína.²¹ A Figura 5 apresenta um resumo das estruturas das proteínas.

Figura 5: A sequência de aminoácidos da cadeia de polipeptídeos de uma proteína é chamada de estrutura primária. Diferentes regiões dessa sequência formam uma estrutura secundária regular local, como a hélice alfa (α) ou folha beta (β). A estrutura terciária é formada pelo empacotamento de tais elementos estruturais em uma ou várias unidades globulares compactas chamadas domínios. A proteína final pode conter várias cadeias de polipeptídeos organizadas em uma estrutura quaternária.²⁰

3.2. Genoma e Proteoma

O genoma, o qual consiste em todos os genes de um organismo, determina como os aminoácidos serão unidos para formar uma proteína.²² Após sua formação, a proteína pode atuar sozinha ou em complexos protéicos na execução das mais variadas funções biológicas (Figura 6).

Figura 6: Formação da proteína. As células são as unidades fundamentais de funcionamento de todo sistema vivo; nelas estão os genes que possuem as instruções para a produção das proteínas.²³

Alguns cientistas acreditam que o genoma não trará todas as respostas às questões biológicas, pois os processos orgânicos estão na dependência de outras substâncias, como as proteínas, as quais exercem papéis essenciais em virtualmente todos os processos biológicos, sendo elas que determinam o padrão de transformações químicas dentro das células.²⁵

Portanto, o estudo das proteínas (projeto proteoma) tornou-se a próxima e necessária etapa, depois do projeto genoma.

O termo proteoma foi cunhado por Wilkins e Williams como significando o conjunto de todas as *proteínas* expressas por um *genoma*.^{22,26} Portanto, o proteoma é definido como o conjunto das proteínas expressas numa célula, tecido ou organismo, e que intervêm nos processos biológicos de uma espécie.²⁶⁻²⁸

O proteoma apresenta, com certeza, muito maior dificuldade e extensão que o genoma. O velho dogma de que um gene tem o código de uma proteína já foi superado e, hoje, sabe-se que um gene pode, de uma maneira ou de outra, dar origem a muitas proteínas. Assim, há mais proteínas em um proteoma que genes em um genoma (o DNA consiste de quatro bases químicas: a adenina (A), a citosina (C), a guanina (G) e a timina (T); já as proteínas são compostas de aminoácidos, dos quais existem 20 tipos diferentes).^{22,29}

O genoma de um organismo individual é essencialmente estático e o DNA, do qual pode ser extraído em princípio de uma única célula, contém toda a informação. Em contraste, o proteoma observável de uma única célula é altamente dinâmico e dependente do estado da célula.³⁰ Portanto, o proteoma não é estacionário, mudando com o desenvolvimento do organismo e com qualquer mudança no seu ambiente.²²

3.3. Proteômica

A proteômica é o estudo do proteoma^{22,31} e tem como objetivo a identificação e a determinação das funções de todas as proteínas expressas numa célula, tecido ou organismo.

Para a identificação das proteínas, a proteômica faz uso de duas plataformas metodológicas que se complementam. A primeira plataforma tem como meta uma separação preliminar dos componentes a serem analisados na segunda plataforma. O principal objetivo da primeira plataforma é facilitar o poder de resolução da segunda plataforma. As técnicas mais utilizadas para essa primeira plataforma são a cromatografia líquida uni (1D ou LC) ou bidimensional (2D ou LC/LC) e a eletroforese bidimensional em gel de poliacrilamida (2D PAGE).

A segunda plataforma visa a caracterização dos compostos separados na primeira. A técnica ideal para a segunda plataforma é a espectrometria de massas em “tandem” (MS/MS) (Figura 7).²⁴

Figura 7: Análise esquemática do proteoma. A entrada do sistema é o proteoma em toda sua complexidade. Proteínas e peptídeos são separados e analisados pelas ferramentas listadas. Os dados brutos, obtidos pela associação MS/MS, são posteriormente processados por um software para produção de informação a respeito da identidade das proteínas detectadas.³²

3.3.1. Eletroforese bidimensional em gel de poliacrilamida (2D PAGE)

A eletroforese bidimensional em gel de poliacrilamida (2D PAGE) e a identificação de proteínas por espectrometria de massas representam uma tecnologia integrada, através da qual milhares de espécies de proteínas podem ser separadas, detectadas e quantificadas. Além disso, centenas de proteínas podem ser identificadas pela análise seqüencial da mistura de peptídeos gerada pela digestão de pontos individuais produzidos no gel, durante a etapa de separação.³³ Um diagrama esquemático ilustrando as etapas envolvidas na metodologia típica usada em eletroforese bidimensional, empregando gel de poliacrilamida (2D PAGE), é mostrada na Figura 8.

Figura 8: Metodologia usada para a realização da eletroforese bidimensional em gel de poliacrilamida (2D PAGE). Após a aplicação da amostra, o gel de poliacrilamida é submetido a um campo elétrico para separação bidimensional. Na primeira dimensão, a separação ocorre de acordo com os pontos isoelétricos das proteínas (focalização isoelétrica). Já na segunda dimensão, a separação se dá de acordo com suas massas moleculares. Em seguida, proteínas individuais são recortadas do gel e fragmentadas (com tripsina, por exemplo). Os peptídeos gerados da digestão enzimática são analisados por espectrometria de massas.³⁴

No entanto, um exame mais detalhado das classes de proteínas rotineiramente identificadas pela técnica de 2D PAGE-MS/MS sugere que ela não representa uma técnica definitiva e verdadeiramente global, pois, sabe-se que classes específicas de proteínas são excluídas ou sub-representadas no gel 2D. Estas incluem proteínas muito ácidas ou muito básicas, proteínas excessivamente grandes ou excessivamente pequenas e proteínas existentes na membrana celular. Além disso, foi mostrado que a abordagem 2D PAGE-MS/MS é incapaz de detectar proteínas com baixa abundância, sem um enriquecimento anterior a separação na placa de gel. Existe ainda o problema da co-migração de proteínas que provoca sérias dificuldades de identificação.^{32,35}

Concluindo, 2D PAGE-MS/MS não serve como tecnologia base para a análise do proteoma quando os mais variados tipos de proteínas celulares são de interesse (análise global).³²

3.3.2. Cromatografia líquida (LC)

Para minimizar algumas das limitações da eletroforese bidimensional em gel de poliacrilamida, a cromatografia líquida têm sido integrada com a espectrometria de massas, compondo uma nova plataforma para análise do proteoma.

A cromatografia líquida (LC) é uma técnica analítica de separação amplamente usada no estudo de substâncias de grande interesse para a indústria e para a ciência, como, por exemplo, aminoácidos, proteínas, ácidos nucleicos, carboidratos, drogas, pesticidas, etc.

Nas separações cromatográficas, a amostra é transportada por uma fase móvel (líquido), a qual é forçada através de uma fase estacionária imiscível fixa, colocada na coluna ou numa superfície sólida. Os componentes da amostra mais fortemente retidos na fase estacionária movem-se muito lentamente no fluxo da fase móvel, enquanto os que se ligam mais fracamente à fase estacionária movem-se mais rapidamente. Como consequência dessas diferentes mobilidades, ocorre uma separação dos componentes em bandas ou zonas discretas que podem, então, ser analisados.

Os componentes básicos de um cromatógrafo líquido são: reservatório de fase móvel, sistemas de bombeamento, sistemas de injeção de amostras, colunas cromatográficas e detectores.³⁶

3.3.3. Cromatografia líquida de fase reversa (RP LC)

A cromatografia líquida de fase reversa (RP LC) se tornou uma ferramenta bem estabelecida e largamente utilizada para análise e purificação de biomoléculas. O principal motivo que explica esse fato é a alta resolução a qual pode ser conseguida com tal técnica. A RP LC é

capaz de separar polipeptídeos de sequências muito próximas tanto de pequenos peptídeos quanto de grandes proteínas. Até mesmo polipeptídeos que diferem por apenas um resíduo de aminoácido podem ser separados.³⁷⁻³⁸

A cromatografia líquida de fase reversa pode ser usada para separação de peptídeos derivados de uma digestão enzimática³⁹⁻⁴² e para purificação de peptídeos naturais e sintéticos em grandes ou em pequenas quantidades (g ou mg).⁴³⁻⁴⁵

A separação de moléculas pequenas, em RP LC, envolve uma contínua partição das moléculas entre a fase móvel e a fase estacionária hidrofóbica. Os polipeptídeos, porém, são muito grandes para que haja partição e, por isso, eles adsorvem na superfície hidrofóbica, depois de entrar na coluna, e permanecem adsorvidos até que a concentração de solvente orgânico atinja um valor crítico, necessário para promover a dessorção. Após dessorverem, muito pouca interação acontece entre os polipeptídeos e a superfície da fase reversa, até que eles eluam da coluna (essas interações subsequentes tem pouco efeito na separação).⁴⁶

Somente uma parte da molécula do polipeptídeo, chamada de “pé hidrofóbico”, fica em contato com a superfície da fase reversa, enquanto que a maior parte da molécula fica exposta à fase móvel. A RP LC consegue separar os polipeptídeos devido a sutil diferença (que resulta da diferença na sequência de aminoácidos e da diferença na conformação) nas suas partes hidrofóbicas⁴⁷ (Figura 9).

Figura 9: O polipeptídeo entra na coluna com a fase móvel. A parte hidrofóbica do polipeptídeo adsorve na superfície hidrofóbica da fase reversa e, ali permanece, até que a concentração de solvente orgânico atinja o valor crítico e desorva o polipeptídeo.

Como o número de moléculas, chamado de número “Z” por Geng e Regnier⁴⁶, de solvente orgânico para dissolver um polipeptídeo é muito preciso, a desorção acontece dentro de uma faixa muito estreita de concentração de solvente orgânico. Isso significa que o polipeptídeo permanece completamente retido até que a concentração crítica de solvente orgânico seja atingida; quando isso acontece, repentinamente ocorre a desorção (Figura 10).⁴⁷

Figura 10: A retenção de moléculas pequenas, como a bifenil, diminui gradualmente a medida que a concentração de solvente orgânico aumenta, pois a bifenil é retida por partição. Já a retenção de polipeptídeos, como a lisozima, muda repentinamente e drasticamente quando a concentração de solvente orgânico atinge o valor crítico necessário para dessorver o polipeptídeo.⁴⁷

No caso de pequenos polipeptídeos, o mecanismo de retenção é uma mistura de partição e adsorção. Eles desorvem mais rapidamente, com mudanças na concentração de solvente orgânico, do que pequenas moléculas mas, dessorvem mais gradualmente do que as proteínas.⁴⁷

3.3.4. Cromatografia de troca iônica

A cromatografia de troca iônica é largamente usada para purificação de polipeptídeos, pois emprega condições não-desnaturantes e próximas às condições fisiológicas. Além disso, esse tipo de cromatografia, geralmente, possui alta capacidade de amostragem e estabilidade em uma larga faixa de pH (quando as resinas de troca são de base polimérica).

A cromatografia de troca iônica separa os polipeptídeos por interação eletrostática entre as cadeias laterais carregadas dos aminoácidos e a superfície, também carregada, da resina de troca. A retenção leva em conta a carga total do polipeptídeo no pH da fase móvel.⁴⁸

Entretanto, Kopaciewicz e colaboradores desenvolveram um mecanismo mais abrangente para a cromatografia de troca iônica de polipeptídeos. Eles mostraram que ocorre uma significativa retenção de polipeptídeos mesmo no seu pI (pH no qual a carga total é zero), e que a correlação entre a carga total e a retenção de polipeptídeos é geralmente pobre. Eles concluíram que a assimetria de carga melhor explica a retenção de polipeptídeos em cromatografia de troca iônica e que isso explica o fato de a estrutura terciária dos polipeptídeos afetarem sua retenção.⁴⁹

Resinas de troca iônica com caráter hidrofóbico podem resultar em um modo misto de separação (interação eletrostática e hidrofóbica). Isso é possível por causa dos espaços entre o substrato das partículas e o grupo responsável pela interação iônica. Para diminuir essa interação hidrofóbica, a qual pode causar alargamento de picos e desnaturação do polipeptídeo, solventes orgânicos são adicionados a fase móvel.^{48,50}

A cromatografia de troca iônica pode ser empregada tanto na troca de ânions (cromatografia de troca aniônica) como na troca de cátions (cromatografia de troca catiônica). No caso da cromatografia de troca

catiônica, a retenção dos polipeptídeos se dá pela interação dos grupos de ácidos sulfônicos, da superfície da resina de troca iônica, com aminoácidos como histidina, lisina e arginina. A fase móvel é tamponada para que o valor do pH fique abaixo de 6 ou 7 para manter as cadeias laterais básicas protonadas⁴⁸ (Figura 11).

Figura 11: Cadeias laterais básicas protonadas.

3.3.5. Cromatografia líquida bidimensional

Em caso de amostras de menor complexidade, o uso de cromatografia líquida unidimensional (1D ou LC) já é o suficiente. No entanto, quando as misturas são extremamente complexas há a necessidade de se utilizar a cromatografia bidimensional (2D ou LC/LC).^{32,51}

O critério mais rigoroso na análise multidimensional estabelece que as colunas acopladas devem ter diferentes seletividades, ou seja, um método ortogonal de separação.⁵² Um exemplo de atendimento a esse

critério seria um sistema bidimensional composto por cromatografia de troca catiônica (separação ocorre com base na carga dos polipeptídeos) e cromatografia de fase reversa (separação ocorre com base na hidrofobicidade dos polipeptídeos) (Figura 12).

