

## SEQÜENCIAMENTO DE PEPTÍDEOS USANDO ESPECTROMETRIA DE MASSAS: UM GUIA PRÁTICO

Marcelo Delmar Cantú e Emanuel Carrilho\*

Instituto de Química de São Carlos, Universidade de São Paulo, CP 780, 13560-970 São Carlos – SP, Brasil

Nelson Arno Wulff

Fundo de Defesa da Citricultura, Av. Adhemar Pereira de Barros, 201, 14807-040 Araraquara - SP, Brasil

Mario Sérgio Palma

Departamento de Biologia, Instituto de Biociências de Rio Claro, Universidade Estadual Paulista, Av. 24 A, 1515, 13506-900 Rio Claro - SP, Brasil

Recebido em 6/10/06; aceito em 17/8/07; publicado na web em 19/3/08

PEPTIDE SEQUENCING USING MASS SPECTROMETRY: A PRACTICAL GUIDE. This paper introduces the basics of peptide mass spectra interpretation applied to proteomics and is directed to chemists, biochemists and biologists. The manuscript presents a well detailed protocol aiming to serve as a first choice guide for understanding peptide sequencing. The tutorial was elaborated based on both a thorough bibliographic revision and the author's experience. In order to prove the applicability of the proposed guide, spectra obtained on different instruments have been successfully interpreted by applying the presented rational.

Keywords: proteomic analysis; mass spectrometry; peptide sequencing.

### INTRODUÇÃO

A análise proteômica, definida como sendo o conjunto de metodologias analíticas empregadas para caracterizar (quali e quantitativamente) um proteoma, trata-se de uma área interdisciplinar da ciência, a qual agrega principalmente química, biologia e informática. O sinergismo oriundo de tamanha interdisciplinaridade faz-se necessário num cenário onde se pretende estudar a função/comportamento dos genes com base nas identificações das proteínas por eles expressas. Neste contexto, muitas vezes é necessário não somente determinar o conjunto de proteínas presentes em uma amostra, o que por si só é algo bastante desafiador, mas também caracterizar as inúmeras e comumente presentes isoformas das proteínas, produtos de modificações pós-traducionais sofridas pelas mesmas e, por fim, como essas proteínas interagem entre si.<sup>1,2</sup> Devidamente dimensionada a complexidade do assunto, a espectrometria de massas (MS) emerge como uma tecnologia indispensável para a interpretação da informação codificada pelos genes, ou seja, o proteoma.

Uma das forças que impulsiona a proteômica é a habilidade de usar dados de espectrometria de massas inerentes a peptídeos para identificar proteínas em bancos de dados. Para tal fim, dois tipos de resultados são usados. O primeiro usa a informação relativa à massa molecular dos peptídeos oriundos da digestão enzimática (*Peptide Mass Fingerprint* – PMF), enquanto o segundo faz uso de resultados obtidos pela fragmentação de peptídeos individuais previamente detectados.<sup>3</sup>

### ESPECTROMETRIA DE MASSAS APLICADA À ANÁLISE PROTEÔMICA

Em linhas gerais, a MS é uma técnica capaz de determinar a relação entre massa e carga ( $m/z$ ) de espécies ionizadas em fase gasosa.<sup>2</sup> Um espectrômetro de massas é um instrumento constituído por uma fonte de íons, um analisador de massas, um detector e um

sistema de aquisição de dados. As fontes de ionização empregadas em MS aplicada à análise proteômica são *Electrospray* (ESI)<sup>4</sup> e *MALDI* (*Matrix-Assisted Laser Desorption Ionization*),<sup>5</sup> tendo a função de ionizar (de maneira suave, preservando assim a estrutura polipeptídica) e transferir as espécies a serem analisadas para a fase gasosa. Os analisadores de massas, como o próprio nome indica, têm como função básica separar os íons formados de acordo com suas relações  $m/z$ . Diversos analisadores de massas, tais como, quadrupolos, *ion-traps* (tridimensionais e lineares), *time-of-flight* (TOF), *Fourier-transform ion cyclotron resonance* (FT-ICR), *orbitrap*, entre outros, são comercialmente disponíveis e cada um possui aspectos positivos e negativos, de acordo com o experimento planejado e o resultado experimental requerido. Estes analisadores podem ser usados “sozinhos” e de maneira independente ou acoplados entre si, dando origem a equipamentos classificados como híbridos, os quais fazem uso das vantagens inerentes a cada analisador. Tais equipamentos permitem que experimentos em seqüência (*tandem*) sejam realizados, isto é, sendo possível detectar um determinado íon e posteriormente submetê-lo a uma etapa de fragmentação. Uma vez separados, esses íons são detectados por eletromultiplicadoras que constituem os detectores mais largamente usados. Uma ilustração dos processos é mostrada na Figura 1S em material suplementar

De maneira geral, é possível descrever um estudo proteômico empregando MS em seis etapas, descritas a seguir:

- i) as proteínas a serem analisadas devem ser primeiramente isoladas ou extraídas de lisados celulares ou tecidos. Tal procedimento comumente emprega metodologias de extração ou fracionamento, definindo o “sub-proteoma” a ser estudado. Como exemplo prático pode-se citar uma área que tem a cada dia recebido maior atenção, a busca de biomarcadores para doenças, fazendo uso da análise proteômica aplicada ao plasma sanguíneo. A primeira e mais importante etapa consiste da eliminação nas proteínas mais abundantes do plasma, uma vez que as 10 proteínas mais abundantes no plasma humano correspondem a aproximadamente 90% do conteúdo total. Assim, quando biomarcadores para doenças são o alvo do estudo é necessário reduzir a quantidade das proteínas altamente abun-

