

PROTEOMA

Ilustrações cedidas pelas autoras

Avanços Recentes em Técnicas de Eletroforese Bidimensional e Espectrometria de Massa

Proteoma indica as PROTEÍNAS expressas em um genOMA ou tecido. Enquanto o genoma representa a soma de todos os genes de um indivíduo, o proteoma não é uma característica fixa de um organismo. O proteoma altera com o estado de desenvolvimento, do tecido ou mesmo sob as condições nas quais o indivíduo se encontra. Portanto, há muito mais proteínas no proteoma do que genes no genoma, especialmente para eucaritos. Isto porque há várias maneiras do gene expresso, o RNA total, sofrer redução (*splicing*) para construir o RNA “maduro”. Este então vai servir de molde para a tradução de uma proteína, a qual sofre modificação pós-tradução.

Investigar diretamente os produtos dos genes é uma forma de estudar doenças e qualquer problema biológico complexo. Por exemplo, para entender molecularmente como uma célula funciona em um indivíduo doente e em um sadio é preciso ter conhecimento das proteínas e de outros componentes celulares que estão presentes, como eles interagem e o resultado de suas interações.

Portanto, análise ao nível de proteína é necessária, pois o estudo dos genes através do sequenciamento dos genes, ou seja, o estudo genômico, não pode adequadamente prever a estrutura dinâmica das proteínas, uma vez que é ao nível das proteínas que muitos dos processos de uma célula ocorrem, onde processos de doenças inicialmente acontecem e aonde muitas das drogas medicinais atuam. Sendo assim, proteômica é o método direto

para identificar, quantificar e estudar as modificações pós-traducionais das proteínas em uma célula, tecido ou mesmo organismo.

O termo proteômica foi introduzido em 1995 para descrever todas as proteínas que são expressas em um genoma (Anderson et al., 1996; Wilkins et al., 1996). Definir todos os aspectos da proteômica é difícil, pois atualmente este termo é mais um conceito do que uma ciência bem definida.

Proteômica pode ser vista como uma metodologia de seleção da biologia molecular, a qual tem como objetivo documentar a distribuição geral de proteínas da célula, identificar e caracterizar proteínas individuais de interesse e principalmente elucidar as suas associações e funções. Sendo assim, proteômica fundamenta-se em princípios bioquímicos, biofísicos e de bioinformática para quantificar e identificar as proteínas expressas, pois elas se alteram conforme o desenvolvimento de um organismo assim como em resposta aos fatores do ambiente (Anderson & Anderson, 1996; Celis et al., 1996; Wilkins et al., 1996; Wilkins et al., 1997).

Há uma forte e sinérgica correlação entre os estudos proteômicos e genômicos uma vez que ambas áreas de estudo investigam a organização celular ao nível complementar, proteínas e genes, e cada área fornece informação que aumenta a eficiência da outra. Por exemplo, proteômica conecta as descobertas da genômica com o desenvolvimento das drogas que as companhias farmacêuticas in-