A cromatografia de troca catiônica é mais favorável que cromatografia de fase reversa para a aplicação como primeira dimensão, uma vez que oferece uma alta capacidade da coluna e uma alta biorecuperação. Outro ponto importante é que o eluente da primeira coluna (1ª dimensão) deve ser compatível com as condições da separação na segunda coluna (2ª dimensão).

O efluente da coluna de troca catiônica é rico em sal. Desta forma, quando as frações dos efluentes da coluna de troca catiônica são transferidas para o topo da coluna de fase reversa, o sal deverá eluir juntamente com o volume morto dessa última coluna. Usualmente, o sal não atrapalha a separação na coluna de fase reversa. Porém, caso haja algum distúrbio uma coluna de “desalting” pode ser instalada entre as duas dimensões.

Figura 12: Separação de uma mistura de peptídeos por cromatografia de troca catiônica em linha com cromatografia de fase reversa.⁵³

3.3.6. Espectrometria de massas

A espectrometria de massas surgiu na primeira metade do século passado, e se tornou uma ferramenta analítica de extrema importância em pesquisa envolvendo a identificação e/ou quantificação de compostos orgânicos e inorgânicos. No entanto, somente com o desenvolvimento de técnicas de espectrometria de massas direcionadas para a análise de biomoléculas, passou a ser uma técnica analítica importante em pesquisas biológicas, como o proteoma. Ela é capaz de fornecer de modo rápido, preciso e sensível, a massa de compostos como proteínas e peptídeos.²⁴

Um espectrômetro de massas é um instrumento que separa íons em movimento, com base em suas razões massa/carga (m/z). Ele é constituído por uma fonte de íons, um separador ou filtro de massa/carga (m/z) e um detector.^{36,54}

A primeira etapa ou etapa de ionização, a qual ocorre na fonte de íons, era um grande problema no caso de moléculas biológicas, pois, nos primeiros métodos de ionização, as moléculas a serem analisadas deveriam estar em fase gasosa, alto vácuo e altas temperaturas, condições incompatíveis com a maioria das biomoléculas. Esse problema foi resolvido com o surgimento, na década de 1980, dos métodos brandos de ionização, os quais permitiram a produção de íons de moléculas não voláteis.^{24,36}

Uma desses métodos foi a ionização do tipo ESI, do inglês “electrospray ionization”, na qual a amostra, em solução, passa por um tubo capilar submetido à alta voltagem. Na saída do tubo há a formação de um processo de nebulização, e as moléculas, presas nas gotículas suspensas em vácuo, são ionizadas.

Essa técnica de ionização é largamente empregada, atualmente, e um de seus pesquisadores, John Fenn, foi ganhador do prêmio nobel de química de 2002, por tal desenvolvimento.²⁴

3.3.7. Ionização por “electrospray”

O “electrospray”, apesar de ser considerado um método de ionização, é, na verdade, um processo de transferência de íons, já existentes em solução, para a fase gasosa.⁵⁴

Em 1968, Dole sugeriu o “electrospray” como um método de ionização para espectrometria de massas.^{54,55} Seus experimentos, que visavam analisar espécies poliméricas, como poliestireno, que não estão ionizados em solução, não foram convincentes. Assim, em 1984, Yamashita e Fenn conseguiram demonstrar o uso da fonte de “electrospray” como um método brando de ionização para espectrometria de massas.^{54,56}

De maneira resumida, a ionização por “electrospray” envolve a formação de um “spray” eletrostático, a partir do qual são produzidas gotículas carregadas, e destas são liberados os íons (Figura 13).^{54,57}

Figura 13: Esquema de uma fonte de “electrospray” e do processo de formação do “spray”. A região I está sob pressão atmosférica, a região II, sob pressão intermediária e a região III, sob alto vácuo.^{54,58,59}

De acordo com a Figura 13, ao se aplicar um potencial positivo na solução, os íons positivos tendem a se afastar em direção ao contra-eletrodo. Deste modo, a gota que se forma na ponta do capilar é rica em íons positivos. A esta separação de carga dá-se o nome de processo eletroforético. O aumento da densidade de carga na gota aumenta o campo elétrico formado entre o capilar e o contra eletrodo, causando a deformação da gota. A gota adquire a forma de um cone (chamado de cone de Taylor) e permanece presa ao capilar. Quando a densidade de carga na superfície da gota e o aumento da repulsão entre os íons superam a tensão superficial do líquido, há a liberação de pequenas gotas com alta densidade de carga. Este processo depende do valor do campo elétrico, da tensão superficial do solvente e da condutividade da solução.

As gotas liberadas passam então por um processo de dessolvatação (evaporação do solvente; tal evaporação é favorecida pela energia térmica do ambiente e auxiliada por um gás secante, em geral, nitrogênio), que ocorre na região entre a ponta do capilar e o contra-eletrodo. A perda gradativa de solvente faz com que a densidade de carga, na gota, aumente até um valor em que as forças de repulsão superam a tensão superficial, e pequenas gotas são liberadas através da ruptura da gota inicial (Figura 14).⁵⁴

Figura 14: Esquema do processo de geração de microgotas por “electrospray”. As setas brancas representam a evaporação do solvente, e o tempo indicado sobre elas é aquele necessário para a redução do diâmetro até o valor no qual ocorre a liberação das microgotas (seta preta). O número, dentro das gotas maiores, indica a quantidade de cargas em excesso. Já o valor numérico ao lado das gotículas representa o número de cargas de cada uma delas. Portanto, o resultado da fissão é sempre uma gota principal e várias outras de menor tamanho. O esquema A, em amarelo, mostra que as gotas menores também sofrem o processo de fissão. Nesse esquema, há a geração de gotículas com diâmetro em torno de 6nm e com duas cargas em excesso. Tomando-se, como exemplo, o diâmetro de um íon sulfato hidratado que é da ordem de 0,4nm, pode-se perceber que as gotas já atingem o nível molecular, após poucas fissões. No esquema B, também em amarelo, é apresentado o processo de deformação e fissão de uma gota, com excesso de carga, numa região de alto campo elétrico. Em I, têm-se uma gota, com excesso de carga, que acabou de ser inserida em uma região com um campo elétrico. Já em II, a gota começa a se deformar e o excesso de carga, antes distribuído uniformemente sobre sua superfície, começa a se concentrar na direção estabelecida pelo campo. Em III, observa-se uma significativa deformação da região da gota de maior densidade de carga. Finalmente, em IV, há o aumento da deformação e a conseqüente liberação das gotículas, que possuem maior densidade de carga que a gota original. O tempo que decorre até a produção dos íons, na fase gasosa, é da ordem de microssegundos.^{54,59,60}

Todo o mecanismo de dessolvatação da gota até que ela atinja um raio de cerca de 30nm, está bem consolidado. Porém, ainda existem dúvidas em relação à geração do íon na fase gasosa. Há dois mecanismos propostos que tentam explicar tal fenômeno.

Um deles, proposto por Dole em 1968, foi denominado modelo da carga residual (do inglês, “charge residue model” - CRM). Segundo o CRM, a evaporação do solvente de gotas extremamente pequenas, 1nm de raio, e que contenham somente um íon, a transformará em um íon na fase gasosa.^{54,55}

Já o outro mecanismo, proposto por Thomson e Iribarne, foi chamado de modelo da evaporação do íon (do inglês, “ion evaporation model” - IEM). Este modelo propõe que haja emissão de íons (sem ruptura da gota) diretamente de gotas muito pequenas (raio de 8nm) e bastante carregadas (aproximadamente 70 cargas).^{54,61,62}

Dados experimentais auxiliam na comprovação da existência do CRM e também do IEM. Há ainda a possibilidade de que ambos os mecanismos ocorram simultaneamente.

3.3.7. Analisadores de massas

Após a geração dos íons, vêm os processos de separação e detecção dos mesmos. Tais processos são, geralmente, selecionados de acordo com características como resolução, sensibilidade, precisão e custo.

O processo de separação é realizado no analisador de massas que separa os íons de acordo com suas relações massa/carga (m/z).

Há diferentes tipos de analisadores de massas interfaceados com a ionização do tipo “electrospray”. O mais comum deles é o analisador quadrupolar, devido à facilidade de uso e ao menor custo.

O quadrupolo é constituído por quatro barras cilíndricas paralelas duas a duas. Cada par de barras opostas é conectado eletricamente estabelecendo assim, um campo quadrupolar bidimensional no plano x-y. Os íons viajam na direção z.^{36,63} Enquanto viajam nesta direção, eles também oscilam no plano x-y, devido ao potencial aplicado às barras. Portanto, em condições elétricas adequadas, os íons com uma única razão massa/carga ou que estão dentro de uma faixa de m/z selecionada, terão uma trajetória estável durante todo o caminho do quadrupolo.

A seletividade do quadrupolo é estabelecida variando-se as magnitudes das voltagens de corrente contínua DC (U), do inglês “direct current”, e de corrente alternada AC, do inglês “alternating current”, ou de rádio frequência RF (V), aplicadas a cada par de barras. Um par (plano x-z) é conectado a uma voltagem DC positiva juntamente com uma voltagem RF. O outro par de barras (plano y-z) é conectado a uma voltagem DC negativa e a voltagem RF que é igual, mas de fase oposta, àquela aplicada ao outro par de barras. Tais voltagens aplicadas podem ser representadas pelas equações 2 e 3, abaixo:

$$\phi_{x-z} = +(U + V \cos \Omega t) \quad (\text{equação 2})$$

$$\phi_{y-z} = -(U + V \cos \Omega t) \quad (\text{equação 3})$$

onde Ω é a frequência angular ($2\pi f$) do potencial RF (Figura 15).⁶³

Para se obter o espectro de massas, os íons são acelerados para dentro do espaço entre as barras, por um potencial de 5 a 10V. Apenas os íons dentro da faixa dos valores de m/z especificados chegam ao detector. Os demais colidem com as barras e são neutralizados.³⁶

Figura 15: Esquema de um filtro de massas quadrupolar. Os íons entram e viajam na direção z enquanto oscilam no plano x - y . Tal oscilação é controlada pelos potenciais DC (U) e RF (V) aplicados em cada par de barras. Somente atravessarão o quadrupolo e serão detectados, os íons que tiverem trajetórias estáveis sob os valores de U e V aplicados.⁶³

Há, ainda, outros tipos de analisadores de massas como, por exemplo, os analisadores do tipo tempo de voo (do inglês, “time of flight” - TOF); nele, os íons são acelerados, por um pulso de campo elétrico (10^3 a 10^4 V), para um tubo de separação livre de campo e com aproximadamente um metro de comprimento. Como todos os íons entram com a mesma energia cinética (idealmente), suas velocidades, no tubo, devem variar inversamente com sua massa, ou seja, os íons com massas menores chegam mais rapidamente ao detector que os íons mais pesados (Figura 16).³⁶

Figura 16: Diagrama esquemático de um analisador de massas do tipo TOF.³⁶

3.3.8. Espectrometria de massas em "tandem"

A segunda grande inovação, depois dos métodos brandos de ionização, que provocou um grande avanço nas pesquisas biológicas foi o desenvolvimento da espectrometria de massas em "tandem", ou seja, em seqüência. Comumente chamada de MS/MS ou MS^n , a espectrometria de massas (do inglês, "mass spectrometry" - MS) em "tandem" envolve o acoplamento de mais de um analisador de massas, em um único equipamento. Como exemplos desse tipo de acoplamento pode-se citar o triplo quadrupolo (Q1q2Q3) e o quadrupolo-tempo de voo (Q/TOF). O triplo quadrupolo é constituído por dois quadrupolos, que são usados como analisadores de massas, e por uma cela de colisão (Figura 17).

Figura 17: Diagrama esquemático do funcionamento básico de um triplo quadrupolo. Inicialmente, com o uso do primeiro quadrupolo (Q1), faz-se uma varredura de todos os íons existentes na amostra. Em seguida, um íon da primeira varredura é selecionado. Após isso, o primeiro quadrupolo permite que apenas o íon escolhido entre na cela de colisão (q2). Na cela, tal íon é fragmentado mediante colisão com gás (em geral, argônio ou até mesmo ar). Finalmente, o segundo quadrupolo (Q3) faz a varredura dos íons (chamados de íons filhos) gerados pela colisão com o gás.

Já o quadrupolo-tempo de voo (Q/TOF) é constituído por um analisador do tipo quadrupolo e um do tipo tempo de voo, separados por uma cela de colisão (Figura 18).

Figura 18: Esquema do espectrômetro de massas comercial Q-TOF, da Waters/Micromass. Seu funcionamento é similar ao funcionamento básico do triplo quadrupolo, descrito na Figura 17. A diferença reside no fato de que, no Q-TOF, é no analisador do tipo tempo de voo que ocorre a separação dos íons, gerados pela colisão com o gás.

Com a espectrometria de massas em “tandem” é possível realizar a seleção de um composto ionizado e fragmentá-lo, na cela de colisão, através do choque com um gás inerte (argônio ou até mesmo ar). O estudo das massas dos fragmentos obtidos fornece informações estruturais como, por exemplo, a seqüência de aminoácidos de um peptídeo ou uma alteração química específica em algum resíduo de aminoácido.²⁴

Como descrito anteriormente, a proteômica faz uso de duas plataformas metodológicas que se complementam, para a identificação de proteínas. A primeira plataforma tem como meta uma separação preliminar dos componentes a serem analisados na segunda plataforma. As técnicas mais utilizadas para essa primeira plataforma são a cromatografia líquida uni (LC) ou bidimensional (LC/LC) e a eletroforese bidimensional em gel de poliacrilamida (2D PAGE).