\*e-mail: emanuel@iqsc.usp.br

- dantes, que acabam interferindo na análise.<sup>6</sup>
- ii) Após a etapa de purificação, o próximo passo é converter a(s) proteína(s) isolada(s) em um conjunto de peptídeos. Isso é feito com o uso de enzimas que promovem a clivagem das proteínas em pontos específicos. Quando a análise proteômica é realizada por meio da análise de peptídeos oriundos da digestão enzimática de proteínas, esse arranjo experimental recebe o nome genérico de análise proteômica *bottom-up*.<sup>7,8</sup> Mesmo considerando que os espectrômetros de massas podem determinar a massa molecular de proteínas intactas, existem inúmeras razões que justificam o uso de peptídeos e não proteínas para a análise proteômica. Dentre essas razões estão: de forma geral, proteínas são difíceis de manusear e degradam-se com facilidade, podendo ainda apresentar problemas de solubilidade. Assim, em muitos casos faz-se necessária a adição de tensoativos que comprovadamente interferem com a análise por MS, uma vez que muitos desses componentes ionizam muito bem e quase sempre estão presentes em excesso na amostra. A sensibilidade dos espectrômetros de massas para a análise de proteínas é consideravelmente menor que para peptídeos. Além disso, se o interesse da análise é a identificação das proteínas, informação inerente à sequência é necessária, e nesse sentido os espectrômetros de massas são mais eficientes para obter informação estrutural de peptídeos que possuem até 20 aminoácidos em comparação a proteínas intactas. Entretanto é necessário esclarecer que com o uso de espectrômetros de massas de última geração, tal como o FT-ICR, é possível obter informações parciais da sequência primária de proteínas intactas, que obviamente podem ser usada para fins de identificação ou análise de modificações pós-traducionais, num arranjo experimental referido como análise proteômica *top down*.<sup>8,9</sup>
  - iii) Os peptídeos obtidos podem ser separados por meio das técnicas de cromatografia líquida uni- ou multidimensional, ionizados e transferidos (ESI ou MALDI) para o analisador de massas. ESI aplicada à análise de peptídeos produz preferencialmente espécies duplamente carregadas, enquanto MALDI gera quase que exclusivamente íons monocarregados.
  - iv) Nesta etapa o espectro de massas dos peptídeos oriundos da digestão enzimática é adquirido. Este resultado indica a relação  $m/z$  e, por consequência, a massa molecular dos peptídeos. Para esse resultado dá-se o nome de *peptide mass fingerprint* (PMF).
  - v) Os peptídeos previamente detectados durante o PMF (chamados de íons precursores) são então isolados e submetidos à fragmentação por colisão com moléculas de um gás inerte, tal como argônio, nitrogênio ou hélio. O espectro obtido é chamado espectro de fragmentação ou MS/MS.
  - vi) Ao final do processo, os resultados inerentes a massa molecular (MM) dos peptídeos, obtida a partir do PMF, bem como a informação relativa a sequência de aminoácidos dos peptídeos, contida nos espectros de fragmentação (MS/MS), são usados pelos softwares de busca para “localizar” as proteínas nos bancos de dados. Os softwares mais conhecidos e comumente empregados são o Sequest<sup>10</sup> e o Mascot<sup>11</sup> (veja Figura 2S).

## FERRAMENTAS PARA IDENTIFICAÇÃO DE PROTEÍNAS EM BANCOS DE DADOS

Como apontado no parágrafo anterior, os programas mais comumente empregados para a identificação de proteínas em bancos de dados a partir de dados de MS são o Sequest e o Mascot. Ambos os programas correlacionam espectros de massas de fragmentação (não interpretados) de peptídeos com sequências de aminoácidos de proteínas registradas em bancos de dados.<sup>12,13</sup> Além

disso, esses softwares também têm a capacidade de usar sequências de nucleotídeos para fazer tal correlação. Para tal, eles primeiramente simulam as sequências primária potenciais das proteínas correspondentes àquelas sequências de nucleotídeos encontradas nos bancos de genes, utilizando-se do código genético universal; posteriormente, simulam a fragmentação destas sequências primárias. De forma geral, estes programas têm como objetivo encontrar a sequência de aminoácidos, em um determinado banco de dados, que melhor descreve os íons fragmentos encontrados em um espectro. As sequências “candidatas” são procuradas nos bancos de dados de acordo com a massa do peptídeo intacto (informação adquirida na etapa do PMF) e com o espectro de fragmentação obtido para cada peptídeo.

No Sequest, uma técnica de processamento do sinal chamada autocorrelação é usada a fim de determinar matematicamente a sobreposição entre o espectro teórico, derivado de cada sequência obtida no banco de dados em questão, e o espectro experimentalmente obtido. O resultado de tal sobreposição é expresso quantitativamente em termos de uma *score* para cada peptídeo (Xcorr). O Xcorr é um parâmetro que depende de diversos fatores, tais como o estado de carga do peptídeo bem como do tamanho do banco de dados que está sendo usado para a busca. Assim, a avaliação de um segundo *score*, classificado como  $\Delta Cn$  faz-se necessária para que a confiabilidade do resultado obtido seja aumentada. Esse parâmetro é definido como sendo a diferença entre os valores de Xcorr obtidos para a sequência de aminoácidos que obteve o maior Xcorr e a sequência subsequente. Na literatura, diferentes critérios são usados para classificar uma determinação como satisfatória ou não. De forma geral, estes valores são: Xcorr > 3,75 para peptídeos com carga +3; Xcorr > 2,2 para peptídeos com carga +2 e Xcorr > 1,9 para peptídeos com carga +1. Em todos os casos descritos,  $\Delta Cn > 0,10$  é exigido para que a determinação seja considerada suficientemente confiável.<sup>3,14</sup> O Sequest tem se mostrado uma ferramenta bastante robusta, inclusive quando espectros com baixa relação sinal ruído são submetidos à análise.<sup>3,10</sup>

O Mascot também envolve o cálculo de fragmentos teoricamente preditos para todos os peptídeos de um banco de dados de acordo com a massa do íon precursor, previamente determinada. Os valores de  $m/z$  dos fragmentos preditos são comparados com os fragmentos experimentais sendo que, neste caso, a comparação se inicia com base nos íons  $-b$  e  $-y$  mais intensos. A probabilidade de o valor de  $m/z$  de um fragmento teoricamente obtido coincidir, de maneira randômica, com o valor de  $m/z$  de um fragmento obtido experimentalmente é calculada e expressa como sendo o negativo do logaritmo desse número (*score*). Assim, quanto maior for o valor obtido, menor é a probabilidade de que este resultado seja fruto de uma “coincidência”. Esse software fornece para cada busca submetida um valor limite (dependendo das condições usadas para a busca) a partir do qual o valor obtido indica que a determinação possui probabilidade inferior a 5% de ser um evento randômico.<sup>11</sup>