Luciana Di Ciero
Doutora, Pesquisadora, Esalq-USP
ldctleme@terra.com.br

Cláudia de Mattos Bellato, Ph.D
Pesquisadora, CENA-USP
bellato@cena.usp.br

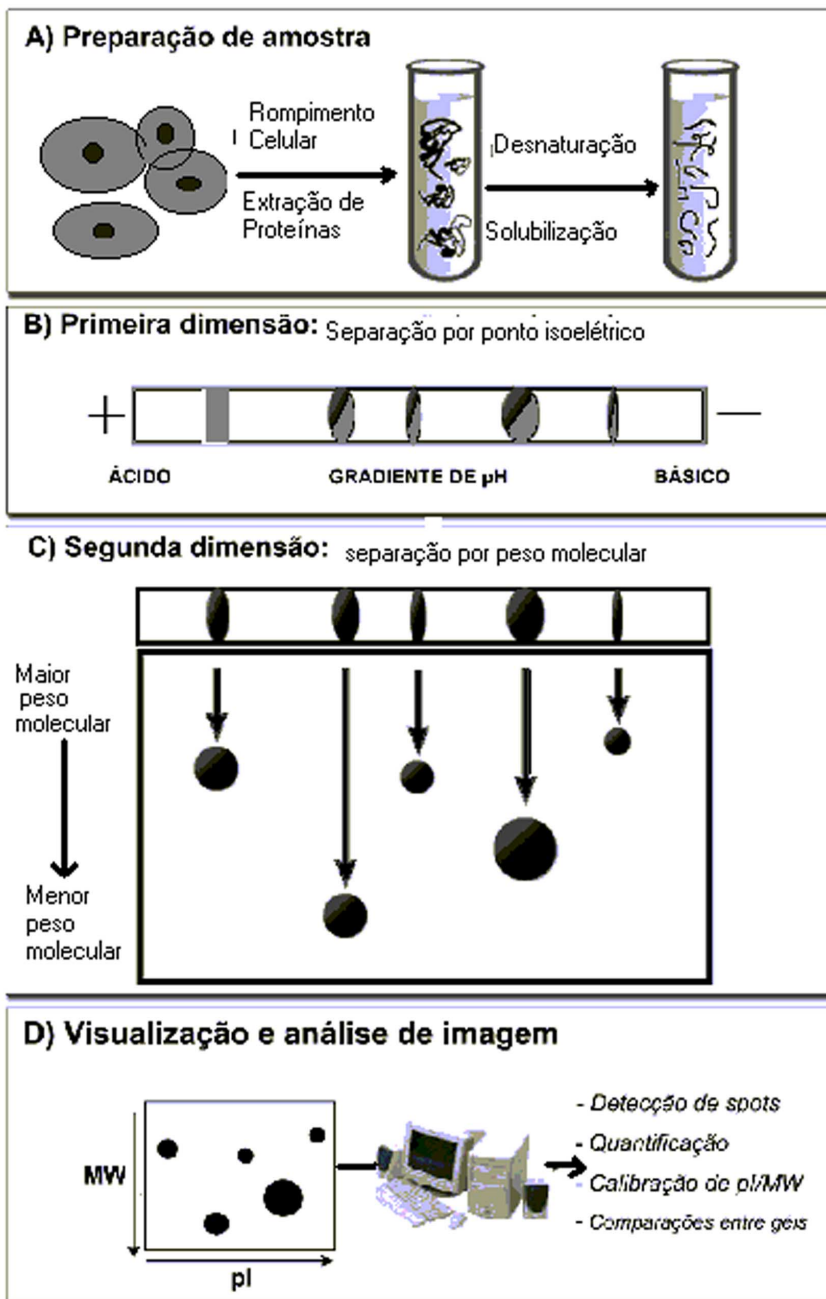


Figura 1. Esquema de procedimento de eletroforese bidimensional (A) Preparação de amostra: rompimento celular seguido de extração e solubilização de proteínas das células (ou tecidos) (B) Separação na primeira dimensão por IEF, onde as proteínas são separadas pelo seu ponto isoelétrico. (C) Separação na segunda dimensão SDS-PAGE, que separa proteínas pela massa molecular aparente. (D) coloração das proteínas separadas, resultando num mapa de pontos (spots). Digitalização da imagem e análise densitométrica por programas de computador

investigam. Sendo assim, proteínas de uma forma geral controlam a vida e a saúde, ou seja, é preciso entender as proteínas e como elas operam celularmente para assim entender como regular os mecanismos da doença para

um posterior tratamento.

A Proteômica pode ser aplicada em diversas áreas de interesse como por exemplo na investigação de marcadores moleculares em determinadas doenças indicando a resposta da célula ou

tecido a estresses externos. Através da proteômica pode-se fazer uma comparação do perfil protéico de uma célula cancerosa com o de uma célula sadia, ou de uma célula cancerosa cujo portador está sob tratamento médico. Várias pesquisas já foram ou estão sendo conduzidas no mundo na área da saúde como o proteoma do fígado, rim, e para diferentes organismos, como *Mycobacterium tuberculosis* (agente causal da tuberculose), *Plasmodium falciparum* (agente causal da malária), *Helicobacter pylori* (agente causal da úlcera e gastrite). Na área agrônômica, há estudos proteômicos de milho, arroz, trigo, cana-de-açúcar, eucalipto entre outras, e de organismos que causam impacto econômico, como aquele que fixa nitrogênio atmosférico para as leguminosas, *Rhizobium* spp., ou que causam doenças em feijoeiro e citros como *Xanthomonas* spp. Aqui no Brasil, pesquisas proteômicas estão sendo realizadas com sucesso. No Estado de São Paulo, graças ao apoio da Fundação de Amparo a Pesquisa (FAPESP), estudos proteômicos estão sendo conduzidos para entender o processo da interação entre a bactéria patogênica, *Xylella fastidiosa*, de citros variedade laranja doce.

Técnica aplicada ao Estudo Proteômica

A eletroforese de duas dimensões em gel de poliacrilamida (2DE) em conjunto com a espectrometria de massa são as duas tecnologias mais usadas atualmente para análise de proteoma. A primeira é aplicada para separação, detecção e quantificação das proteínas e a espectrometria de massa para identificação das proteínas graças aos grandes avanços da bioinformática.