A segunda plataforma visa a caracterização dos compostos separados na primeira. A técnica ideal para a segunda plataforma é a espectrometria de massas em “tandem” (MS/MS).

Assim, a combinação das duas plataformas permite a montagem de sistemas extremamente poderosos para a proteômica, como LC-MS/MS e LC/LC-MS/MS.

3.3.9. Identificação de proteínas por espectrometria de massas

O emprego da espectrometria de massas na identificação de proteínas é uma ferramenta poderosa nos complexos estudos proteômicos. Existem basicamente quatro metodologias para se identificar proteínas usando dados de espectrometria de massas. Tais metodologias estão sumarizadas nos itens abaixo:

- “Peptide Mass Fingerprint” (PMF) ou “Peptide Mass Map Analysis”: inicialmente, realiza-se a digestão enzimática da proteína a ser identificada para a produção dos seus respectivos peptídeos. As massas de tais peptídeos são então determinadas por espectrometria de massas. Essas massas obtidas constituem uma espécie de impressão digital ("peptide mass fingerprinting") da proteína. Com auxílio de softwares especiais, é possível comparar essa impressão digital com aquelas geradas, teoricamente, para todas as seqüências protéicas presentes nos bancos de dados. Assim, a proteína estudada será prontamente identificada se sua seqüência estiver no banco de dados.¹⁸
- “Peptide Sequence” ou “Peptide Sequence Tag Query”: neste caso, emprega-se a espectrometria de massas em “tandem” (MS/MS), a qual permite a fragmentação dos peptídeos, gerados por digestão enzimática, das proteínas a serem identificadas. Utilizando-se os espectros de fragmentação obtidos, pode-se conseguir, manualmente, via interpretação espectral, uma seqüência parcial de um ou mais peptídios das proteínas estudadas. Novamente, empregando-se softwares especiais, é possível realizar as identificações protéicas pela comparação de uma seqüência parcial, aliada a massa total do peptídeo que gerou tal seqüência e a diferença de massa entre a seqüência parcial obtida e as porções não seqüenciadas do peptídeo, com aquelas presentes em banco de

dados. Tal método de identificação é considerado mais específico que a metodologia do "peptide mass fingerprinting".

- "MS/MS Ion Search Analysis": espectros não interpretados de MS/MS de um simples peptídeo ou de uma série deles são comparados, com o uso de softwares, com banco de dados de proteínas. Essa metodologia de identificação ocorre sem nenhuma interpretação manual de sequência.⁶⁴
- "De Novo Sequencing": envolve a análise manual, assistida por softwares, dos espectros de MS/MS. É similar ao "peptide sequence tag query", com a ressalva de que a análise manual não termina quando uma sequência curta é obtida, uma vez que o objetivo é conseguir a sequência total do peptídeo representado no espectro de MS/MS. Essa metodologia é empregada quando não há informações sobre a proteína, da qual o peptídeo sequenciado deriva, em banco de dados, ou seja, quando se trata de proteína desconhecida.

A escolha do método de identificação a ser utilizado depende de alguns fatores. Para amostras que possuem uma única proteína, "peptide mass fingerprinting" é um método satisfatório. No caso de misturas simples, ou seja, com poucas proteínas, há a necessidade de se determinar as massas dos peptídeos com alta precisão, para se aumentar a especificidade.

A metodologia do "peptide mass fingerprinting", no entanto, apresenta uma séria limitação que pode ser ilustrada pelo seguinte exemplo: ao se realizar uma pesquisa no banco de dados com o valor, medido com alta precisão, de 1529,7384 para a relação massa/carga, encontram-se dois peptídeos, mostrados na Tabela 2, como resultado. Apesar de apresentarem a mesma relação m/z , a seqüência de aminoácidos mostra que se tratam de peptídeos diferentes, pois "VGA" é diferente de "IGG", e "HAG" é semelhante, mas não igual, a "HGA". Isso mostra que o fato de o método do "peptide mass fingerprinting" usar apenas as massas dos peptídeos na identificação, faz com que ele não consiga distinguir entre dois peptídios diferentes, mas com a mesma m/z .⁵³

Tabela 2: Peptídeos com a mesma razão massa/carga (para carga igual a um), mas com seqüência de aminoácidos e fonte de origem diferentes⁵³

	Seqüência de Aminoácidos	$[M+H]^+$	Fonte
Peptídeo 1	VGA HAGEYGAEALER	1529,7384	Homem
Peptídeo 2	IGG HGAIEYGAEALER	1529,7384	Rato

Para análise de misturas protéicas mais complicadas ou para confirmação dos resultados obtidos por PMF, "peptide sequence" ou "MS/MS ion search analysis" podem ser empregados. Em termos de identificação, essas duas metodologias oferecem maior especificidade, pois se baseiam nas seqüências de aminoácidos dos peptídeos e não somente nas massas, e menos ambigüidade do que o "peptide mass

fingerprinting", apesar de serem mais lentas e mais caras. A pesquisa em banco de dados usando "peptide sequence" é mais rápida que "MS/MS ion search analysis", mas como é preciso uma prévia interpretação manual dos dados, o tempo total da análise acaba sendo maior para "peptide sequence". Para espectros de MS/MS de difícil interpretação manual (com bastante ruído, ausência de íons importantes, entre outras dificuldades), pode-se optar por "MS/MS ion search analysis". Tal técnica também pode ser empregada em espectros de boa qualidade.⁶⁴

As três primeiras metodologias de identificação de proteínas, discutidas anteriormente, usam softwares para a pesquisa em banco de dados. Uma lista desses softwares, que utilizam dados de espectrometria de massas na identificação, está mostrada na Tabela 3.

Esses três primeiros métodos necessitam que a seqüência das proteínas estudadas esteja em banco de dados de boa qualidade (com dados confiáveis). Quando isso não ocorre, faz-se uso do procedimento "de novo sequencing", o qual requer espectros de MS/MS de alta qualidade (derivados de peptídeos de alta pureza, com pouco ruído e com a presença do maior número de íons, de intensidade razoável, possíveis), bastante experiência do analista e, em geral, bastante tempo; além disso, em muitos casos, é necessário usar mais de um tipo de protease para se completar ou confirmar resultados. Em relação à sensibilidade, as quatro técnicas descritas a apresentam em alto grau.

Tabela 3: Algumas ferramentas (softwares), que usam dados de espectrometria de massas, para identificação de proteínas via pesquisa em banco de dados⁶⁴

NOME DO SOFTWARE	FONTE
Peptide Mass Fingerprint ou Peptide Mass Map Analysis	
Mascot (www.matrixscience.com)	Matrix Science Ltd.
MassSearch (cbrg.inf.ethz.ch/Server/MassSearch.html)	ETH
Mowse (www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse)	UK Human Genome Mapping Project Resource Centre
MS-FIT (prospector.ucsf.edu)	UCSF
PepMAPPER (wolf.bms.umist.ac.uk/mapper)	UMIST
PepSea (195.41.108.38/PepSeaIntro.html)	Protana
PeptideSearch (www.narrador.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html)	EMBL
Peptident (www.expasy.ch/tools/peptident.html)	ExPASy
ProFound (prowl.rockefeller.edu/PROWL/prowl.html), (also at service.proteometrics.com/prowl/profound.html)	Rockefeller University and Proteometrics
Peptide Sequence ou Peptide Sequence Tag Query	
Mascot (www.matrixscience.com)	Matrix Science Ltd.
MS-Seq (prospector.ucsf.edu)	UCSF
PepSea (195.41.108.38/PepSeaIntro.html)	Protana
PeptideSearch (www.narrador.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html)	EMBL
TagIdent (www.expasy.ch/tools/tagident.html)	ExPASy
MS/MS Ion Search Analysis	
Mascot (www.matrixscience.com)	Matrix Science Ltd.
MS-Tag (prospector.ucsf.edu)	UCSF
PepFrag (prowl.rockefeller.edu/PROWL/prowl.html)	Rockefeller University
Sequest (fields.scripps.edu/sequest)	Scripps Research Institute
Sonars MS/MS (www.proteometrics.com)	Proteometrics

Capítulo II

Parte Experimental

1. LC-MS/MS

1.1. Preparação das Amostras

1.1.1. Reagentes

As soluções utilizadas em toda parte experimental foram preparadas usando água purificada em sistema “Milli-Q” (Millipore, USA) e empregando os reagentes descritos na Tabela 4.

Tabela 4: Reagentes utilizados na parte experimental

REAGENTES	ORIGEM
Acetonitrila (grau HPLC)	Mallinckrodt (França)
Ácido Acético Glacial	Mallinckrodt (França)
Ácido Cítrico	Mallinckrodt (França)
Ácido Clorídrico	Mallinckrodt (França)
Ácido Fórmico	Mallinckrodt (França)
Ácido Iodo Acético	Acros Organics (EUA)
Cloreto de Amônio	Mallinckrodt (França)
Cloreto de Magnésio	Mallinckrodt (França)
Cloreto de Potássio	Mallinckrodt (França)
Ditiotreitol (DTT)	Mallinckrodt (França)
Sacarose	Mallinckrodt (França)
Tris (Hidroximetil) Aminometano	Mallinckrodt (França)
Tripsina (modificada; grau sequenciamento)	Promega (EUA)
Uréia	Mallinckrodt (França)

1.1.2. Obtenção dos corpos protéicos do milho

Os corpos protéicos da variedade de milho BR 473 foram obtidos de acordo com o seguinte método⁶⁵: os grãos de milho moídos, com maturação de 28 DAP (dias após polinização), foram homogeneizados em solução tampão A (tris HCl 200mmol/L; sacarose 200mmol/L; KCl 60mmol/L; MgCl₂ 50mmol/L; pH 8,5) e centrifugados a 500xg por cinco minutos; em seguida, o sobrenadante dessa centrifugação foi novamente centrifugado a 40000xg por uma hora e trinta minutos; o resíduo desta última centrifugação foi lavado, congelado e liofilizado.

1.1.3. Digestão das proteínas

Os corpos protéicos obtidos da variedade de milho BR 473 foram submetidos à digestão enzimática segundo o protocolo descrito a seguir.⁶⁶

Protocolo de Digestão

1. Suspender a proteína (10mg/mL) numa solução contendo uréia a 6 mol/L e tampão tris (pH 7,8) a 100mmol/L.
2. Adicionar 5μL de agente redutor (ditiotretitol, DTT, a 200mmol/L) a uma alíquota de 100μL da amostra da proteína (em uréia e tris). Deixar reduzir por uma hora em temperatura ambiente.

3. Adicionar 20 μ L de agente alquilante (ácido iodo acético a 200mmol/L).
Alquilar a proteína por uma hora em temperatura ambiente.
4. Adicionar 20 μ L de agente redutor para “consumir” o ácido iodo acético que não reagiu. Deixar reagir por uma hora.
5. Reduzir a concentração de uréia diluindo para 0,6mol/L.
6. Adicionar 100 μ L de solução de enzima (tripsina) contendo 20 μ g da mesma. Digerir durante toda a noite a 37°C.
7. Interromper a reação ajustando-se o pH para menor do que 6, com ácido acético.

1.2. Equipamentos e Condições Experimentais

1.2.1. Cromatografia de fase reversa

A amostra digerida foi cromatograficamente separada num sistema de cromatografia líquida (Figura 19) acoplado a um espectrômetro de massas.

Figura 19: Foto (A) e esquemas representativos (B e C) do sistema cromatográfico “capLC”, acoplado ao espectrômetro de massas, da Waters/Micromass.⁶⁷ (B) A válvula se encontra na posição de lavagem da amostra, a qual está retida na pré-coluna. (C) A válvula se encontra na posição de análise. A fase móvel empregada nas análises cromatográficas foi: A – Água/0,1%Ácido Fórmico, B – Acetonitrila/0,1%Ácido Fórmico e C – Água. O volume de injeção foi de 6,4 μ l e o fluxo na coluna analítica foi de 200nl/min. As colunas cromatográficas, de modelo Waters Symmetry 300, tiveram as seguintes especificações: pré-coluna de fase reversa (C₁₈) - 5mmx0,350mm e d_p (diâmetro das partículas) de 5 μ m; coluna analítica de fase reversa (C₁₈) - 150mmx0,075mm e d_p (diâmetro das partículas) de 3 μ m. Utilizou-se o seguinte gradiente: 10min – 10%B, 70min – 45%B, 80min – 60%B, 85min – 10%B; a etapa de lavagem da amostra com o solvente C, a 10 μ l/min, durou 10 minutos.

1.2.2. Espectrometria de massas

A coluna analítica de fase reversa, descrita na Figura 19, foi diretamente acoplada a um espectrômetro de massas da Waters/Micromass, modelo ULTIMA API (ESI-Q/TOF).⁶⁷ Um esquema representativo completo do sistema de LC-MS/MS empregado está descrito na Figura 20.

Figura 20: Esquema representativo do sistema "on-line" de LC-MS/MS.⁶⁷ O sistema capLC (A), com a válvula na posição de análise, está diretamente acoplado ao espectrômetro de massas do tipo ESI-Q/TOF (B). As análises por espectrometria de massas foram feitas no modo positivo usando os seguintes parâmetros: voltagem do cone 35V, voltagem do capilar 3500V, temperatura da fonte 100°C, varredura (modo MS) 200-2500 (m/z) e varredura (modo MS/MS) 50-4000 (m/z). O espectrômetro foi operado no modo DDA ("Data Direct Analysis") segundo o qual os peptídeos foram selecionados automaticamente para o modo MS/MS baseado na carga (+2, +3 e +4) e na intensidade (os três íons mais intensos detectados no modo MS). Empregou-se argônio como gás de colisão.