Uma vez entendida a sistemática aplicada pelos softwares para a identificação de proteínas em bancos de dados usando dados de espectrometria de massas, faz-se extremamente necessário e de suma importância o completo entendimento de como ocorre a fragmentação dos peptídeos. Além disso, a interpretação manual de espectros de MS/MS é recomendada em todos os casos e indispensável em algumas situações. Por fim, existem situações nas quais o genoma de uma determinada espécie ainda não está completamente sequenciado ou disponível e, neste cenário, é necessário derivar a sequência primária de aminoácidos de um determinado peptídeo baseado única e exclusivamente nos dados obtidos por espectrometria de massas, isto é, sem recorrer a banco de dados (sequenciamento “*de novo*”).<sup>15</sup>

## FRAGMENTAÇÃO E SEQÜENCIAMENTO DE PEPTÍDEOS

A fragmentação de peptídeos por espectrometria de massas para a posterior análise de sua seqüência de aminoácidos é comumente realizada por meio do processo de dissociação induzida por colisão (*collision induced dissociation* – CID),<sup>16</sup> também referida por alguns autores como dissociação ativada por colisão (*collision activation dissociation* – CAD). Apesar de outras metodologias para a fragmentação de peptídeos, tais como *Electron Capture Dissociation* (ECD); *Electron Transfer Dissociation* (ETD)<sup>17</sup> terem sido desenvolvidas, CID é sem dúvida o mais largamente empregado, além de ser o método mais freqüentemente aplicado nos espectrômetros de massas comercialmente disponíveis. Neste processo, os peptídeos são inicialmente introduzidos em uma região de vácuo do espectrômetro de massas por meio dos processos de *electrospray* ou MALDI. Usando uma descrição simplista e abrangente, cabível para a maior parte dos equipamentos comercialmente disponíveis, os peptídeos ionizados são acelerados para uma região do espectrômetro preenchida com um gás inerte (hélio, argônio ou nitrogênio) proporcionando, assim, a colisão entre os peptídeos ionizados e as moléculas do gás inerte. Como resultado, a energia translacional transferida em cada colisão é convertida em energia interna. O modelo da mobilidade do próton<sup>18</sup> descreve como a energia interna adquirida induz a transferência intramolecular dos prótons em cada peptídeo, culminando na desestabilização das ligações do esqueleto polipeptídico e, por conseqüência, induzindo a formação de dois íon-fragmentos,<sup>19</sup> que são classificados como íons que retêm a carga residual (próton) no lado N-terminal (gerando fragmentos *-a*, *-b* e *-c*, dependendo da ligação que é fragmentada); íons que retêm a carga residual (próton) na região C-terminal (gerando os fragmentos *-x*, *-y* e *-z*, dependendo da ligação que é fragmentada), segundo a nomenclatura proposta por Roepstorff–Fohlmann–Biemann.<sup>20</sup> É importante enfatizar que os pares de íons *a/x*, *b/y* e *c/z* serão sempre íons correspondentes aos fragmentos opostos e complementares entre si. Considerando-se que as ligações peptídicas são aquelas menos energéticas, espera-se que a formação do par de fragmentos *-b/-y* seja mais freqüente que os demais pares de fragmentos, facilitando muito a interpretação dos espectros.

Apesar de diversos estudos no sentido de definitivamente compreender o mecanismo de fragmentação de peptídeos usando CID terem sido realizados, o mecanismo é ainda não completamente entendido.<sup>16,21,22</sup>

Como resultado da fragmentação das ligações peptídicas mencionada acima, uma série de íons *-b* e *-y* complementares é obtida, de modo que a diferença de valores de *m/z* entre dois íons consecutivos do mesmo tipo pode revelar a identidade do resíduo de aminoácido em questão. Enquanto as séries *-b* e *-y* resultam diretamente da clivagem das ligações peptídicas, os íons *-a* são formados pela perda neutra de monóxido de carbono dos íons *-b* (diferença de 27.9949 u relativo ao íon *-b* correspondente).<sup>16,18</sup> Considerando todos os íons que teoricamente podem ser produzidos em condições de CID, os íons *-b* e *-y* correspondem a grande maioria dos íons observados, enquanto os íons *-a* são menos comuns. Vale ainda dizer que para as situações mais comumente enfrentadas em um estudo proteômico, ou seja, a análise de peptídeos oriundos de digestão triptica (**R** ou **K** na posição C-terminal) a formação da série de íons *-y* é favorecida (em relação à série *-b*) devido à elevada basicidade desses resíduos de aminoácidos.<sup>16</sup> A energia comumente usada para induzir a fragmentação dos peptídeos em CID geralmente é insuficiente para romper as ligações entre o carbono- $\alpha$  e o carbono da carbonila, bem como entre o nitrogênio e o

carbono- $\alpha$  adjacente, de modo que os íons *-c*, *-x* e *-z* são tipicamente não observados no espectro.<sup>16</sup>

Quando a fragmentação ocorre simultaneamente nas posições amino e carboxi-terminal do mesmo resíduo de aminoácido, íons imônio são produzidos (Tabela 1). Esses íons servem como íons diagnóstico, podendo indicar a presença ou ausência de determinados aminoácidos na seqüência em estudo. Em certos resíduos de aminoácidos, perdas neutras de moléculas de H<sub>2</sub>O e NH<sub>3</sub> são freqüentemente verificadas. Por exemplo, **S**, **T** e **E** são aminoácidos que quando presentes em peptídeos submetidos à fragmentação por CID apresentarão perda neutra de H<sub>2</sub>O bastante pronunciada. Por outro lado, **R**, **K**, **Q** e **N** são exemplos de aminoácidos que apresentam pronunciada perda neutra de NH<sub>3</sub>. Além disso, determinadas modificações pós-traducionais ocorridas nas cadeias laterais de certos aminoácidos, tais como fosforilação de **S** e **T**; glicosilação e/ou oxidação de **M** tornam tais grupos laterais lábeis, de modo que a perda neutra destes íons pode ser observada. Como exemplo, pode-se verificar se a **S** constituinte de um determinado peptídeo apresenta ou não uma fosforilação; para isso, deve-se verificar se há um pico com massa 98 u inferior ao íon correspondente (*-b* ou *-y*) sendo que para esse íon dá-se o nome de íon satélite<sup>15</sup> (veja Figura 3S).