Eletroforese Bidimensional (2DE)

A técnica de eletroforese bidimensional foi introduzida na década de 70. Entretanto foi nos últimos 10 anos que ocorreram os grandes avanços nos métodos que possibilitaram identificar proteínas separadas através de 2DE de forma sensível, rápida e conclusiva.

Desde a introdução da técnica de

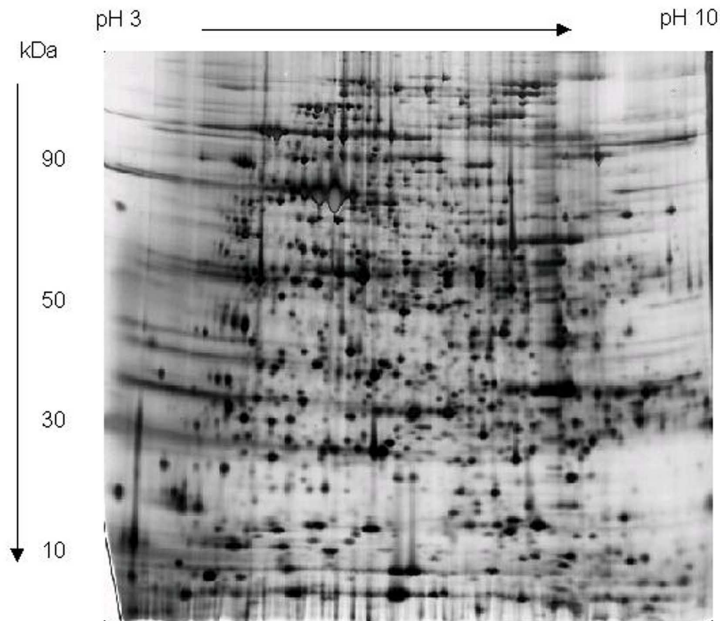


Figura 2. Perfil da proteína total (150 ug) de *Xylella fastidiosa*, bactéria patogênica de citros var. laranja doce. IEF ocorreu entre pH3-10 e a 2DE em gel gradiente (9-18%) de poliacrilamida. Gel corado com nitrato de prata. Projeto financiado pela FAPESP



Figura 3. Esquema de “Peptide mass fingerprinting” (PMF) – Os “spots” são recortados do gel 2DE e submetidos a digestão enzimática, geralmente usando-se tripsina. Os peptídeos resultantes são incorporados a uma matriz e identificados por espectrometria de massa (MALDI-TOF). A massa dos peptídeos são comparadas a um banco de dados para identificar seqüências de amonácidos, cujas massas previstas batem com a massa dos peptídeos obtidos

2DE, diversas modificações foram feitas, principalmente na primeira dimensão. Utilizava-se os anfólitos para estabelecer o gradiente de pH para a separação das proteínas pelo ponto isoelétrico. Entretanto, o uso de anfólitos com esta finalidade tem diversos problemas incluindo a incapacidade de correr grandes quantidades de proteínas necessárias para a micro seqüência, pouca estabilidade do gradiente de pH durante a eletroforese e freqüente falta de reprodutibilidade dos géis entre os laboratórios. O uso dos gradientes de pH imobilizados (IPG) para a separação por cargas na 2DE solucionou muitos dos problemas. Encontra-se no mercado fitas de géis IPG desidratadas (Amersham Biosciences) feitas de uma matriz de poliacrilamida, a qual é modificada covalentemente com grupos ácidos e básicos que formam o gradiente de pH imobilizado em toda a extensão do gel. Há diversas faixas de pH, desde ampla (3-10), como faixas estreitas, como por exemplo de 4,5-5,5 ou 7,0-11,0. Esta flexibilidade na escolha das faixas de pH a serem utilizadas é de grande utilidade para separar eficientemente o maior número de proteínas possível.

A 2DE combina duas técnicas de separação: a primeira separação (primeira dimensão) é a focalização isoelétrica (IEF), e a segunda dimensão (2DE) é a eletroforese em gel de poliacrilamida, chamada de SDS-PAGE. No IEF as proteínas são separadas, através de uma alta tensão, em um gradiente de pH imobilizado em um gel de acrilamida até alcançarem a posição estacionária onde a carga total é zero. O pH no qual a proteína tem carga líquida zero é chamado de ponto isoelétrico (pI). Na 2DE, as proteínas são separadas pelo peso molecular, através de uma tensão, também em gel de poliacrilamida contendo ou não o detergente sulfato dodecil de sódio (SDS-PAGE). A mobilidade eletroforética se dá de acordo com o tamanho da proteína sendo restringida pelo tamanho do poro da matriz de poliacrilamida, de uma forma que as proteínas migram em uma velocidade proporcional ao seu tamanho. A matriz de poliacrilamida usada pode ser homo-