1.3. Tratamento dos Dados

1.3.1. Pesquisa no banco de dados

Os dados brutos obtidos, resultantes dos experimentos de LC-MS/MS, foram processados usando o programa "ProteinLynx" (versão 2.0) da Waters/Micromass.⁶⁷ Em seguida, os dados processados foram submetidos à pesquisa no banco de dados do NCBI ("National Center for Biotechnology Information").⁶⁸ Para tal, foi utilizado o "software" Mascot (versão 2.1)⁶⁹ com os parâmetros apresentados na Tabela 5.

Tabela 5: Parâmetros empregados para a pesquisa no banco de dados do NCBI, com o uso do "software" Mascot, utilizando os dados resultantes das análises por LC-MS/MS

PARÂMETROS DE PESQUISA	PARÂMETROS DE PESQUISA EMPREGADOS
Tipo de Pesquisa	"MS/MS Ion Search"
Enzima	Tripsina
Modificação Fixa	Carbamidometil (Cisteína)
Modificação Variável	Oxidação (Metionina)
Valores de Massa	Monoisotópicos
Tolerância de Massa (Peptídios)	±0,10Da
Tolerância de Massa (Fragmentos)	±0,14Da
Número Máximo de Clivagem Perdida	1
"Score" de Corte	12
Banco de Dados	NCBI

1.3.2. Validação das identificações

Para validar os peptídeos e proteínas identificadas através da pesquisa no banco de dados do NCBI, com o uso do “software” Mascot, foram empregados alguns critérios⁷⁰⁻⁷⁴, descritos a seguir, os quais, por serem considerados rigorosos, conferiram um alto grau de confiança para as identificações.

O programa Mascot atribui um “score” (pontuação) para cada íon (peptídeo) baseado na qualidade da correspondência entre os espectros experimentais de MS/MS e os espectros de MS/MS teóricos, gerados a partir dos peptídeos do banco de dados pesquisado. O programa também fornece um “score” limite segundo o qual os íons que apresentarem um “score” que supere esse valor, são considerados como identificações confiáveis.⁷⁵⁻⁷⁶

Neste trabalho, foram considerados, como identificações válidas, os peptídeos que tiveram um “score” acima do limite estabelecido pelo programa Mascot e/ou os peptídeos que atenderam aos demais critérios descritos a seguir. Primeiro, o espectro de MS/MS deveria ser de boa qualidade, ou seja, deveria possuir pelo menos três íons fragmentos “b” ou “y” adjacentes claramente acima do ruído da linha de base. Segundo, deveria haver continuidade entre os íons “b” ou “y”, ou seja, fragmentos de três ou mais aminoácidos adjacentes.^{74,77} Por último, se o aminoácido prolina estivesse presente na sequência do peptídeo, o íon fragmento “y”

correspondente a ela deveria ser um pico intenso quando comparado ao pico do íon fragmento precedente.⁷⁸⁻⁸¹

Proteínas identificadas por um único peptídeo foram consideradas como identificações válidas se este peptídeo apresentou um íon “score” acima do limite estabelecido pelo Mascot, e atendeu aos critérios descritos anteriormente.

2. LC/LC-MS/MS

2.1. Preparação das Amostras

As análises por LC/LC-MS/MS também foram executadas utilizando-se os corpos protéicos da variedade de milho BR 473. Tais amostras foram obtidas utilizando-se os mesmos métodos descritos nos itens 1.1.2 e 1.1.3.

2.2. Equipamentos e Condições Experimentais

2.2.1. Cromatografia de troca catiônica forte

Para os experimentos com LC/LC-MS/MS utilizou-se, como primeira dimensão, a cromatografia de troca catiônica forte (SCX). Ela foi usada para o fracionamento dos corpos protéicos digeridos do milho BR 473 e fez uso de um equipamento cujas características estão descritas na Tabela 6.

Tabela 6: Especificações do cromatógrafo líquido “2796 Alliance Bio”, da empresa Waters⁶⁷, usado para as análises de cromatografia de troca catiônica forte

	EQUIPAMENTO	
	Modelo	Marca
Bomba A	1525μ	Waters
Bomba B	1525μ	Waters
Detector UV	996 PDA	Waters
Coletor de Frações	Fraction Collector III	Waters
Coluna SCX Comercial	BioBasic SCX (150mmx2,1mm; 5μm)	Thermo Electron Corporation

A fase móvel foi constituída de: (A) acetonitrila, (B) ácido cítrico 20mM (pH 2,65) e (C) B + cloreto de amônio 1,0M. O volume de injeção foi de 200µL, o fluxo da fase móvel foi de 200µL/min, o comprimento de onda monitorado de 214nm e o gradiente utilizado está representado na Tabela 7.

Tabela 7: Gradiente usado para a análise, por SCX, do digerido protéico da variedade de milho BR 473

Tempo (min)	%A	%B	%C
0	25	75	0
40	25	5	70
50	25	75	0

O efluente das análises por SCX foi coletado em intervalos de dois minutos até o final da corrida gerando, assim, vinte e cinco frações.

Com o objetivo de diminuir o número de frações a ser posteriormente analisado, tornando, deste modo, o método de LC/LC-MS/MS mais rápido, realizou-se um reagrupamento das vinte e cinco frações (Tabela 8). Esse procedimento foi baseado nas regiões do cromatograma de SCX onde houve absorção, no comprimento de onda monitorado (214nm). As frações pertencentes aos intervalos de tempo desconsiderados são referentes às regiões, do cromatograma de SCX, onde não houve absorção claramente distinta do ruído da linha de base.

Tabela 8: Reagrupamento das frações geradas pela cromatografia de troca catiônica forte e respectivos intervalos de tempo

FRAÇÃO	INTERVALO DE TEMPO (MIN.)
Fração 1	6-8
Fração 2	8-10
Fração 3	10-12
Fração 4	12-14
Fração 5	14-18
Fração 6	18-22
Fração 7	22-26

As frações reagrupadas tiveram seu volume reduzido, em “speed-vac” (SPD1010 SpeedVac System, ThermoSavant), a aproximadamente 30µL.

Em seguida, as 7 frações geradas após o reagrupamento foram analisadas por cromatografia de fase reversa acoplada a espectrometria de massas (LC-MS/MS).

2.2.2. Cromatografia de fase reversa

As frações de SCX da variedade de milho BR 473 foram individualmente separadas por cromatografia de fase reversa acoplada a espectrometria de massas. Para tal, empregou-se o mesmo sistema descrito na Figura 19, mudando-se apenas o gradiente da análise. O novo gradiente, para todas as frações, foi o seguinte: 20min – 10%B, 60min – 100%B, 80min – 100%B, 85min – 10%B, 90min – 10%B; a etapa de lavagem da amostra com o solvente C, a 10 μ L/min, durou 20 minutos.

2.2.3. Espectrometria de massas

As análises por espectrometria de massas das frações da SCX foram realizadas no mesmo sistema “on-line”, de LC-MS/MS, descrito na Figura 20; os parâmetros empregados também foram os mesmos.

2.3. Tratamento dos Dados

2.3.1. Pesquisa no banco de dados

Os dados brutos obtidos, resultantes dos experimentos de LC/LC-MS/MS, também foram processados usando o programa “ProteinLynx” (versão 2.0) da Waters/Micromass.⁶⁷ Em seguida, os dados processados

foram submetidos à pesquisa no banco de dados do NCBI (“National Center for Biotechnology Information”).⁶⁸ Para tal, foi utilizado o “software” Mascot (versão 2.1)⁶⁹ com os parâmetros apresentados na Tabela 9.

Tabela 9: Parâmetros empregados para a pesquisa no banco de dados do NCBI, com o uso do “software” Mascot, utilizando os dados resultantes das análises por LC/LC-MS/MS

PARÂMETROS DE PESQUISA	PARÂMETROS DE PESQUISA EMPREGADOS
Tipo de Pesquisa	“MS/MS Ion Search”
Enzima	Tripsina
Modificação Fixa	Carbamidometil (Cisteína)
Modificação Variável	Oxidação (Metionina)
Valores de Massa	Monoisotópicos
Tolerância de Massa (Peptídios)	±0,09Da
Tolerância de Massa (Fragmentos)	±0,10Da
Número Máximo de Clivagem Perdida	1
“Score” de Corte	8
Banco de Dados	NCBI

2.3.2. Validação das identificações

Para validar os peptídeos e proteínas identificadas, por LC/LC-MS/MS, foram empregados os mesmos critérios descritos no item 1.3.2.

3. Esquema Geral

Um esquema geral resumido de todas as análises efetuadas está mostrado na Figura 21.

De acordo com o esquema apresentado, os corpos protéicos foram obtidos e purificados da variedade de milho BR 473; as proteínas deles derivadas foram digeridas com tripsina para produção dos seus respectivos peptídeos. Em seguida, os peptídeos foram submetidos a dois métodos de análise: LC-MS/MS e LC/LC-MS/MS. No primeiro, eles foram separados pela cromatografia líquida de fase reversa (C₁₈); já no segundo, eles passaram, inicialmente, por um fracionamento por troca catiônica, antes da separação também pela cromatografia líquida de fase reversa (C₁₈). Após as separações cromatográficas, os peptídeos foram analisados por espectrometria de massas seqüencial (MS/MS). Finalmente, os dados obtidos por LC-MS/MS e LC/LC-MS/MS foram processados e submetidos à pesquisa no banco de dados do NCBI, com o auxílio do programa Mascot. Os resultados fornecidos pelo programa foram validados antes de se obter uma lista final de proteínas identificadas.

Figura 21: Esquema geral resumido das análises realizadas.

Capítulo III

Resultados e Discussões

1. LC-MS/MS

Os corpos protéicos obtidos da variedade de milho BR 473 passaram por um processo de digestão com tripsina e os peptídeos gerados foram separados e detectados empregando-se o método de LC-MS/MS. A seguir, estão mostrados os cromatogramas (BPI, do inglês “base peak intensity”) dos peptídeos selecionados para o modo MS/MS do espectrômetro de massas (Figura 22).

Após a conclusão do experimento de LC-MS/MS, os dados brutos obtidos foram processados, submetidos à pesquisa no banco de dados do NCBI⁶⁸, com o uso do “software” Mascot⁶⁹, e validados.

Assim, o resultado final obtido da análise, por LC-MS/MS, dos corpos protéicos digeridos, está sumarizado na Tabela 10.

Figura 22: Cromatogramas (BPI) dos peptídeos selecionados para o modo MS/MS; as funções 2, 3 e 4 são as responsáveis pela seleção automática dos peptídeos, para o modo MS/MS, com base na carga (+2, +3 e +4) e na intensidade (os três íons mais intensos detectados no modo MS - função 1). Os peptídeos são derivados das proteínas, digeridas com tripsina, dos corpos protéicos do milho BR 473.

Tabela 10: Informações sobre as proteínas identificadas, da variedade de milho BR 473, usando o método de LC-MS/MS

Proteína	N. A.	M. M.	pI	% C. P.	“Score”	Peptídeos Identificados
Protein brittle-1, chloroplast precursor [Zea mays]	gil231654	47054	8,51	15	211	1. R.LVSGAIAGAVSR.T 2. R.TFVAPLETIR.T 3. K.DVYDNVAHAFVK.I 4. R.RPGADVGPVATLLIGSAAGAIASSATFPLEVAR.K
Legumin 1 [Zea mays]	gil16305144	53346	6,31	9	152	1. R.VVVDAMGLLLPR.Y 2. K.FLLAGGFSK.G 3. K.GQPHFAENIFK.G 4. R.AGQLLIVPQGYLVATK.A
Glucose-1-phosphate adenylyltransferase [Zea mays]	gil1707924	57890	6,16	11	102	1. R.VSAILGGGTGSQFLPTSTR.A 2. K.SIDNIVILSGDQLYR.M 3. R.VLQFFEKPK.G 4. K.YTQLHDFGSEILPR.A

Observações:

- O “score” do íon é dado por $-10 \cdot \log(P)$, onde P é a probabilidade de que a combinação observada é um evento aleatório.
- “Score” de íons individuais acima de 46 indicam identidade ou extensa homologia ($p < 0,05$).
- O “score” da proteína é derivado do “score” dos íons e tem base não probabilística.
- Abreviaturas: N.A. – número de acesso da proteína no NCBI, M.M. – massa molecular teórica (Da), pI – ponto isoelétrico teórico, %C P. – porcentagem de cobertura da proteína.

Tabela 10 (continuação): Informações sobre as proteínas identificadas, da variedade de milho BR 473, usando o método de LC-MS/MS

Proteína	N. A.	M. M.	pI	% C. P.	“Score”	Peptídeos Identificados
Glyceraldehyde-3-phosphate dehydrogenase (GAPC3/GAPC4) [Zea mays]	gil6166167/ gil1184776	36597/ 36656	7,01/ 6,61	14/ 14	93/ 93	1. K.FGIVEGLMTTVHAITATQK.T 2. K.GILGYVEEDLVSTDFQGDSR.S 3. K.AGIALNGNFVK.L
Globulin-1 (GLB1-L/GLB1-S) [Zea mays]	gil22284/ gil121205	66635/ 65446	6,23/ 6,63	7/ 7	65/ 65	1. R.NPESFLSSFSK.S 2. R.LSPGTAfVVPAGHPFVAVASR.D 3. K.VFLAGADNVLQK.L
16kD gamma zein [Zea mays]	gil16305111	20230	7,48	3	48	1. R.QVEPLHR.Y

Observações:

- O “score” do íon é dado por $-10 \cdot \log(P)$, onde P é a probabilidade de que a combinação observada é um evento aleatório.
- “Score” de íons individuais acima de 46 indicam identidade ou extensa homologia ($p < 0,05$).
- O “score” da proteína é derivado do “score” dos íons e tem base não probabilística.
- Abreviaturas: N.A. – número de acesso da proteína no NCBI, M.M. – massa molecular teórica (Da), pI – ponto isoelétrico teórico, %C P. – porcentagem de cobertura da proteína.