No entanto, apesar da determinação da seqüência de aminoácidos em um peptídeo ser possível por meio do simples cálculo da diferença de massa entre picos vizinhos em uma série de íons, tal trabalho é bastante difícil devido a uma série de fatores, dentre os quais pode-se citar o conjunto de íons fragmento esperado pode não estar presente na íntegra, ou em outras palavras, pode haver a ausência de alguns íons das séries *-b* e *-y*; alguns fragmentos podem sofrer rearranjos internos e subsequente fragmentação; os íons podem estar presente com diferentes estados de carga, dificultando a correta atribuição dos íons (tal dificuldade aplica-se na interpretação de espectros que não são de-convoluídos); alguns fragmentos podem sofrer rearranjo neutro de hidrogênios durante a fragmentação. Assim, o somatório destes fatores pode induzir a atribuição errada das séries de íons, tornando a interpretação do espectro bastante desafiadora.

Desta forma, uma série de regras básicas, compiladas a partir de informações adquiridas na literatura<sup>15,23</sup> bem como fazendo uso da experiência própria adquirida durante o trabalho realizado nessa área do conhecimento, foi elaborada e apresentada a seguir, com o objetivo de estabelecer um guia geral para a interpretação/confirmação de seqüências de aminoácidos obtidos pela fragmentação de peptídeos por espectrometria de massas. Tais informações são apresentadas de maneira bastante abrangente, de modo que podem ser aplicadas para a interpretação de espectros obtidos por diferentes instrumentos, tais como Q-TOF, TOF-TOF, triplo quadrupolos e *ion-traps*. A Tabela 1 apresenta informações inerentes à massa molecular dos resíduos de aminoácidos e íons imônio enquanto a Tabela 2 traz a massa de dipeptídeos, uma informação bastante útil para a atribuição de íons *-b<sub>2</sub>*.

## DETALHAMENTO DO PROCEDIMENTO SUGERIDO PARA O SEQÜENCIAMENTO “DE NOVO” DE PEPTÍDEOS

### Composição de aminoácidos – diagnóstico dos aminoácidos constituintes do peptídeo

Verificar a presença dos íons imônio (Tabela 1) na região de baixas massas do espectro. Tais íons podem fornecer informações inerentes à composição de aminoácidos de um peptídeo. No entanto, é importante ter em mente que se o íon imônio para um determinado aminoácido não estiver presente, isso não significa que o aminoácido

**Tabela 1.** Lista dos 20 resíduos de aminoácidos mais comuns<sup>26</sup>

Aminoácido			Massa média	Massa monoisotópica	Íon imônio	Íon relacionados
Glicina	Gly	G	57,052	57,02146	30	
Alanina	Ala	A	71,079	71,03711	44	
Serina	Ser	S	87,078	87,03203	60	
Prolina	Pro	P	97,117	97,05276	70	
Valina	Val	V	99,133	99,06841	72	
Treonina	Thr	T	101,105	101,04768	74	
Cisteína	Cys	C	103,145	103,00919	76	
Leucina	Leu	L	113,160	113,08406	86	72
Isoleucina	Ile	I	113,160	113,08406	86	72
Asparagina	Asn	N	114,104	114,04293	87	70
Ácido aspártico	Asp	D	115,089	115,02694	88	
Glutamina	Gln	Q	128,131	128,05858	101	84, 129
Lisina	Lys	K	128,174	128,09496	101	70, 84, 112, 129
Ácido glutâmico	Glu	E	129,116	129,04259	102	
Metionina	Met	M	131,199	131,04048	104	61
Histidina	His	H	137,141	137,05891	110	82, 121, 123, 138, 166
Fenilalanina	Phe	F	147,177	147,06841	120	91
Arginina	Arg	R	156,188	156,10111	129	59, 70, 73, 87, 100, 112
Tirosina	Tyr	Y	163,176	163,06333	136	91, 107
Triptofano	Trp	W	186,213	186,07931	159	117, 130, 170, 171

**Tabela 2.** Lista das massas de dipeptídeos, úteis para a determinação de íons  $-b_2$ . Os valores correspondem à soma das massas dos resíduos dos aminoácidos acrescidos de uma unidade<sup>23</sup>

	G	A	S	P	V	T	C	I/L	N	D	Q/K	E	M	H	F/M*	R	C**	Y	W	
	57	71	87	97	99	101	103	113	114	115	128	129	131	137	147	156	161	163	186	
G	57	114																		
A	71	129	143																	
S	87	145	159	175																
P	97	155	169	185	195															
V	99	157	171	187	197	199														
T	101	159	173	189	199	201	203													
C	103	161	175	191	201	203	205	207												
I/L	113	171	185	201	211	213	215	217	227											
N	114	172	186	202	212	214	216	218	228	229										
D	115	173	187	203	213	215	217	219	229	230	231									
Q/K	128	186	200	216	226	228	230	232	242	243	244	257								
E	129	187	201	217	227	229	231	233	243	244	245	258	259							
M	131	189	203	219	229	231	233	235	245	246	247	260	261	263						
H	137	195	209	225	235	237	239	241	251	252	253	266	267	269	275					
F/M*	147	205	219	235	245	247	249	251	261	262	263	276	277	279	285	295				
R	156	214	228	244	254	256	258	260	270	271	272	285	286	288	294	304	313			
C**	161	219	233	249	259	261	263	265	275	276	277	290	291	293	299	309	318	323		
Y	163	221	235	251	261	263	265	267	277	278	279	292	293	295	301	311	320	325	327	
W	186	244	258	274	284	286	288	290	300	301	302	315	316	318	324	334	343	348	350	361

\*Oxidação; \*\* carbamidometilcisteína

está ausente da sequência. Seguindo a mesma linha de raciocínio, a presença de um íon com massa concordante com algum íon imônio não determina por certo a presença do aminoácido, uma vez que tal íon pode corresponder, por exemplo, ao um íon-fragmento oriundo de rearranjo sofrido pelo peptídeo que por coincidência possui valor de massa igual à de um determinado íon imônio.