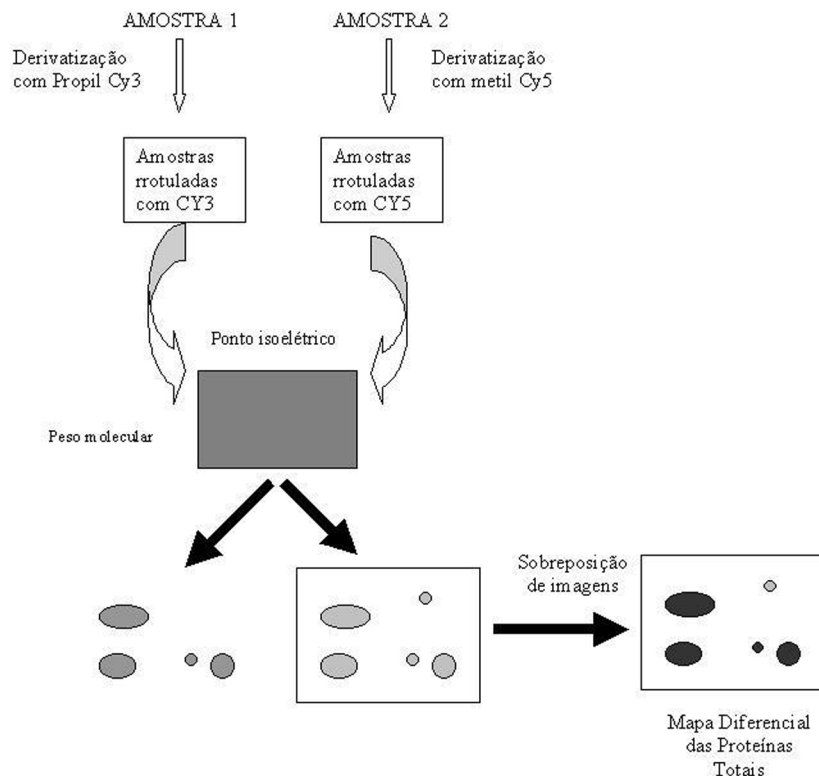


Figura 4. Ilustração Esquemática de DIGE (adaptado de Patton, 2002) – Duas amostras diferentes são derivatizadas em dois fluoróforos diferentes, combinadas e submetidas a 2DE. AS proteínas são detectadas em um scanner duplo a laser com filtros de excitação diferentes para gerar duas imagens separadas. As imagens são, então sobrepostas com a ajuda de um programa de computador, os sinais são normalizados e os spots são quantificados. As diferenças na expressão das proteínas são identificadas pela avaliação de uma imagem pseudo-colorida e os dados são comparados

gênea ou em gradiente. Já os parâmetros elétricos usados na corrida dependem do tipo e da quantidade de amostra aplicada, influenciando na qualidade da focalização. Estes parâmetros devem ser cuidadosamente monitorados e controlados. Em aparelhos do tipo IPGphor (Amersham Biosciences), geralmente utiliza-se uma corrente de 50 uA por strip e uma voltagem que vai de 100 a 8000 (80kVh-150kVh).

Para aplicar a técnica de 2DE (Figura 1), as amostras biológicas são, de uma maneira geral, tratadas da seguinte forma. Primeiramente, as células são rompidas através de métodos químicos (detergentes, solventes, etc) ou físicos (homogeneizadores, moínhos, “french press”). Seguido da ruptura celular, as proteínas são solubilizadas e

desnaturadas em uma solução contendo detergentes não iônicos ou zwitteriônicos, caotropes (uréia, tiouréia), e pequenas quantidades de agentes redutores, como o ditioneitol (DTT). Uma solução comum é a composta de uréia (8M), CHAPS (4%), DTT (70mM), anfólito (0,5 a 2%) e azul de bromofenol (0,005%). A concentração do anfólito deve ser determinada para cada amostra empiricamente.

A solubilização das proteínas é uma etapa fundamental para o sucesso da eletroforese bidimensional. Estas soluções devem ser otimizadas baseadas nas amostras utilizadas no experimento. Sabe-se que proteínas de membrana são de difícil visualização em gel devido, principalmente a dificuldade de solubilizá-las, já que o uso do SDS é

limitante para a 2DE. Recentemente foram disponibilizados diversos detergentes zwitteriônicos que melhoraram muito o desempenho das soluções de solubilização para proteínas hidrofóbicas (Chevallet et al., 1998). Estes detergentes são sulfobetáinas com cauda de cadeia geralmente com 8 a 16 carbonos. A Calbiochem comercializa estes detergentes com os nomes comerciais de SB 3-10, SB 3-12, ASB 14, ASB 16. Fizemos um estudo comparativo da eficiência de 4 detergentes na solubilização de proteínas de membrana de *Xyella fastidiosa* (projeto Proteoma de Membrana de *Xyella fastidiosa* financiado pela FAPESP), e mostramos que o ASB 14 é mais eficiente quando comparado com o SB 3-10, seguido do CHAPS e Triton X-100 (Di Ciero et al., resultados recentes).