2. LC/LC-MS/MS

Os peptídeos gerados dos corpos protéicos do milho BR 473 também foram submetidos ao método de LC/LC-MS/MS. A introdução da cromatografia de troca catiônica forte (SCX), como primeira dimensão, teve como objetivo a realização de um fracionamento dos peptídeos, antes da análise pela cromatografia de fase reversa (segunda dimensão) acoplada a espectrometria de massas.

Na Figura 23, a seguir, está mostrado o cromatograma de SCX e os cromatogramas (BPI) dos peptídeos selecionados para o modo MS/MS do espectrômetro de massas. As funções 2, 3 e 4 são as responsáveis pela seleção automática dos peptídeos, para o modo MS/MS, baseada na carga (+2, +3 e +4) e na intensidade (os três íons mais intensos detectados no modo MS - função 1).

Após a análise individual de cada uma das frações, os dados brutos obtidos foram processados, submetidos à pesquisa no banco de dados do NCBI⁶⁸, com o uso do “software” Mascot⁶⁹, e validados.

O resultado final obtido da análise, por LC/LC-MS/MS, dos corpos protéicos digeridos, encontra-se sumarizado na Tabela 11.

Figura 23: Cromatograma de SCX com indicação das sete frações produzidas da amostra de peptídeos dos corpos protéicos digeridos do milho BR 473. Para cada fração são mostrados os cromatogramas (BPI) dos peptídeos selecionados para o modo MS/MS do espectrômetro de massas.

Tabela 11: Informações sobre as proteínas identificadas, da variedade de milho BR 473, usando o método de LC/LC-MS/MS

Proteína	N. A.	M. M.	pI	% C. P.	“Score”	Peptídeos Identificados
Globulin-1 (GLB1-L) [Zea mays]	gil22284	66635	6,23	23	674	1. R.VLRPFDEVSR.L 2. K.QGHV FVAPAGAVTYLANTDGR.K 3. R.NPESFLSSFSK.S 4. R.GPYSLLDQRPSIANQHGQLYEADAR.S 5. R.SEEEEESSEEQEEAGQGYHTIR.A 6. R.LSPGTAFVVPAGHPFVAVASR.D 7. K.VFLAGADNVLQK.L 8. K.GFLPGPEESGGHEER.E
Globulin-1 (GLB1-S) [Zea mays]	gil121205	65446	6,63	21	563	1. R.VLRPFDEVSR.L 2. K.QGHV FVAPAGAVTYLANTDGR.K 3. R.NPESFLSSFSK.S 4. R.GPYSLLDQRPSIANQHGQLYEADAR.S 5. R.SEEEEESSEEQEEVGQGYHTIR.A 6. R.LSPGTAFVVPAGHPFVAVASR.D 7. K.VFLAGADNVLQK.L

Observações:

- O “score” do íon é dado por $-10 \cdot \log(P)$, onde P é a probabilidade de que a combinação observada é um evento aleatório.
- “Score” de íons individuais acima de 46 indicam identidade ou extensa homologia ($p < 0,05$).
- O “score” da proteína é derivado do “score” dos íons e tem base não probabilística.
- Abreviaturas: N.A. – número de acesso da proteína no NCBI, M.M. – massa molecular teórica (Da), pI – ponto isoelétrico teórico, %C P. – porcentagem de cobertura da proteína

Tabela 11 (continuação): Informações sobre as proteínas identificadas, da variedade de milho BR 473, usando o método de LC/LC-MS/MS

Proteína	N. A.	M. M.	pI	% C. P.	“Score”	Peptídeos Identificados
Protein brittle-1, chloroplast precursor [Zea mays]	gil231654	47054	8,51	20	500	1. R.LVSGAIAGAVSR.T 2. R.TFVAPLETIR.T 3. K.AIEHFTYDTAK.K 4. K.DVYDNVAHAFVK.I 5. K.ILRDEGPSELYR.G 6. R.RPGADVGPVATLLIGSAAGAIASSA TFPLEVAR.K
Legumin 1 [Zea mays]	gil16305144	53346	6,31	21	458	1. K.VQSEAGSVQYFSR.F 2. R.KFLLAGGFSK.G 3. K.FLLAGGFSK.G 4. K.GQPHFAENIFK.G 5. R.FLSEALGVSMHVAEK.L Oxidation (M) 6. R.LDQADVYSPGAGR.I 7. R.AGQLLIVPQGYLVATK.A 8. R.KHELAVLTPAGSGSYQQGQAGSAQQ.-

Observações:

- O “score” do fon é dado por $-10 \cdot \log(P)$, onde P é a probabilidade de que a combinação observada é um evento aleatório.
- “Score” de fons individuais acima de 46 indicam identidade ou extensa homologia ($p < 0,05$).
- O “score” da proteína é derivado do “score” dos fons e tem base não probabilística.
- Abreviaturas: N.A. – número de acesso da proteína no NCBI, M.M. – massa molecular teórica (Da), pI – ponto isoelétrico teórico, %C P. – porcentagem de cobertura da proteína

Tabela 11 (continuação): Informações sobre as proteínas identificadas, da variedade de milho BR 473, usando o método de LC/LC-MS/MS

Proteína	N. A.	M. M.	pI	% C. P.	“Score”	Peptídeos Identificados
Glucose-1-phosphate adenylyltransferase [Zea mays]	gil1707924	57890	6,16	19	440	1. R.VSAIILGGGTGSQLFPLTSTR.A 2. K.FIWVLEDYYSHK.S 3. K.SIDNIVILSGDQLYR.M 4. R.VLQFFFEKPK.G 5. K.KDALLDLLK.S 6. K.DALLDLLK.S 7. K.YTQLHDFGSEILPR.A 8. K.FDFYDPK.T 9. K.GIQEADHP EEGYYIR.S
Glyceraldehyde-3-phosphate dehydrogenase (GAPC3/GAPC4) [Zea mays]	gil6166167/ gil1184776	36597/ 36656	7,01/ 6,61	21/ 21	426/ 426	1. K.KVVISAPSK.D 2. K.FGIVEGLMTTVHAITATQK.T 3. R.VPTVDVSVVDLTVR.L 4. K.GILGYVEEDLVSTDFQGDSR.S 5. K.AGIALNGNFVK.L
Glyceraldehyde-3-phosphate dehydrogenase (GAPC2) [Zea mays]	gil312179	36633	6,41	14	276	1. K.TLLFGEKPVTVFGIR.N 2. K.KVVISAPSK.D 3. R.VPTVDVSVVDLTVR.L 4. K.AGIALNDHFIK.L

Observações:

- O “score” do íon é dado por $-10 \cdot \log(P)$, onde P é a probabilidade de que a combinação observada é um evento aleatório.
- “Score” de íons individuais acima de 46 indicam identidade ou extensa homologia ($p < 0,05$).
- O “score” da proteína é derivado do “score” dos íons e tem base não probabilística.
- Abreviaturas: N.A. – número de acesso da proteína no NCBI, M.M. – massa molecular teórica (Da), pI – ponto isoelétrico teórico, %C P. – percentagem de cobertura da proteína.

Tabela 11 (continuação): Informações sobre as proteínas identificadas da variedade de milho BR 473, usando o método de LC/LC-MS/MS

Proteína	N. A.	M. M.	pI	% C. P.	“Score”	Peptídeos Identificados
Alpha subunit of translation elongation factor 1 (eEF1A) [Zea mays]	gil1321656/ gil2282584/ gil7230397	49544/ 49601/ 49405	9,19/ 9,19/ 9,20	9/ 9/ 9	198/ 198/ 198	1. R.LPLQDVYK.I 2. K.IGGIGTVPVGR.V 3. R.VETGVIKPGMVVTFGPTGLTTEVK.S Oxidation (M)
Glyceraldehyde-3-phosphate dehydrogenase (GAPC1) [Zea mays]	gil295853	36614	6,46	11	179	1. K.TLLFGDKPVTVFGR.N 2. K.KVVISAPSK.D 3. R.VPTVDVSVDLTVR.L
Starch branching enzyme IIB [Zea mays]	gil1169911	90916	5,86	4	158	1. K.NEFGVWEIFLPNNADGTSPIPHGSR.V 2. R.FDLGDADYLR.Y
Adenine nucleotide translocator [Zea mays]	gil22162/ gil22164	42593/ 42477	9,85/ 9,85	5/ 5	115/ 115	1. R.QFNGLVVDVYR.K 2. K.SSLDAFQQILK.K
Plasma membrane integral protein (ZmPIP2-1) [Zea mays]	gil13447801	30424	7,68	4	76	1. R.YGGGANSLASGYSR.G

Observações:

- O “score” do íon é dado por $-10 \cdot \log(P)$, onde P é a probabilidade de que a combinação observada é um evento aleatório.
- “Score” de íons individuais acima de 46 indicam identidade ou extensa homologia ($p < 0,05$).
- O “score” da proteína é derivado do “score” dos íons e tem base não probabilística.
- Abreviaturas: N.A. – número de acesso da proteína no NCBI, M.M. – massa molecular teórica (Da), pI – ponto isoelétrico teórico, %C P. – porcentagem de cobertura da proteína.

Tabela 11 (continuação): Informações sobre as proteínas identificadas da variedade de milho BR 473, usando o método de LC/LC-MS/MS

Proteína	N. A.	M. M.	pI	% C. P.	“Score”	Peptídeos Identificados
Brittle 2 [Zea mays]	gil23664291	52928	5,65	10	72	1. K.AAANDSTYLNPAHDSVLGIIILGGGAGTR.L 2. R.EQFPEANDFGSEVIPGATSIGKR.V
Sucrose synthase [Zea mays]	gil22488	92073	5,96	5	67	1. K.ESLYPLLNFLK.A 2. R.VLDTLHLLLDLLEAPDPANLEK.F 3. K.AADILVNFFDK.C
Geranylgeranyl-diphosphate synthase [Zea mays]	gil33340602	40444	5,13	4	66	1. K.TASGQLLDLITTHEGEK.D
Globulin 2 (GLB2) [Zea mays]	gil228310	50234	6,16	5	61	1. R.FTHELLEDAVGNYR.V 2. K.GEITTASEEQIR.E

Observações:

- O “score” do íon é dado por $-10 \cdot \log(P)$, onde P é a probabilidade de que a combinação observada é um evento aleatório.
- “Score” de íons individuais acima de 46 indicam identidade ou extensa homologia ($p < 0,05$).
- O “score” da proteína é derivado do “score” dos íons e tem base não probabilística.
- Abreviaturas: N.A. – número de acesso da proteína no NCBI, M.M. – massa molecular teórica (Da), pI – ponto isoelétrico teórico, %C P. – porcentagem de cobertura da proteína.

3. Comparação entre os Métodos LC-MS/MS e LC/LC-MS/MS

3.1. Número de Proteínas e Peptídeos Identificados

O número de proteínas e peptídeos identificados nos dois métodos desenvolvidos, LC-MS/MS e LC/LC-MS/MS, está apresentado na Figura 24 e na Figura 25 a seguir.

Figura 24: Número de proteínas identificadas em cada um dos dois métodos empregados: LC-MS/MS e LC/LC-MS/MS. No método LC-MS/MS, foram identificadas 6 proteínas no total sendo que 5 delas também foram identificadas no outro método. Já no método LC/LC-MS/MS, foram identificadas 16 proteínas no total. A área de interseção dos círculos indica o número de proteínas identificadas comuns nos dois métodos. No total, foram identificadas 17 proteínas únicas considerando-se ambos os métodos.

Figura 25: Número de peptídeos identificados em cada um dos dois métodos empregados: LC-MS/MS e LC/LC-MS/MS. No método LC-MS/MS, foram identificados 19 peptídeos no total sendo que 17 deles também foram identificados no outro método. Já no método LC/LC-MS/MS, foram identificados 55 peptídeos no total. A área de interseção dos círculos indica o número de peptídeos identificados comuns nos dois métodos. No total, foram identificados 57 peptídeos únicos considerando-se ambos os métodos.

De acordo com as figuras apresentadas, pode-se afirmar que o método de LC/LC-MS/MS mostrou-se extremamente superior, pois identificou mais peptídeos e, consequentemente, mais proteínas. Essa superioridade está relacionada ao fato desse método empregar duas dimensões, ou seja, duas colunas cromatográficas com mecanismos de separação diferentes; isso possibilita a separação e a posterior identificação de mais componentes da amostra.

Apesar da superioridade de um dos métodos, eles podem ser considerados complementares, uma vez que foram identificados peptídeos e proteínas únicas em cada um dos métodos. Portanto, em conjunto, eles contribuem para o aumento da quantidade de informação final obtida do

sistema estudado, em relação ao número total de proteínas e peptídeos únicos identificados.