Obs.: Para instrumentos do tipo *ion-traps*, tal informação é em muitos casos total ou parcialmente perdida devido à limitação que estes equipamentos apresentam para a determinação de íons com valores de  $m/z$  inferiores a 1/3 da relação  $m/z$  do íon precursor (regra do 1/3).<sup>24</sup> No entanto, avanços recentes nessa área estão sendo realizados de modo que tal limitação será contornada num futuro bem próximo.

### Determinação do aminoácido presente na posição C-terminal – aplicada a peptídeos obtidos por clivagem enzimática

Se os peptídeos a serem sequenciados são oriundos de digestão trípica, deve-se verificar a existência dos íons diagnóstico  $-y_i$ : 147 para **K** ou 175 para **R**. Um vez verificada a presença de um destes íons ( $-y_i$ ), determinar o correspondente íon  $b_{n-1}$  (na região de alta massa) por meio da seguinte relação:  $b_{n-1} = (M+H)^{1+} - y_i + 1$

### Verificação/confirmação da presença do íon $-b_2$

Para tal, pode-se fazer uso da Tabela 2, a qual traz uma lista



das massas de dipeptídeos ionizados. Geralmente esses íons podem ser identificados por meio da seguinte razão: íon  $-b_2$ / íon  $-a_2$  separados por 28 u (inerente à perda neutra de CO por parte de um íon **b**). Novamente, uma vez encontrada a razão  $m/z$  do íon  $-b_2$ , esta é usada para calcular a  $m/z$  do correspondente íon  $-y_{m-2}$  fazendo uso da relação:  $y_{m-2} = (M+H)^{1+} - b_2 + 1$

Em instrumentos do tipo *ion-traps* tal informação pode, em alguns casos, não ser medida.

### Extensão das seqüências de íons $-b$ e $-y$

Tendo definido os aminoácidos posicionados nas extremidades do peptídeo e usando a massa dos resíduos dos aminoácidos, iniciar o seqüenciamento analisando a região de altas massas do espectro. O menor número de picos nessa região do espectro irá tornar o trabalho mais simples. No entanto, deve-se ter cuidado com a região em torno de 60 u abaixo do íon precursor, que pode ser confundida com picos referentes a múltiplas perdas de água e amônia. Todavia, não se pode descartar a hipótese que **G** pode ser o primeiro resíduo de aminoácido da seqüência, de modo que o pico inerente a essa possibilidade pode estar presente.

A partir desse ponto pode-se sistematicamente estender as seqüências de íons  $-b$  ou  $-y$ . Em outras palavras, a partir de um determinado íon (seja ele  $-b$  ou  $-y$ ) basta acrescentar ou subtrair (dependendo da massa do íon em questão) a massa dos resíduos de aminoácidos sucessivamente a partir da **G** até o **W**. Uma vez determinada a massa de um íon  $-b$  ou  $-y$ , o correspondente íon  $-y$  ou  $-b$  pode ser calculado usando as seguintes relações gerais:  $y_{m-n} = (M+H)^{1+} - b_n + 1$  e  $b_{n-m} = (M+H)^{1+} - y_m + 1$ .

Sempre que um íon determinado (por exemplo, um íon  $-b$ ) apresenta o íon correspondente (íon  $-y$ ), a determinação ganha muito em confiabilidade.

Considerando que o íon  $-b_1$  é raramente observado no espectro, a determinação da ordem dos dois primeiros aminoácidos da região N-terminal é bastante difícil. Uma solução para tal problema é a determinação do aminoácido N-terminal empregando-se a Química Degradativa de Edman, estratégia plausível desde que a proteína em questão não apresente o aminogruppo N-terminal bloqueado.

### INFORMAÇÕES PERTINENTES QUE CORROBORAM PARA A CORRETA DETERMINAÇÃO DA SEQÜÊNCIA DE AMINOÁCIDOS

#### Perda neutra de amônia (NH<sub>3</sub>) e água (H<sub>2</sub>O)

As informações apresentadas a seguir são bastante importantes no tocante à confirmação de identificação de certos aminoácidos: íons fragmento  $-y$  e  $-b$  contendo os resíduos de aminoácidos **R**, **K**, **Q** e **N** podem apresentar perda neutra de amônia (-17 u). O íon inerente a essa perda neutra é não raramente mais intenso que os próprios íons  $-b$  ou  $-y$  correspondentes e, íons-fragmento  $-y$  e  $-b$  contendo os resíduos de aminoácidos **S**, **T** e **E** podem apresentar perda neutra de água (-18 u). No caso do ácido glutâmico, tal fato será mais notório caso esse aminoácido esteja na posição do N-terminal do fragmento. Tais informações corroboram para que a certeza inerente a uma determinação seja aumentada.

#### Intensidades dos picos no espectro

Nos peptídeos gerados a partir de digestão tríptica, os íons da série  $-y$  geralmente serão os de maior intensidade no espectro. Sempre que um peptídeo tríptico contiver **D** em sua seqüência (não importando a posição) e o número de cargas for igual ou menor ao

número de resíduos de **R**, haverá uma fragmentação preferencial na posição C-terminal do **D**, de modo que o pico inerente a tal fragmentação será o mais intenso do espectro. Quando a clivagem ocorrer em uma ligação peptídica de modo a posicionar um resíduo de **P** na posição N-terminal do peptídeo, a intensidade do íon  $-b$  (o qual não conterà a **P**) será bastante reduzida em relação ao íon  $-y$  correspondente (o qual terá a **P** na posição N-terminal). Tal evento também poderá ser observado, ainda que em menor extensão, com os resíduos de **G**, **H**, **K** e **R**.