Após serem submetidas a 2DE, as proteínas são visualizadas através de métodos de detecção específicos para cada objetivo, como por exemplo detecção de proteínas totais, modificações pós-traducionais, etc. A maioria desses métodos se baseia em coloração de proteínas com corantes ou reagentes fluorescentes, ou detecção de proteínas marcadas radioativamente por fluorografia ou autoradiografia.

Dentre os diversos métodos para coloração de proteínas, os mais usados são coloração por Azul de Coomassie (Coomassie Blue) e coloração por prata, devido ao custo, facilidade de uso e compatibilidade com métodos de caracterização microquímica, tais como sequência automática de aminoácidos e espectrometria de massa. Coomassie Blue detecta proteínas com concentração superior a 100 ng, enquanto a prata é mais sensível que o Coomassie, podendo detectar proteínas em concentrações de até 1 ng.

Os métodos de marcação radioativa baseia-se na incorporação de ^3H , ^{14}C , ^{35}S , ^{32}P , ^{33}P ou ^{125}I às proteínas. Após a eletroforese a detecção do sinal pode ser feita usando-se filme para autoradiografia ou fluorografia direta. Esses métodos são aplicados em proteoma juntamente com a coloração por Azul de Coomassie ou nitrato de prata para análise simultânea do níveis