3.2. Qualidade do Espectro de MS/MS

Aproximadamente 94% dos espectros de MS/MS, derivados do método de LC/LC-MS/MS, apresentaram uma qualidade superior quando comparados aos espectros, dos mesmos peptídeos, originados pelo método de LC-MS/MS. Tal fato pode ser verificado na Figura 26, a qual mostra a comparação entre os espectros de MS/MS, do peptídeo “K.DVYDNVAHAFVK.I” de m/z 689,27 (2+), gerados pelos métodos de LC-MS/MS (A) e LC/LC-MS/MS (C). Os símbolos, (*), nos cromatogramas (E) e (F), indicam o pico de eluição do peptídeo considerado em ambos os métodos. Os espectros (B) e (D) mostram uma ampliação de (A) e (C), respectivamente, além das séries de íons “b” e “y”. O “score”, atribuído pelo programa Mascot, para o espectro (A) ou (B) foi de 76, enquanto que para o espectro (C) ou (D), foi de 102.

Figura 26: Comparação entre os espectros de MS/MS, do peptídeo “K.DVYDNVAHAFVK.I” de m/z 689,27 (2+), gerados pelos métodos de LC-MS/MS e LC/LC-MS/MS.

Nesse exemplo, a melhor qualidade espectral observada para C ou D, que se deve principalmente a sua maior intensidade em relação a A ou B, proporcionou uma identificação mais confiável. Isso foi possível porque houve uma alta confiabilidade na atribuição das séries de íons “b” e “y”, devido a fácil distinção dos picos correspondentes a elas em relação ao ruído da linha de base. Além disso, o “score”, fornecido pelo programa Mascot e que também é uma medida de confiabilidade, foi maior para a identificação do peptídeo derivado do espectro de melhor qualidade (C ou D).

Apesar da melhor qualidade espectral e do conseqüente aumento da confiabilidade observados para o método de LC/LC-MS/MS, os dois métodos, quando empregados em conjunto, contribuem para um aumento ainda maior da confiabilidade, uma vez que determinados peptídeos são identificados não em apenas um, mas em ambos os métodos.

3.3. Comparação de Parâmetros

Alguns parâmetros são de extrema importância na identificação de proteínas, tais como: o “score”, a porcentagem de cobertura da proteína e o número de íons “b” e “y”. Com o objetivo de comparar os métodos de LC-MS/MS e LC/LC-MS/MS utilizados, tais parâmetros, obtidos para todas as proteínas identificadas nos dois métodos, foram confrontados em um único gráfico (Figura 27).

Figura 27: Comparação entre os métodos LC-MS/MS e LC/LC-MS/MS, considerando-se os valores normalizados, para todas as proteínas identificadas, dos seguintes parâmetros: o "score", a porcentagem de cobertura da proteína (%CP) e o número de íons "b" e "y" (b+y).

O gráfico mostra um aumento no valor de todos os parâmetros para o método de LC/LC-MS/MS, no caso das proteínas comuns identificadas nos dois métodos. Observa-se, ainda, que os parâmetros que sofreram os maiores aumentos foram o “score” e a quantidade de íons “b” e “y” (b+y).

Como citado anteriormente, o programa Mascot fornece um “score” para a proteína, que é derivado dos “scores” dos peptídeos que, por sua vez, são baseados na qualidade da correspondência entre os espectros experimentais de MS/MS e os espectros de MS/MS teóricos. Assim, um aumento no “score” significa um aumento na qualidade dessa correspondência e, conseqüentemente, um aumento na confiabilidade das identificações.⁷⁵⁻⁷⁶

Ainda em relação à qualidade da correspondência entre os espectros de MS/MS, ela é aumentada quando é realizada entre espectros de boa qualidade, ou seja, espectros que possuam pelo menos três íons fragmentos “b” ou “y” adjacentes claramente acima do ruído da linha de base.^{74,77} Deste modo, desde que estejam acima do ruído da linha de base, um número maior de íons “b” ou “y” melhora a qualidade do espectro e, conseqüentemente, da correspondência entre eles. Novamente, isso aumenta o “score” e como resultado melhora a confiabilidade das identificações.

Por último, apesar da porcentagem de cobertura da proteína (%CP) ter sofrido um aumento menor no seu valor, em relação aos outros parâmetros, ele é importante, pois indica que foi identificado um número maior de peptídeos elevando, assim, a segurança das identificações; isso é válido, desde que esses peptídeos tenham sido identificados com uma confiabilidade aceitável, ou seja, valores de “score” e de (b+y) razoáveis.

Portanto, pode-se concluir que o método de LC/LC-MS/MS produziu identificações mais confiáveis, quando se comparou as proteínas comuns identificadas nos dois métodos, devido ao aumento dos três parâmetros relatados anteriormente.

4. Proteínas Identificadas

4.1. Zeínas

Como descrito anteriormente, as proteínas do milho podem ser classificadas como não-zeínas e zeínas. De acordo com a solubilidade e homologia de sequência, as zeínas são classificadas em α , β , γ e δ .¹⁻³ Segundo os resultados apresentados (Tabela 10), houve apenas a identificação da zeína γ de 16kDa (“16kDa gamma zein”). Esse fato pode ser explicado pelo uso intencional da enzima tripsina no processo de digestão.

A tripsina cliva as proteínas na direção C-terminal dos aminoácidos lisina e arginina, desde que eles não sejam seguidos por uma prolina.⁵³ Como as zeínas são pobres em lisina e arginina, o processo de digestão pode gerar peptídeos muito grandes, o que dificulta ou impede a fragmentação dos mesmos na cela de colisão do espectrômetro de massas. Além disso, a presença de muitos outros peptídeos, das proteínas não-zeínas, pode provocar uma supressão dos poucos peptídeos obtidos das zeínas.

Assim, o uso intencional da tripsina possibilitou a identificação majoritária de proteínas não-zeínas, estando de acordo com o objetivo principal desse trabalho que era identificar proteínas desse tipo, as quais pudessem ser relacionadas ao alto conteúdo protéico da variedade de milho estudada.

4.2. Seqüência das Proteínas Identificadas

Após a identificação das proteínas únicas nos dois métodos desenvolvidos, suas seqüências foram individualmente analisadas para se verificar a quantidade e a porcentagem de cada aminoácido, em relação ao número ao número total de aminoácidos de cada proteína (Tabela 12).⁸¹

Tabela 12: Aminoácidos que compõem as 17 proteínas únicas identificadas pelos métodos de LC-MS/MS e LC/LC-MS/MS. Nos casos em que houve mais de uma identificação potencial (adenine nucleotide translocator, eEF1A e GAPC3/GAPC4), segundo a Tabela 11, foi considerada apenas uma (gi|22162, gi|1321656 e gi|6166167, respectivamente), devido a grande semelhança de sequência entre elas.

Abreviatura: N°AA – número de aminoácidos

	Adenine nucleotide translocator		eEF1A		Brittle-1		Brittle-2		GAPC1		GAPC2		GAPC3/GAPC4	
	N° AA	%	N° AA	%	N° AA	%	N° AA	%	N° AA	%	N° AA	%	N° AA	%
Ala	38	9,8191	31	6,9351	55	12,6146	45	9,5744	29	8,6053	30	8,9020	30	8,9020
Cys	4	1,0335	6	1,3422	8	1,8348	6	1,2765	2	0,5934	2	0,5934	3	0,8902
Asp	17	4,3927	26	5,8165	17	3,8990	31	6,5957	26	7,7151	23	6,8249	21	6,2314
Glu	10	2,5839	28	6,2639	33	7,5688	25	5,3191	15	4,4510	18	5,3412	19	5,6379
Phe	25	6,4599	15	3,3557	14	3,2110	20	4,2553	14	4,1543	14	4,1543	13	3,8575
Gly	44	11,3695	39	8,7248	46	10,5504	34	7,2340	30	8,9020	30	8,9020	29	8,6053
His	2	0,5167	11	2,4608	5	1,1467	7	1,4893	8	2,3738	7	2,0771	6	1,7804
Ile	22	5,6847	30	6,7114	19	4,3577	39	8,2978	22	6,5281	24	7,1216	19	5,6379
Lys	27	6,9767	49	10,9619	23	5,2752	29	6,1702	28	8,3086	28	8,3086	29	8,6053
Leu	32	8,2687	27	6,0402	29	6,6513	39	8,2978	20	5,9347	19	5,6379	20	5,9347
Met	11	2,8423	13	2,9082	11	2,5229	14	2,9787	8	2,3738	8	2,3738	7	2,0771
Asn	19	4,9095	14	3,1319	8	1,8348	26	5,5319	13	3,8575	13	3,8575	14	4,1543
Pro	12	3,1007	22	4,9217	34	7,7981	26	5,5319	12	3,5608	12	3,5608	10	2,9673
Gln	14	3,6175	9	2,0134	10	2,2935	14	2,9787	3	0,8902	4	1,1869	5	1,4836
Arg	23	5,9431	18	4,0268	31	7,1100	20	4,2553	11	3,2640	11	3,2640	11	3,2640
Ser	30	7,7519	21	4,6979	21	4,8165	25	5,3191	23	6,8249	23	6,8249	26	7,7151
Thr	14	3,6175	32	7,1588	22	5,0458	19	4,0425	23	6,8249	23	6,8249	21	6,2314
Val	21	5,4263	40	8,9485	34	7,7981	25	5,3191	36	10,6824	34	10,0890	40	11,8694
Trp	5	1,2919	3	0,6711	4	0,9174	4	0,8510	5	1,4836	5	1,4836	5	1,4836
Tyr	17	4,3927	13	2,9082	12	2,7522	22	4,6808	9	2,6706	9	2,6706	9	2,6706
Total	387	100	447	100	436	100	470	100	337	100	337	100	337	100

Tabela 12 (continuação): Aminoácidos que compõem as 17 proteínas únicas identificadas pelos métodos de LC-MS/MS e LC/LC-MS/MS.

Nos casos em que houve mais de uma identificação potencial (adenine nucleotide translocator, eEF1A e GAPC3/GAPC4), segundo a Tabela 11, foi considerada apenas uma (gi|22162, gi|1321656 e gi|6166167, respectivamente), devido a grande semelhança de sequência entre elas. Abreviatura: N°AA – número de aminoácidos

	Geranylgeranyl-diphosphate synthase		GLB1-L		GLB1-S		GLB2		Glucose-1-phosphate adenylyltransferase	
	N° AA	%	N° AA	%	N° AA	%	N° AA	%	N° AA	%
Ala	26	7,4285	44	7,5601	43	7,5043	31	6,8888	36	6,9767
Cys	8	2,2857	9	1,5463	8	1,3961	6	1,3333	15	2,9069
Asp	27	7,7142	24	4,1237	23	4,0139	19	4,2222	34	6,5891
Glu	27	7,7142	81	13,9175	76	13,2635	45	10,0000	26	5,0387
Phe	16	4,5714	23	3,9518	22	3,8394	21	4,6666	22	4,2635
Gly	25	7,1428	51	8,7628	51	8,9005	42	9,3333	39	7,5581
His	6	1,7142	27	4,6391	26	4,5375	13	2,8888	11	2,1317
Ile	21	6,0000	17	2,9209	17	2,9668	15	3,3333	40	7,7519
Lys	26	7,4285	21	3,6082	22	3,8394	19	4,2222	30	5,8139
Leu	39	11,1428	32	5,4982	33	5,7591	32	7,1111	41	7,9457
Met	4	1,1428	5	0,8591	4	0,6980	9	2,0000	12	2,3255
Asn	13	3,7142	12	2,0618	12	2,0942	10	2,2222	24	4,6511
Pro	8	2,2857	24	4,1237	24	4,1884	18	4,0000	20	3,8759
Gln	14	4,0000	22	3,7800	21	3,6649	16	3,5555	20	3,8759
Arg	16	4,5714	73	12,5429	72	12,5654	39	8,6666	25	4,8449
Ser	13	3,7142	45	7,7319	45	7,8534	49	10,8888	41	7,9457
Thr	15	4,2857	18	3,0927	17	2,9668	15	3,3333	29	5,6201
Val	21	6,0000	40	6,8728	42	7,3298	37	8,2222	30	5,8139
Trp	5	1,4285	3	0,5154	3	0,5235	5	1,1111	3	0,5813
Tyr	20	5,7142	11	1,8900	12	2,0942	9	2,0000	18	3,4883
Total	350	100	582	100	573	100	450	100	516	100

Tabela 12 (continuação): Aminoácidos que compõem as 17 proteínas únicas identificadas pelos métodos de LC-MS/MS e LC/LC-MS/MS.