Clivagens internas podem ocorrer nos resíduos **P** e **H**. Um fragmento de clivagem interna é o fragmento que parece ser um peptídeo reduzido apresentando **P** e/ou **H** em sua posição N-terminal. Por exemplo, o peptídeo **EFGLPGLQNK** pode apresentar os íons  $-b$  referentes aos fragmentos de seqüência: **PGLQNK**, **PGLQN**, **PGLQ**, etc. Esses são resultados de uma dupla clivagem e são normalmente designados como fragmentos internos.

### Espécies isobáricas

**L** e **I** são isóbaros e não podem ser diferenciados usando CID como mecanismo de dissociação. Quando essa diferença de massa for verificada no espectro, deve-se usar o símbolo **X** ou **Lxx** (**L/I** é outra notação comumente usada), de acordo com a nomenclatura de Hunt.<sup>25</sup>

**K** e **Q** são aminoácidos “quase” isobáricos, uma vez que possuem massa 128,09496 e 128,05858, respectivamente. A diferença de massa é 0,03638 u e pode ser usada para diferenciar **K** e **Q** se um espectrômetro for capaz de gerar resultados com alta exatidão de massa e resolução, tais como Q-TOF, Orbitrap e FT-ICR. Geralmente triplo quadrupolos e *ion traps* são incapazes de realizar tal feito.

Existem situações onde dois resíduos de aminoácidos irão apresentar massa molecular bastante próxima da de um terceiro aminoácido (Tabela 3), ou mesmo de outros dois aminoácidos específicos.

**Tabela 3.** Combinação de resíduos de aminoácidos que possuem a mesma massa de um único aminoácido

Aminoácido (s)	Massa (Da)
GG	114,04293
N	114,04293
AG	128,05858
Q	128,05858
K	128,09496
GV	156,08988
R	156,10111
AD	186,06406
EG	186,06406
W	186,07931
SV	186,10044

### EXEMPLOS PRÁTICOS DE INTERPRETAÇÃO DE ESPECTROS DE FRAGMENTAÇÃO DE PEPTÍDEOS

Uma vez apresentadas uma série de regras básicas bem como informações práticas que objetivam auxiliar na interpretação de espectros de fragmentação de peptídeos por MS em *tandem*, dois exemplos práticos serão apresentados a fim de melhor ilustrar os procedimentos descritos. Escolheram-se dois espectros inerentes ao mesmo peptídeo tríptico, obtidos por dois instrumentos diferentes. No primeiro exemplo, o espectro foi adquirido num instrumento do tipo TOF-TOF (Proteomics 4700 – Applied Biosystems), equipado com uma fonte de ionização do tipo MALDI (que gera peptídeos

majoritariamente monocarregados). No segundo, um *ion-trap* tridimensional (LCQ Deca XP Plus – Thermo) foi empregado. Nesse instrumento uma fonte de ionização do tipo *electrospray* foi utilizada. Assim, além de apresentar a interpretação dos espectros, será possível verificar as diferenças entre os espectros obtidos por esses dois diferentes instrumentos.

O espectro de fragmentação obtido para o peptídeo tríptico em estudo (íon precursor monocarregado na forma  $[M+H]^+$ ,  $m/z=1623,7$ ) usando o instrumento TOF-TOF ilustra bem o perfil dos espectros usualmente obtidos com esse tipo de instrumento. Em outras palavras, os espectros são pobres em relação ao número de picos principalmente nas regiões de altas massas, sendo bastante difícil a verificação de picos inerentes aos íons  $-a$ , assim como perdas neutras de  $H_2O$  e  $NH_3$ . Por outro lado, os espectros obtidos nesse tipo de instrumento apresentam as vantagens de serem obtidos com alta resolução (da ordem de 3000) e exatidão de massa (melhor que 40 ppm). Tais características permitem que ambigüidades em algumas situações sejam eliminadas. Como exemplo, pode-se citar a distinção entre os resíduos de aminoácidos **Q** e **K**, que possuem massas bastante próximas e não são distinguidos em equipamentos com baixa exatidão de massa, tais como os *ion-traps*. Por fim, é necessário afirmar que o espectro de fragmentação obtido no instrumento TOF-TOF em uso é submetido a uma etapa de pós-tratamento, onde sofre um processo de desconvolução, que exclui possíveis picos inerentes a espécies multicarregadas (se presentes), bem como elimina o padrão isotópico.

O mesmo espectro de fragmentação com um *zoom* na região de baixas massas, seguindo o procedimento apresentado, identificou os íons imônio (Tabela 1) correspondentes aos seguintes resíduos de aminoácidos: **S** (60), **P** (70), **IL** (86), **N** (112), **Y** (136) e **W** (159). Dessa forma, é possível ter uma idéia acerca da composição de aminoácidos do peptídeo. Além disso, o pico com  $m/z$  101, também presente, pode indicar a presença do resíduo de aminoácido **Q** ou **K**.

Como próximo passo, sabendo que o peptídeo em estudo é oriundo de digestão tríptica, espera-se que o resíduo de aminoácido C-terminal seja **R** ou **K**. Nesse caso, o aminoácido C-terminal é facilmente identificado como sendo a **R**, visto que além do íon imônio referente a **R** ter sido verificado, o pico com  $m/z$  175 ( $-y_1$ ) confirma tal fato. Com essa informação, automaticamente pode se retornar ao parágrafo anterior e inferir que o íon com  $m/z$  101 corresponde a **Q** e não a **K**, a menos que a tripsina tenha “falhado” em um ponto de clivagem, o que *a priori* pode ser descartado. De posse da informação inerente ao íon  $-y_1$ , pode-se calcular a relação  $m/z$  do íon  $-b_{n-1}$  correspondente, que nesse caso se trata do íon 1449,7, o qual não pode ser visualizado no espectro (Figura 4SD - Material Suplementar).