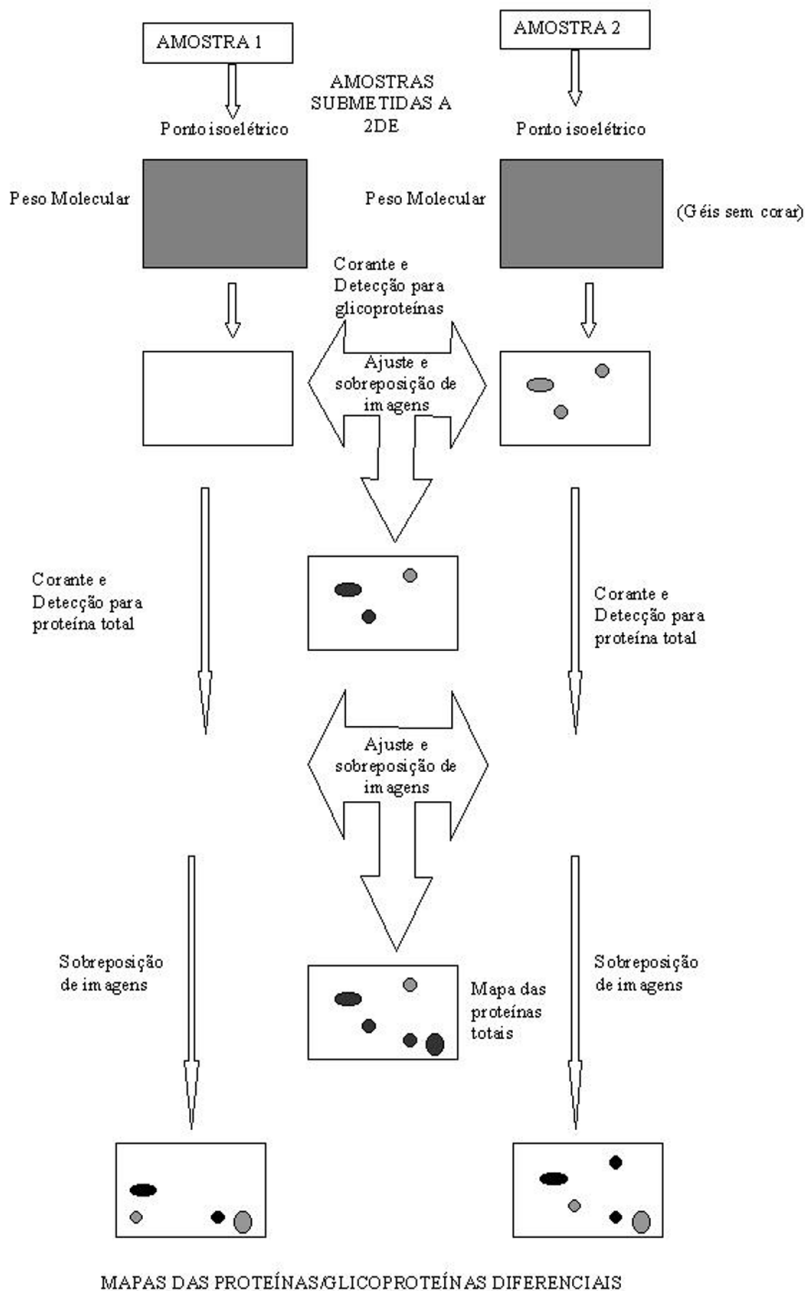


Figura 5. Ilustração Esquemática de Proteoma Múltiplo (MP) (Adaptado de Patton, 2002) – Duas amostras diferentes são submetidas a 2DE. Os géis são corados por um atributo funcional particular, tal como glicosilação, e digitalizados. Os géis são então corados para proteínas totais usando o corante SYPRO Ruby e digitalizados novamente. As duas imagens são então sobrepostas com a ajuda de um programa de computador, e os spots são quantificados. As diferenças na glicosilação e expressão de proteínas totais são identificadas pela análise de imagens pseudo-coloridas e comparação de dados. O registro do perfil de proteína total com padrões de glicosilação são feitos pela sobreposição das imagens

de expressão da proteína total juntamente com a taxa de síntese de proteína, bem como fornecer regiões marcadas para a localização de proteínas

de baixa abundância em 2DE. Embora os métodos de detecção por marcação radioativa sejam bastante sensíveis e quantitativos por uma maior faixa de

abundância, eles tendem a ser mais caros e perigosos de manipulação, e limitados a amostras que podem ser marcados radioativamente (geralmente por marcação metabólica).

Métodos de coloração reversa têm sido usados com menor frequência para a visualização de proteínas em 2DE, os quais empregam cloreto de potássio, cloretos de cobre, cloreto de zinco, acetato de cobalto cloreto de níquel ou cloreto de zinco imidazol. Dentre elas, as mais usadas em proteoma são cloreto de potássio, cloreto de cobre e cloreto de zinco. O cloreto de cobre é um pouco mais sensível que o Azul de Coomassie, entretanto o cloreto de potássio é menos sensível. O método que emprega o cloreto de zinco imidazol é o mais sensível entre eles e pode ser avaliados por densitometria, embora a faixa linear dinâmica está restrita a 10-100 ng. Os géis corados com zinco-imidazol podem ser eletrotransferidos (“electroblotted”) ou eletroeluídos (“electroeluted”) pela adição de um agente quelante ao tampão de transferência. Outra vantagem deste método é que ele é compatível com métodos de microsequência baseados em Edman e a digestão trípica *ingel* para a identificação dos peptídeos por espectrometria de massa MALDI-TOF (“matrix-assisted laser desorption ionization time-of-flight”).

A detecção de proteínas por fluorescência em eletroforese em gel é muito usada em laboratórios que pesquisam proteoma em larga escala. A faixa de detecção linear dinâmica normalmente é superior aos métodos de coloração. Os corantes disponíveis no mercado são o Vermelho do Nilo (Nile Red), SYPRO Vermelho, SYPRO Laranja e SYPRO Tangerina. Os corantes SYPRO são fáceis de serem aplicados, através de somente uma etapa de coloração de 30 a 60 minutos sem a necessidade de descoloração. Estes corantes detectam pequenas quantidades de proteína 4-8 ng e há metodologia para quantificação por análise de imagem.

Para análise de modificações pós-traducionais são usados métodos de detecção específicos para glicoproteínas (autoradiografias com a incor-

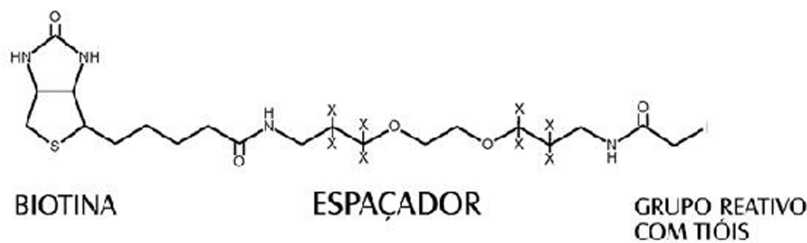


Figura 6. Estrutura do reagente ICAT, composto de três elementos: um grupamento reativo à tióis, que se liga à resíduos de cisteínas reduzidas; um espaçador que pode conter deutérios (reagente pesado, com os deutérios nas posições indicadas por X) ou não (reagente leve, com hidrogênios nas posições indicadas por X); e um grupamento de biotina utilizado para o isolamento seletivo dos peptídios contendo cisteína (marcados com ICAT)

poração de ^3H e ^{14}C ; fluorescência, método de Schiff, Azul de Alcian, etc.), fosforilação (autoradiografias com a incorporação de ^{32}P ou ^{33}P na cultura de células ou após dentro de frações subcelulares por proteínas cinases, detecção *in gel* por hidrólise alcalina de ésters de fosfato de serina ou treonina precipitando-se o fosfato liberado com cálcio formando um complexo de fosfomolibdato e então visualizando-se com um corante, tal como Verde de

Metil), proteólise (ensaios zimográficos, corantes fluorescentes BODIPY), nitrosilação (método de Biotina), metilação (fluorografia), ribolização (anticorpos específicos anti-ADP ribose).