Nos casos em que houve mais de uma identificação potencial (adenine nucleotide translocator, eEF1A e GAPC3/GAPC4), segundo a Tabela 11, foi considerada apenas uma (gi|22162, gi|1321656 e gi|6166167, respectivamente), devido a grande semelhança de sequência entre elas. Abreviatura: N°AA – número de aminoácidos

	Legumin 1		Starch branching enzyme IIB		Sucrose synthase		ZmPIP2-1		16kDa gamma zein	
	N° AA	%	N° AA	%	N° AA	%	N° AA	%	N° AA	%
Ala	45	9,3167	55	6,8836	51	6,3591	44	15,1724	20	10,9289
Cys	7	1,4492	8	1,0012	7	0,8728	4	1,3793	12	6,5573
Asp	21	4,3478	58	7,2590	48	5,9850	13	4,4827	0,0000	0,0000
Glu	25	5,1759	45	5,6320	56	6,9825	5	1,7241	3	1,6393
Phe	22	4,5548	48	6,0075	40	4,9875	16	5,5172	7	3,8251
Gly	43	8,9026	73	9,1364	47	5,8603	37	12,7586	15	8,1967
His	21	4,3478	34	4,2553	23	2,8678	6	2,0689	4	2,1857
Ile	17	3,5196	34	4,2553	49	6,1097	22	7,5862	2	1,0928
Lys	12	2,4844	31	3,8798	48	5,9850	9	3,1034	1	0,5464
Leu	41	8,4886	54	6,7584	89	11,0972	25	8,6206	19	10,3825
Met	11	2,2774	26	3,2540	20	2,4937	3	1,0344	4	2,1857
Asn	17	3,5196	31	3,8798	30	3,7406	6	2,0689	1	0,5464
Pro	21	4,3478	40	5,0062	35	4,3640	13	4,4827	25	13,6612
Gln	34	7,0393	19	2,3779	30	3,7406	4	1,3793	31	16,9398
Arg	27	5,5900	50	6,2578	42	5,2369	10	3,4482	3	1,6393
Ser	44	9,1097	50	6,2578	52	6,4837	16	5,5172	11	6,0109
Thr	13	2,6915	35	4,3804	40	4,9875	14	4,8275	6	3,2786
Val	41	8,4886	54	6,7584	47	5,8603	25	8,6206	10	5,4644
Trp	4	0,8281	14	1,7521	9	1,1221	5	1,7241	1	0,5464
Tyr	17	3,5196	40	5,0062	39	4,8628	13	4,4827	8	4,3715
Total	483	100	799	100	802	100	290	100	183	100

Uma vez determinada a quantidade de cada aminoácido que constitui as proteínas únicas identificadas, o objetivo seguinte foi verificar quais delas poderiam contribuir para o alto conteúdo de lisina e de triptofano da variedade de milho BR 473. Para tal, foi construído, com os dados da Tabela 12, um gráfico que mostra a contribuição de cada proteína única identificada para a maior qualidade nutricional do milho estudado (Figura 28).

De acordo com a Figura 28, apenas a proteína “16kDa gamma zein”, que é uma zeína, não faz parte da fração não-zeína da qual se busca encontrar proteínas que sejam responsáveis pelo alto conteúdo de lisina e triptofano apresentado pelo milho BR 473. Todas as outras proteínas não-zeínas identificadas podem contribuir para o conteúdo final desses aminoácidos, uma vez que apresentam quantidades variando entre ~2,5-11% de lisina e ~0,5-1,8% de triptofano.

Para o conteúdo de lisina, a contribuição dessas proteínas não-zeínas é, quantitativamente, superior à contribuição para o conteúdo de triptofano. No entanto, levando-se em consideração o aumento verificado na Tabela 1, em ambos os casos houve um aumento de aproximadamente 50% no conteúdo desses dois aminoácidos.

Considerando-se as proteínas individualmente, as não-zeínas “alpha subunit of translation elongation factor 1” e “glyceraldehyde-3-phosphate dehydrogenase” são as que mais podem contribuir para o conteúdo de lisina⁸², enquanto que as proteínas “starch branching enzyme IIB” e “ZmPIP2-1”, podem apresentar uma maior contribuição para a quantidade de triptofano.

É importante ressaltar que nesse estudo não foram verificadas as concentrações dessas proteínas não-zeínas; no entanto, apesar dos diferentes níveis de concentração que elas possivelmente apresentam, qualquer uma delas, em baixos ou em altos níveis, podem contribuir para a quantidade superior de lisina e triptofano observada para a variedade de milho estudada.

Figura 28: Contribuição de cada proteína única identificada, pelos métodos de LC-MS/MS e LC/LC-MS/MS, para o alto conteúdo de lisina e triptofano da variedade de milho BR 473.

4.3. Proteínas com mais de uma Identificação Potencial

A proteína “glyceraldehyde-3-phosphate dehydrogenase” é derivada de uma pequena família de genes formada por gpc1, gpc2, gpc3 e gpc4. As proteínas dela derivada podem ser divididas em dois grupos: grupo I, constituído por GAPC1 e GAPC2 que apresentam 97% de identidade, e grupo II, formado por GAPC3 e GAPC4 que são 99.4% idênticas.⁸³ As proteínas do grupo I foram individualmente identificadas nesse estudo, enquanto que as do grupo II foram identificadas em conjunto pois os peptídeos delas identificados não foram referentes as pequenas regiões que as diferenciam.

O milho possui de 10 a 15 genes “alpha subunit of translation elongation factor 1 (eEF1A)”, os quais são quase idênticos e codificam proteínas extremamente similares.⁸⁴ Nesse estudo, foram identificadas três proteínas relacionadas a esses genes.

O milho também possui dois genes “adenine nucleotide translocator” que codificam duas proteínas muito similares.⁸⁵ Na Tabela 11, foram relacionados dois números de acessos que se referem a essas proteínas, uma vez que os peptídeos delas identificados foram referentes as regiões comuns entre elas, não sendo possível a diferenciação entre as duas.

Capítulo IV

Conclusões

As principais conclusões deste trabalho foram:

- ✓ O desenvolvimento do método de LC-MS/MS envolveu a otimização dos seguintes aspectos: digestão das proteínas, cromatografia líquida de fase reversa, espectrometria de massas, pesquisa no banco de dados e validação das identificações; com ele foi possível identificar 19 peptídeos e, através deles, 6 proteínas dos corpos protéicos da variedade de milho BR 473.
- ✓ O desenvolvimento do método de LC/LC-MS/MS envolveu a otimização dos mesmos aspectos da metodologia de LC-MS/MS, além da cromatografia de troca catiônica forte; com ele foi possível identificar 55 peptídeos e, através deles, 16 proteínas dos corpos protéicos da variedade de milho BR 473.
- ✓ Os resultados mostraram a superioridade do método de LC/LC-MS/MS, pois identificou um número maior de proteínas; além disso, quando foram comparadas as proteínas comuns identificadas pelos dois métodos, observou-se uma maior confiabilidade das identificações por LC/LC-MS/MS, pois foi baseada em espectros de melhor qualidade e valores maiores de parâmetros como o “score”, a porcentagem de cobertura da proteína e o número de íons “b” e “y”.

- ✓ Apesar da superioridade de um dos métodos, eles podem ser considerados complementares, uma vez que foram identificados peptídeos e proteínas únicas em cada um deles. Portanto, em conjunto, eles contribuem para o aumento da quantidade de informação final obtida do sistema estudado, em relação ao número total de proteínas e peptídeos únicos identificados, e para o aumento da confiabilidade, uma vez que determinadas proteínas são identificadas não em apenas um, mas em ambos os métodos.
- ✓ O uso da tripsina para a digestão das proteínas do corpo protéico do milho possibilita a identificação majoritária de proteínas não-zeínas.
- ✓ Todas as proteínas não-zeínas identificadas, independente de suas concentrações, podem contribuir para o aumento de aproximadamente 50% no conteúdo de lisina e triptofano em relação ao milho comum, apresentado pela variedade de milho BR 473.

Por último, como conclusão geral, pode-se estabelecer que:

- ➡ Os métodos de LC-MS/MS e LC/LC-MS/MS são ferramentas analíticas complementares, poderosas e confiáveis na identificação de proteínas.
- ➡ Esforços para se elevar os níveis das proteínas não-zeínas identificadas nesse estudo podem ter uma grande contribuição no aprimoramento nutricional do milho.

Capítulo V

Referências Bibliográficas

1. LÁSZTITY, R. **The chemistry of cereal proteins**. Boca Raton: CRC, 1996. 336 p.
2. SHEWRY, P. R.; TATHAM, A. S. The prolamin storage proteins of cereal seeds. Structure and evolution. **Biochemistry Journal**, v. 267, p. 1-12, 1990.
3. TATHAM, A. S.; FIELD, J. M.; MORRIS, V. J.; J'ANSON, K.; CARDLE, L.; DUFTON, M. J.; SHWERY, P. R. Solution conformational analysis of the α zein proteins of maize. **The Journal of Biological Chemistry**, v. 268, p. 26253-26259, 1993.
4. LENDING, C. R.; LARKINS, B. A. Effect in the zein composition of protein bodies during maize endosperm development. **The Plant Cell**, v. 1, p. 1011-1023, 1989.
5. PAULIS, J. W.; BIETZ, J. A. Separation of alcohol-soluble maize proteins by reversed-phase high performance liquid chromatography. **Journal of Cereal Science**, v. 4, p. 205-216, 1986.
6. COLEMAN, C. E.; CLORE, A. M.; RANCH, J. P.; HIGGINS, R.; LOPES, M. A. Expression of a mutant α -zein creates the floury2 phenotype in transgenic maize. **Plant Biology**, v. 94, p. 7094-7097, 1997.
7. OR, E.; BOYER, S. K.; LARKINS, B. A. Opaque2 modifiers act post-transcriptionally and in a polar manner on γ -zein gene expression in maize endosperm. **The Plant Cell**, v. 5, p. 1599-1609, 1993.
8. DANNENHOFFER, J. M.; BOSTWICK, D. E.; OR, E.; LARKINS, B. A. Opaque-15, a maize mutation with properties of a detective opaque-2 modifier. **Plant Biology**, v. 92, p. 1931-1935, 1995.
9. MERTZ, E. T.; BATES, L. S.; NELSON, O. E. Mutant gene that changes protein composition and increases lysine content of maize endosperm. **Science**, v. 145, p. 279-280, 1964.
10. SPRAGUE, G. F.; DUDLEY, J. W. **Corn and corn improvement**. Madison: ASA/CSSA/SSSA, 1988. 986 p.
11. TORRENT, M.; ALVAREZ, I.; GELI, M. I.; DALCOL, I.; LUDEVID, D. Lysine-rich modified gamma-zein accumulate in protein bodies in transiently transformed maize endosperm. **Plant Molecular Biology**, v. 34, p. 139-149, 1997.

12. OLIVEIRA, J. P.; CHAVES, L. J.; DUARTE, J. B.; BRASIL, E. M.; FERREIRA JÚNIOR, L. T.; RIBEIRO, K. O. Teor de proteína no grão em populações de milho de alta qualidade protéica e seus cruzamentos. **Pesquisa Agropecuária Tropical**, v. 34, p. 45-51, 2004.
13. SALDIVAR, S. O. S.; ROONEY, L. W. Quality protein maize processing and perspectives for industrial utilization. In: LARKINS, B.A.; NELSON, O.F. (Eds.). **Quality Protein Maize**: 1964-1994. Washington: Purdue University Press, 1994. p. 89-120.
14. CHANDRASHEKAR, A.; MAZHART, H. The biochemical basis and implications of grain strength in sorghum and maize. **Journal of Cereal Science**, v. 30, p. 193-207, 1999.
15. EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA) Milho e Sorgo. Disponível em: <<http://www.cnpms.embrapa.br/produtos/produtos/br473.html>>. Acesso em: 5 abr. 2007.
16. UNITED STATES DEPARTMENT OF AGRICULTURE. Cooperative State Research, Education, and Extension Service. Disponível em: <<http://www.csrees.usda.gov/index.html>>. Acesso em: 1 mar. 2007.
17. STRYER, L. **Bioquímica**. Rio de Janeiro: Guanabara Koogan, 1995. 1000 p.
18. SOUSA, M. V.; FONTES, W.; RICART, C. A. O. Análise de proteomas. **Biotecnologia Ciência & Desenvolvimento**, v. 7, p. 12-14, 1999.
19. VOET, D.; VOET, J. G. **Biochemistry**. New York: John Wiley, 1995. 1361 p.
20. NELSON, D. L.; COX, M. M. **Lehninger principles of biochemistry**. New York: Worth Publishers, 2000. 1152 p.
21. BELIZÁRIO, J. E. Proteoma: o desafio tecnológico do século 21. **Ciência Hoje**, v. 27, p. 78-81, 2000.
22. KAZMI, S.; KRULL, I. S. Proteomics and the current state of protein separations science, Part one. **Pharma Genomics**, p. 14-29, 2001.
23. CASEY, D.; CANTOR C.; SPENGLER, S. Primer on Molecular Genetics. Washington: U.S. Department of Energy, 1992. p. 1-44 (Human Genome Program). Disponível em: <http://www.ornl.gov/sci/techresources/Human_Genome/publicat/primer/toc.html>. Acesso em: 1 mar. 2007.
24. PIMENTA, A. M. C. Os desafios do proteoma. **Ciência Hoje**, v. 32, p. 16-22, 2003.

25. BRADBURY, J. Proteomics: the next step after genomics? **The Lancet**, v. 356, p. 50, 2000.
26. PENNINGTON, S. R.; WILKINS, M. R.; HOCHSTRASSER, D. F.; DUNN, M. J. Proteome analysis: from protein characterization to biological function. **Trends in Cell Biology**, v. 7, p. 168-173, 1997.
27. IDEKER, T.; THORSSON, V.; RANISH, J. A.; CHRISTMAS, R.; BUHLER, J.; ENG, J. K.; BUMGARNER, R.; GOODLETT, D. R.; AEBERSOLD, R.; HOOD, L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. **Science**, v. 292, p. 929-934, 2001.
28. PANDEY, A.; MANN, M. Proteomics to study genes and genome. **Nature**, v. 405, p. 837-846, 2000.
29. EZZELL, C. O universo mutante das proteínas. **Scientific American Brasil**, v. 1, p. 54-61, 2002.
30. GODOVAC-ZIMMERMANN, J.; BROWN, L. R. Perspectives for mass spectrometry and functional proteomics. **Mass Spectrometry Reviews**, v. 20, p. 1-57, 2001.
31. DI CIERO, L.; BELLATO, C. M. Proteoma. **Biotecnologia Ciência & Desenvolvimento**, v. 29, p. 158-164, 2002.
32. PENG, J.; GYGI, S. P. Proteomics: the move to mixtures. **Journal of Mass Spectrometry**, v. 36, p. 1083-1091, 2001.
33. SHEVCHENKO, A.; JENSEN, O. N.; PODTELEJNIKOV, A. V.; SAGLIOCCO, F.; WILM, M.; VORM, O.; MORTENSEN, P.; SHEVCHENKO, A.; BOUCHERIE, H.; MANN, M. Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. **Proceedings of the National Academy of Sciences of the United States of America**, v. 93, p. 14440-14445, 1996.
34. LANÇAS, F. M.; SILVA, J. C. R.; BICUDO, R. C.; BENASSI NETO, M. A química analítica do proteoma. **Revista Analytica**, v. 6, p. 60-66, 2003.
35. GYGI, S. P.; CORTHALS, G.L.; ZHANG, Y.; ROCHON, Y.; AEBERSOLD, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, p. 9390-9395, 2000.

36. SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. **Princípios de análise instrumental**. Porto Alegre: Bookman, 2002. 836 p.
37. RIVIER, J.; MCCLINTOCK R. Reversed-Phase High Performance Liquid Chromatography of Insulins from Different Species. **Journal of Chromatography**, v. 268, p. 112–119, 1983.
38. KUNITANI, M.; JOHNSON, D. Model of Protein Conformation in the Reversed-Phase Separation of Interleukin-2 Muteins. **Journal of Chromatography**, v. 371, p. 313–333, 1986.
39. TEMPST, P.; WOO, D.; TEPLow, D. B.; AEBERSOLD, R.; HOOD, L. E.; KENT, S. B. H. Microscale Structure Analysis of a High Molecular Weight, Hydrophobic Membrane Glycoprotein Fraction with Platelet-Derived Growth Factor-Dependent Kinase Activity. **Journal of Chromatography**, v. 359, p. 403–412, 1986.
40. GARNICK, R. L.; SOLLI, N. J.; PAPA, P. A. The Role of Quality Control in Biotechnology: An Analytical Perspective. **Analytical Chemistry**, v. 60, p. 2546–2557, 1988.
41. CHRISTIANSON, T.; PAECH, C. Peptide Mapping of Subtilisins as a Practical Tool for Locating Protein Sequence Errors during Extensive Protein Engineering Projects. **Analytical Biochemistry**, v. 223, p. 119–129, 1994.
42. CHANG, J. P.; RICHARDSON, J. M.; RIGGIN, R. M. - Characterization of Tryptic Peptides of a Potent Growth Hormone Releasing Hormone Analog by Reversed Phase High Performance Liquid Chromatography-Ionspray Mass Spectrometry. **Journal of Liquid Chromatography**, v. 17, p. 2881–2894, 1994.
43. SCARBOROUGH, R. M.; JAMIESON, G. C.; KALISH, F.; KRAMER, S. J.; MCENROE, G. A.; MILLER, C. A.; SCHOOLEY, D. A. - Isolation and primary structure of two peptides with cardioacceleratory and hyperglycemic activity from the corpora cardiaca of *Periplaneta americana*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 81, p. 5575–5579, 1984.
44. RIVIER, J.; MCCLINTOCK, R.; GALYEAN, R.; ANDERSON, H. Reversed-Phase High Performance Liquid Chromatography: Preparative Purification of Synthetic Peptides. **Journal of Chromatography**, v. 288, p. 303–328, 1984.

45. HOEGER, C. A.; GALYEAN, R.; MCCLINTOCK, R. A.; RIVIER, J. E. Practical Aspects of Preparative Reversed-phase Chromatography of Synthetic Peptides. In: MANT C.T.; RODGES R.S. (Eds.). **HPLC of peptides and proteins: separation, analysis and conformation**. Boca Raton: CRC, 1991. p. 753-764.
46. GENG, X.; REGNIER, F. E. Retention Model for Proteins in Reversed-Phase Liquid Chromatography. **Journal of Chromatography**, v. 296, p. 15-30, 1984.
47. CARR, D. **The handbook of analysis and purification of peptides and proteins by reversed-phase HPLC**. p. 1-70, 2002. Disponível em: <<http://www.vydac.com/vydacpubs/brindex.html>>. Acesso em: 7 mar. 2007.
48. VYDAC, G. **Principles and applications of high-performance ion-exchange chromatography for bioseparations**. p. 1-14, 1998. Disponível em: <<http://www.vydac.com/vydacpubs/brindex.html>>. Acesso em: 7 mar. 2007.
49. KOPACIEWICZ, W.; ROUNDS, M. A.; FAUSNAUGH, J.; REGNIER, F. E. Retention Model for High-Performance Ion-Exchange Chromatography. **Journal of Chromatography**, v. 266, p. 3-21, 1983.
50. HOLM, A.; STORBRATEN, E.; KARASZEWSKI, A. M. B.; LUNDANES, E.; GREIBROKK, T. Combined solid-phase extraction and 2D LC-MS for characterization of the neuropeptides in rat-brain tissue. **Analytical and Bioanalytical Chemistry**, v. 382, p. 751-759, 2005.
51. YATES, J. R.; CARMACK, E.; HAYS, L.; LINK A. J.; ENG, J. K. Automated protein identification using microcolumn liquid chromatography-tandem mass spectrometry. **Methods in Molecular Biology**, v. 112, p. 553-569, 1999.
52. GIDDINGS, J. C. Concepts and Comparisons in Multidimensional Separations. **Journal of High Resolution Chromatography and Chromatography Communications**, v. 10, p. 319-323, 1987.
53. LIEBLER, D. C. **Introduction to proteomics: tools for the new biology**. Totowa: Humana Press, 2002. 198 p.
54. MORAES, M. C. B.; LAGO, C. L. Espectrometria de massas com ionização por "electrospray" aplicada ao estudo de espécies inorgânicas e organometálicas. **Química Nova**, v. 26, n. 4, p. 556-563, 2003.

55. DOLE, M.; MACK, L. L.; HINES, R. L.; MOBLEY, R. C.; FERGUNSON, L. D.; ALICE, M. B. Molecular beams of microions. **The Journal of Chemical Physics**, v. 49, p. 2240-2249, 1968.
56. YAMASHITA, M.; FENN, J. B. Negative ion production with the electrospray ion source. **Journal of Physical Chemistry**, v. 88, p. 4671-4675, 1984.
57. COLE, R. B. **Electrospray ionization mass spectrometry: fundamentals, instrumentation, and applications**. New York: John Wiley, 1997. 577 p.
58. LANE, C. S. Mass spectrometry-based proteomics in the life sciences. **Cellular and Molecular Life Sciences**, v. 62, p. 848-869, 2005.
59. KEBARLE, P.; TANG, L. From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. **Analytical Chemistry**, v. 65, p. 972A-986A, 1993.
60. TANG, L.; KEBARLE, P. Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analytes in the electrosprayed solution. **Analytical Chemistry**, v. 65, p. 3654-3668, 1993.
61. THOMSON, B. A.; IRIBARNE, J. V. Field induced ion evaporation from liquid surfaces at atmospheric pressure. **The Journal of Chemical Physics**, v. 71, p. 4451-4463, 1979.
62. IRIBARNE, J. V.; THOMSON, B. A. On the evaporation of small ions from charged droplets. **The Journal of Chemical Physics**, v. 64, p. 2287-2294, 1976.
63. BUSTILLOS, O. V.; SASSINE, A.; MARCH, R. **A espectrometria de massas quadrupolar**. São Paulo: Editora Scortecci , 2003. 162 p.
64. CHAKRAVARTI, D. N.; CHAKRAVARTI, B.; MOUTSATSOS, I. Informatic tools for proteome profiling. **Computational Proteomics Supplement**, v. 32, p. 4-15, 2002.
65. BICUDO, T. C. **Purificação e Caracterização das Zeínas γ em Corpos Protéicos de Milho por Técnicas Espectroscópicas**. 2005. 124 f. Tese (Doutorado em Físico-Química) - Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2005.
66. KINTER, M.; SHERMAN, N. E. **Protein Sequencing and Identification Using Tandem Mass Spectrometry**. New York: Wiley, 2000. p. 160.

67. WATERS CORPORATION. Disponível em: <<http://www.waters.com>>. Acesso em: 1 mar. 2007.
68. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (NCBI) Disponível em: <<http://www.ncbi.nlm.nih.gov/>>. Acesso em: 21 set. 2006.
69. MATRIX SCIENCE. Disponível em: <<http://www.matrixscience.com>>. Acesso em: 21 mar. 2007.
70. LINK, A. J.; ENG, J.; SCHIELTZ, D. M.; CARMACK, E.; MIZE, G. J.; MORRIS, D. R.; GARVIK, B. M.; YATES III, J. R. Direct analysis of protein complexes using mass spectrometry. **Nature Biotechnology**, v. 17, p. 676-682, 1999.
71. WASHBURN, M. P.; WOLTERS, D.; YATES III, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. **Nature Biotechnology**, v. 19, p. 242-247, 2001.
72. FLORENS, L.; WASHBURN, M. P.; RAINE, J. D.; ANTHONY, R. M.; GRAINGERK, M.; HAYNES, J. D.; MOCH, J. K.; MUSTER, N.; SACCI, J. B.; TABB, D. L.; WITNEY, A. A.; WOLTERS, D.; WU, Y.; GARDNER, M. J.; HOLDERK, A. A.; SINDEN, R. E.; YATES, J. R.; CARUCCI, D. J. A proteomic view of the Plasmodium falciparum life cycle. **Nature**, v. 419, p. 520-526, 2002.
73. LASONDER, E.; ISHIHAMA, Y.; ANDERSEN, J. S.; VERMUNT, A. M. W.; PAIN, A.; SAUERWEIN, R. W.; ELING, W. M. C.; HALL, N.; WATERSK, A. P.; STUNNENBERG, H. G.; MANN, M. Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. **Nature**, v. 419, p. 537-542, 2002.
74. ADKINS, J. N.; VARNUM, S. M.; AUBERRY, K. J.; MOORE, R. J.; ANGELL, N. H.; SMITH, R. D.; SPRINGER, D. L.; POUNDS, J. G. Toward a Human Blood Serum Proteome – Analysis by multidimensional separation coupled with mass spectrometry. **Molecular & Cellular Proteomics**, v. 1, p. 947-955, 2002.
75. STRITTMATTER, E. F.; RODRIGUEZ, N.; SMITH, R. D. High Mass Measurement Accuracy Determination for Proteomics Using Multivariate Regression Fitting: Application to Electrospray Ionization Time-Of-Flight Mass Spectrometry. **Analytical Chemistry**, v. 75, p. 460-468, 2003.
76. O'NEIL, K. A.; MILLER, F. R.; BARDER, T. J.; LUBMAN, D. M. Profiling the progression of cancer: Separation of microsomal proteins in MCF10 breast epithelial cell lines using nonporous chromatophoresis. **Proteomics**, v. 3, p. 1256-1269, 2003.

77. KAPP, E. A.; SCHUTZ, F.; REID, G. E.; EDDER, J. S.; MORITZ, R. L.; O'HAIR, R. A. J.; SPEED, T. P.; SIMPSON, R. J. Mining a Tandem Mass Spectrometry Database To Determine the Trends and Global Factors Influencing Peptide Fragmentation. **Analytical Chemistry**, v. 75, p. 6251-6264, 2003.
78. VAISAR, T.; URBAN, J. Probing the Proline Effect in CID of Protonated Peptides. **Journal of Mass Spectrometry**, v. 31, p. 1185-1187, 1996.
79. STEEN, H.; MANN, M. The ABC'S (and XYZ'S) of Peptide Sequencing. **Nature Reviews**, v. 5, p. 699-711, 2004.
80. SCHLEY, C.; ALTMAYER, M. O.; SWART, R.; MULLER, R.; HUBER, C. G. Proteome Analysis of *Myxococcus xanthus* by Off-Line Two-Dimensional Chromatographic Separation Using Monolithic Poly-(styrene-divinylbenzene) Columns Combined with Ion-Trap Tandem Mass Spectrometry. **Journal of Proteome Research**, v. 5, p. 2760-2768, 2006.
81. INSTITUTE FOR SYSTEMS BIOLOGY. The Proteomics Tool Kit. Disponível em: <<http://db.systemsbio.org.net:8080/proteomicsToolkit/IsotopeServlet.html>>. Acesso em: 5 abr. 2007.
82. AZAMA, K.; ABE, S.; SUGIMOTO, H. Lysine-containing proteins in maize endosperm: a major contribution from cytoskeleton-associated carbohydrate-metabolizing enzymes. **Planta**, v. 217, p. 628-638, 2003.
83. MANJUNATH, S.; SACHS, M. M. Molecular characterization and promoter analysis of the maize cytosolic glyceraldehyde 3-phosphate dehydrogenase gene family and its expression during anoxia. **Plant Molecular Biology**, v. 33, p. 97-112, 1997.
84. CARNEIRO, N. P.; HUGHES, P. A.; LARKINS, B. A. The eEF1A gene family is differentially expressed in maize endosperm. **Plant Molecular Biology**, v. 41, p. 801-813, 1999.
85. BATHGATE, B.; BAKER, A.; LEAVER, C. J. Two genes encode the adenine nucleotide translocator of maize mitochondria. Isolation, characterisation and expression of the structural genes. **European Journal of Biochemistry**, v. 183, p. 303-310, 1989.