É também possível verificar o íon  $-b_2$ . Tal íon, neste caso o  $m/z$  155, é facilmente identificado devido à presença do íon **a** correspondente ( $m/z$  127). Fazendo uso da Tabela 2, pode-se concluir que o conjunto de dois resíduos de aminoácidos que apresentam  $m/z$  igual a 155 é formado por **PG**, sendo obviamente a ordem desconhecida. Nesse aspecto, o espectro mostra-se bastante simples uma vez que somente uma combinação de resíduos possui a relação  $m/z$  em questão. No entanto, em outros exemplos onde mais de um conjunto de aminoácidos possui o valor de  $m/z$  correspondente ao íon  $-b_2$ , todas as combinações devem ser levadas em consideração. Obviamente, dados inerentes à composição de aminoácidos do peptídeo (íons imônio) podem ajudar a indicar a seqüência que possui a maior probabilidade de ser a verdadeira. Isso significa que se duas seqüências apresentam a mesma massa relativa ao íon  $-b_2$  e, no entanto, em uma delas os íons imônio correspondentes aos aminoácidos constituintes não foram detectados, é provável que essa não seja a seqüência correta.

Conhecendo-se o íon  $-b_2$ , é possível calcular o íon  $-y_{n-2}$  correspondente, que nesse caso se trata do íon de  $m/z$  1569, o qual também não foi observado no espectro. Além disso, pode-se usar tal informação para tentar determinar a seqüência dos resíduos de aminoácidos da região N-terminal. Com as informações obtidas até esse ponto sabe-se que os dois primeiros resíduos da região N-terminal são **P** e **G**. Logo, para determinar a ordem correta basta fazer a suposição que a **P** seja o resíduo N-terminal. Se isso for verdade, um íon com  $m/z$  aproximadamente igual a 98,03 (155,047 – 57,021) deveria ser observado, o que não ocorre. Portanto, não há evidência que a **P** seja o aminoácido N-terminal. A outra possibilidade é que **G** esteja na posição N-terminal e nesse caso o íon com  $m/z$  em torno de 57,99 deveria ser verificado, o que também não ocorre. Assim, apesar de saber que os dois primeiros resíduos de aminoácidos da região N-terminal do peptídeo são a **P** e a **G**, não é possível determiná-los pela interpretação desse espectro de massas.

Uma vez determinados os íons  $-y_1$  e  $-b_2$ , foi possível estender as seqüências de aminoácidos. A série  $-b$  foi estendida desde o íon  $-b_2$  ( $m/z$  155,05) até o íon  $-b_7$  ( $m/z$  782,42). Por outro lado, a série  $-y$  pôde ser estendida a partir do íon  $-y_1$  ( $m/z$  175,07) até o íon  $-y_7$  /  $-b_6$  ( $m/z$  1028,70). Assim sendo, o completo seqüenciamento do peptídeo foi obtido, fazendo uso da série  $-y$  e/ou da série  $-b$  (Figuras 4SC e 4SD). Neste exemplo, íons complementares (os pares  $-y_7$  /  $-b_6$  e  $-y_6$  /  $-b_7$ ) puderam ser identificados, aumentando a confiabilidade da seqüência proposta (veja Figura 4S).

A seqüência de aminoácidos determinada para o peptídeo em estudo, bem como os íons que foram detectados no espectro, permitiu prever a seqüência de aminoácidos do peptídeo, com exceção dos dois primeiros aminoácidos na região N-terminal, que permanece desconhecida (veja Figura 5S).

O espectro de fragmentação para o mesmo peptídeo em questão foi obtido usando-se o instrumento do tipo *ion-trap* tridimensional. Neste caso, o modo de ionização foi o ESI e o íon precursor era duplamente carregado ( $[M+2H]^{2+}$ ,  $m/z = 812,5$ ). Ao contrário do espectro obtido no TOF-TOF, esse espectro apresenta muitos picos, inclusive picos inerentes às perdas neutras de  $H_2O$  e  $NH_3$ , o que facilita o seqüenciamento do peptídeo, porém ao mesmo tempo torna maior a chance de que erros sejam cometidos devido a equívocos na assinalação das séries. Tal possibilidade é ainda aumentada devido ao fato do espectro oriundo desse tipo de instrumento não ser desconvoluído, o que significa que íons relativos a espécies multicarregadas podem estar presentes, dificultando a interpretação.

Deve-se lembrar também que os espectros obtidos em *ion-traps* possuem baixa resolução e exatidão de massas, o que torna impossível, por exemplo, fazer a correta distinção entre os resíduos **Q** ou **K**. A limitação inerente à regra do  $1/3^{24}$  é verificada e por esse motivo íons com relação  $m/z$  inferiores a 270 não são sequer detectados. Nesse caso, como não se tem acesso à região de baixas massas, fica impossível verificar os íons imônio presentes, bem como os íons diagnóstico  $-y_1$ . Sabendo-se de antemão que o resíduo de aminoácido C-terminal se trata de **R** ou **K**, deve-se supor que se trata da **K** (por exemplo) e calcular o íon  $-b_{n-1}$  correspondente, que nesse exemplo teria  $m/z$  igual a 1477,5. Avaliando a Figura 6SC pode-se concluir que esse íon não está presente e, portanto, a suposição de **K** corresponde ao resíduo de aminoácido C-terminal não se mostrou consistente. Como próxima tentativa deve-se supor que **R** é o resíduo C-terminal, situação na qual o íon  $-b_{n-1}$  deveria ter  $m/z$  igual a 1449,4. A análise da Figura 6SC mostra que esse íon está presente e, portanto, o aminoácido C-terminal trata-se da **R**.

Uma análise criteriosa mostra que o íon de  $m/z$  268,11 se trata de um fragmento do tipo  $-b$ . Tal fato pôde ser concluído uma vez que o íon  $-a$  correspondente é facilmente identificado ( $m/z$  240,89).