Após a coloração do gel, observa-se um perfil bidimensional de pontos (“spots”), sendo que em cada ponto há múltiplas cópias de uma proteína. A imagem deste perfil é posteriormente documentada através de programas de computador apropriados (Figura

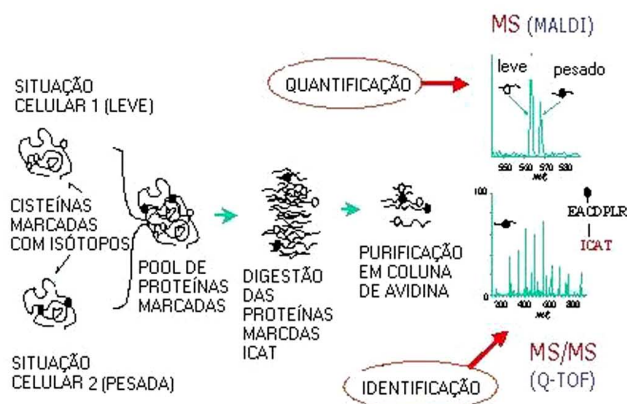


Figura 7. Esquema do procedimento de ICAT - Os resíduos de cisteína reduzida das proteínas em amostras a serem comparadas são rotulados separadamente. Em uma das amostras é usada a forma isotopicamente pesada e na outra amostra a forma leve do reagente ICAT. As duas amostras são combinadas e digeridas com tripsina. Os peptídios marcados, que contém cisteína são isolados por cromatografia de afinidade e, posteriormente analisados por espectrometria de massa para a determinação da identidade da proteína que deu origem ao peptídio e a abundância relativa de cada proteína nas amostras sendo comparadas

2), como o Melanie desenvolvido pela Genebio (Genebra, Suíça), Image Master (Amersham BioSciences) e outros.

Estes programas apresentam várias funções, entre elas a de permitir uma análise da abundância da proteína em cada *spot* através do volume, intensidade e área. Diversos géis representando diferentes proteomas podem ser comparados com o objetivo de detectar proteínas diferencialmente expressas.

Apesar da técnica de 2DE ter o potencial de separar milhares de proteínas e ser a técnica mais utilizada para análise de proteomas, ela apresenta algumas limitações para proteínas muito ácidas (pH menor que 3,5) or básicas (pH maior que 9) proteínas de baixa abundância e raras proteínas hidrofóbicas, geralmente presentes nas membranes celulares. Porém, esta técnica está sendo cada vez mais aperfeiçoada e logo limitações como estas deixarão de existir. No geral, a 2DE é um método de separação eficiente, porque todas as proteínas numa amostra são separadas simultaneamente, fornecendo informações úteis sobre pI, massa molecular, expressão e abundância relativa e modificações pós-traducionais pela alteração da mobilidade eletroforética.

Identificação das Proteínas

Há várias técnicas para identificar uma proteína, como por exemplo, o sequenciamento da proteína através da degradação por Edman, sequenciamento químico, processo imunológico, espectrometria de massa entre outros. Uma das técnicas mais utilizadas para a identificação das proteínas é a análise do “*fingerprinting*” da massa precisa de um número de peptídios derivados da mesma proteína (Figura 3). Para tal, as proteínas de interesse são recortadas do gel, são fragmentadas (obtenção dos peptídeos), geralmente por digestão triptica, e os fragmentos são analisadas no espectrômetro de massa MALDI-TOF auxiliados por programas de informática desenvolvidos por vários grupos (Pappin, 1997; Mann et al., 1993; Henzel, 1993; Yates et al.,

1993; James et al., 1993).

Uma vez conhecida a sequência desta proteína, a sua identificação é geralmente feita por correlação de dados depositados em bancos acessíveis via internet, como o da Swiss-Prot (Suíça).

Técnicas e Opções para Análise de Proteoma Diferencial

O proteoma diferencial compara diferentes perfis de proteínas entre situações de interesse. O objetivo central do proteoma diferencial é melhorar o conteúdo de informações do estudo de proteoma através de análises múltiplas. As três principais técnicas usadas atualmente para análise de proteoma diferencial são eletroforese diferencial em gel (DIGE), "Multiplex Proteome" (MP) e rotulação de proteínas com reagentes ICAT ("isotope-coded affinity tagging").

A técnica DIGE (Figura 4) usa ésters de succinil de corantes de cianina, CY2, CY3 e CY5, que podem ser empregados para rotulação fluorescente para três populações complexas de proteínas antes de misturá-las e resolvê-las simultaneamente na mesma 2DE. Recentemente, foi desenvolvido outro procedimento similar usando corante Alexa Fluor.

O "Multiplex Proteome" (MP) baseia-se na determinação paralela dos níveis de expressão de proteínas bem como em certos atributos funcionais, tais como níveis de glicosilação, capacidade de ligação com drogas ou metabolização de drogas (Figura 5). Esta tecnologia utiliza o mesmo fluoróforo para medir proteínas através de todos os géis no banco de dados, e emprega fluoróforos adicionais com diferentes excitações e/ou emissão máxima para acentuar atributos funcionais específicos da espécie. A vantagem desta técnica sobre a DIGE é que ela dá uma faixa muito mais larga de informações.

A técnica de ICAT permite identificação sistemática das proteínas numa mistura complexa e quantificação precisa das diferenças em abundância de cada proteína presente em duas ou mais amostras protéicas. O reagente ICAT é formado por três componentes funcionais: um grupamento reativo

à tiois, que é seletivo para os grupamentos sulfidrilas das cadeias laterais de cisteínas reduzidas, um grupo de etileno glicol que pode conter átomos de deutério (isotopicamente pesado) ou não (isotopicamente normal) e que possibilita realizar quantificação precisa por MS, e um grupamento de biotina que é a "etiqueta de afinidade" utilizado como base para o isolamento seletivo dos peptídios contendo cisteína (marcados com ICAT) de uma mistura complexa de peptídios, através de cromatografia de afinidade em coluna de avidina (Figura 6).

Os resíduos de cisteína reduzida das proteínas nas duas amostras a serem comparadas são rotulados separadamente durante o experimento. Em uma das amostras é usada a forma isotopicamente pesada e na outra amostra a forma leve do reagente ICAT. As duas amostras são combinadas e digeridas com tripsina. Os peptídios marcados, que contêm cisteína são isolados por cromatografia de afinidade e analisados por espectrometria de massa, determinando assim a identidade da proteína que deu origem ao peptídio e a abundância relativa de cada proteína nas amostras sendo comparadas (Figura 7) (Gygi et al., 1999).

Conclusões

A Proteômica veio para revolucionar os estudos de Genoma, uma vez que utiliza-se de suas ferramentas para identificar as proteínas expressas em situações específicas em um dado momento naquele ambiente. A Proteômica é dinâmica e permite fazer estudos comparativos. Os métodos para análise global da abundância de proteínas, modificações pós-traducionais e as mudanças nesses dois parâmetros como a função de um estado biológico da célula ou tecido se desenvolveram utilizando-se das técnicas de eletroforese, cromatografia e espectrometria de massa como componentes-chave para o estabelecimento de plataformas de estudo. As inovações tecnológicas na detecção, análise e quantificação de proteínas acelerou rapidamente o desenvolvimento da Proteômica nos últimos 4 anos. Isso, sem dúvida, ocorreu devido ao foco que institui-

ções privadas e públicas deram à pesquisa do Genoma ao Proteoma. No Brasil, a Proteômica vem se desenvolvendo desde o início do Projeto Genoma da *Xylorella fastidiosa* financiado pela Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP). Diversos grupos inseridos no Genoma Funcional da FAPESP ainda continuam trabalhando com técnicas proteômicas com o objetivo de mapear as proteínas funcionais importantes no estabelecimento da doença do "amarelinho" e a sobrevivência da bactéria nas laranjeiras variedade doce. Com os resultados obtidos destes estudos espera-se criar estratégias de controle desta doença, a qual já tem causado sérios danos à citricultura paulista.

Literatura Consultada

- Chevallet M, Santoni V, Poinas A, Rouquie D, Fuchs A, Kieffer S, Rossignol M, Lunardi J, Garin J, Rabilloud T. New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. *Electrophoresis* 1998, 19, 1901-1909.
- Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C. et al. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* 1993, 90, 5011-5015.
- James, P., Quadroni, M., Carafoli, E., Gonnet G. Protein identification by mass profile fingerprinting. *Biochem Biophys Res Commun* 1993, 195, 58-64.
- Mann, M., Hojrup, P., Roepstorff, P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* 1993, 22, 338-345.
- Pappin, D. J. Peptide mass fingerprinting using MALDI-TOF mass spectrometry. *Methods Mol Biol* 1997, 64, 165—173.
- Yates, J. R. 3rd, Speicher, S., Griffin, P. R., Hunkapiller T. Peptide mass maps: a highly informative approach to protein identification. *Anal Biochem* 1993, 214, 397-408. †