Além dessa evidência, o cálculo do íon  $-y$  correspondente fornece o íon com  $m/z$  1357,45, também presente no espectro. Assim, o íon de  $m/z$  268,11 trata-se do primeiro íon da série  $-b$  detectado nesse espectro. Dessa maneira é possível estender as séries  $-y$  e  $-b$  usando as massas dos resíduos dos aminoácidos. A série  $-b$  é estendida desde o íon  $-b_{n-1}$  ( $m/z$  1449,5) até o último íon  $b$  detectado ( $m/z$  268,11). Por outro lado, a série  $-y$  pode ser estendida a partir do íon  $-y_7$ , mesmo apesar do fato desse íon não ter sido detectado no espectro. As séries  $-b$  e  $-y$  estendidas bem como alguns íons  $-a$  verificados ( $-a_3$  e  $-a_7$ ) mostram, além de inúmeros íons inerentes, as perdas neutras de  $H_2O$  e  $NH_3$ . Conforme descrito anteriormente, uma intensa perda neutra de  $H_2O$  pode ser verificada em íons fragmento que contêm os resíduos de aminoácidos **S**, neste caso são mostrados como exemplo as perdas neutras relativas aos íons  $-b_6$  e  $-y_8$ . Além disso, diversas perdas neutras de  $NH_3$  são observadas para os fragmentos que contêm os aminoácidos **N**, **Q** e **R**. Como exemplo, são mostradas as perdas neutras dos íons  $-b_8$ ,  $-b_{10}$ ,  $-y_{10}$  entre outras (veja Figura 6S).

A seqüência de aminoácidos obtida para o peptídeo em estudo, bem como todos os íons detectados no espectro, permitiu prever a seqüência de aminoácidos do peptídeo de maneira inequívoca a menos da ordem dos dois primeiros aminoácidos na região N-terminal, a qual continuou desconhecida também nesse equipamento (veja Figura 7S).

Conforme esperado, a interpretação dos espectros de fragmentação do peptídeo em estudo, obtidos pelos dois diferentes instrumentos, gerou resultados concordantes. Assim sendo, a seqüência de aminoácidos obtida (**GPXQXSWNYNYXR** ou **PGXQXSWNYNYXR**) pode agora ser comparada com a seqüência real do peptídeo, que é **GPIQLSWNYNYLR**. Deste modo, pode-se concluir que a interpretação dos espectros possibilitou a determinação fidedigna dos aminoácidos constituintes desse peptídeo, a menos da seqüência dos dois primeiros resíduos de aminoácidos da região N-terminal.

## CONSIDERAÇÕES FINAIS

O presente artigo apresenta um guia prático para a interpretação de espectros de fragmentação de peptídeos obtidos usando espectrometria de massas em *tandem*. O conjunto de regras e informações relatadas foi compilada a partir de uma profunda revisão bibliográfica sobre o assunto, bem como usando o conhecimento prático adquirido pelos autores. A fim de melhor enfatizar a aplicabilidade do guia proposto, dois exemplos foram apresentados. Espectros de fragmentação para um mesmo peptídeo foram obtidos através de dois espectrômetros de massas diferentes, sendo eles um TOF-TOF e um *ion-trap* (tridimensional). Além de exemplificar com exemplos práticos e reais o seqüenciamento de

peptídeos, foi possível verificar as diferenças espectrais inerentes aos instrumentos avaliados. Dessa forma, espera-se que esse artigo sirva como uma fonte de referência para pesquisadores que fazem uso de espectrometria de massas para estudar/identificar proteínas, peptídeos biologicamente ativos, etc.

## MATERIAL SUPLEMENTAR

Está disponível gratuitamente em <http://quimicanova.s bq.org.br>, em forma de arquivo PDF, apresentando as Figuras 1S a 7S.

## REFERÊNCIAS

1. Tyers, M.; Mann, M.; *Nature* **2003**, *422*, 193.
2. Aebersold, R.; Mann, M.; *Nature* **2003**, *422*, 198.
3. Sadygov, R. G.; Cociorva, D.; Yates, J. R.; *Nature Methods* **2004**, *1*, 195.
4. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M.; *Science* **1989**, *246*, 64.
5. Karas, M.; Hillenkamp, F.; *Anal. Chem.* **1988**, *60*, 2299.
6. Ramstrom, M.; Hagman, C.; Mitchell, J. K.; Derrick, P. J.; Hakansson, P.; Bergquist, J.; *J. Proteome Res.* **2005**, *4*, 410.
7. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W.; *J. Am. Chem. Soc.* **1999**, *121*, 806.
8. Bogdanov, B.; Smith, R. D.; *Mass Spectrom. Rev.* **2005**, *24*, 168.
9. Nemeth-Cawley, J. F.; Tangarone, B. S.; Rouse, J. C.; *J. Proteome Res.* **2003**, *2*, 495.
10. Eng, J. K.; McCormack, A. L.; Yates, J. R.; *J. Am. Soc. Mass. Spectrom.* **1994**, *5*, 976.
11. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S.; *Electrophoresis* **1999**, *20*, 3551.
12. Chamrad, D. C.; Korting, G.; Stuhler, K.; Meyer, H. E.; Klose, J.; Bluggel, M.; *Proteomics* **2004**, *4*, 619.
13. Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P.; *Nature Methods* **2005**, *2*, 667.
14. Washburn, M. P.; Wolters, D.; Yates, J. R.; *Nat. Biotechnol.* **2001**, *19*, 242.
15. Steen, H.; Mann, M.; *Nature Reviews* **2004**, *5*, 699.
16. Tabb, D. L.; Smith, L. L.; Brei, L. A.; Vysocki, V. H.; Lin, D.; Yates, J. R.; *Anal. Chem.* **2003**, *75*, 1155.
17. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F.; *PNAS* **2004**, *101*, 9528.
18. Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H.; *J. Am. Chem. Soc.* **1996**, *118*, 8365.
19. Mann, M.; Meng, C. K.; Fenn, J. B.; *Anal. Chem.* **1989**, *61*, 1702.
20. Roepstorff, P.; Fohlman, J.; *J. Biomed. Mass Spectrom.* **1984**, *11*, 601.
21. O'Hair, R. A. J.; *J. Mass Spectrom.* **2000**, *35*, 1377.
22. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Brei, L. A.; *J. Mass Spectrom.* **2000**, *35*, 1399.
23. Kinter, K.; Sherman, N. E.; *Protein Sequencing and Identification Using Tandem Mass Spectrometry*, Wiley-Interscience: New York, 2000.
24. March, R. E.; *J. Mass Spectrom.* **1997**, *32*, 351.
25. Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R.; *PNAS* **1986**, *17*, 6233.
26. Falick, A. M.; Hines, W. M.; Medzihradzky, K. F.; Baldwin, M. A.; Gibson, B. W.; *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 882.