

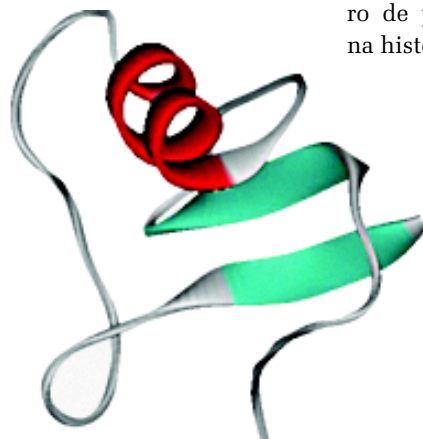
Aprendemos a ler, abrimos um livro e começamos a folheá-lo. Mas sabemos apenas algumas poucas palavras e, para ler e entender totalmente a obra, precisamos obter um vocabulário mais rico e saber ainda o significado das palavras em conjunto na formação de frases, parágrafos e capítulos. Essa analogia revela o quadro atual do conhecimento científico sobre a biologia e o desenvolvimento dos organismos. O seqüenciamento da molécula de DNA de um organismo nos revela o que está escrito no 'livro da vida' dele. No entanto, a informação contida no DNA precisa ser 'traduzida' em proteínas, as palavras do livro, para que seu conteúdo seja utilizado. O papel da pesquisa do proteoma – o conjunto das proteínas – passa a ser, portanto, o de nos ensinar o vocabulário e a compreensão geral do texto.

Adriano Monteiro de Castro Pimenta

Núcleo de Biomoléculas e Laboratório de Venenos e Toxinas Animais, Departamento de Bioquímica e Imunologia (Instituto de Ciências Biológicas), Universidade Federal de Minas Gerais

Os desafios proteômicos

A caracterização dos genomas de organismos foi uma das principais forças motrizes da ciência dos recentes anos 90. Grandes iniciativas – como o Projeto Genoma Humano, de âmbito mundial, ou a Rede Onsa, consórcio brasileiro de cientistas que seqüenciou o genoma da bactéria *Xylella fastidiosa* – possibilitaram esforços conjuntos (reunindo grande número de pesquisadores) sem precedentes na história da ciência.



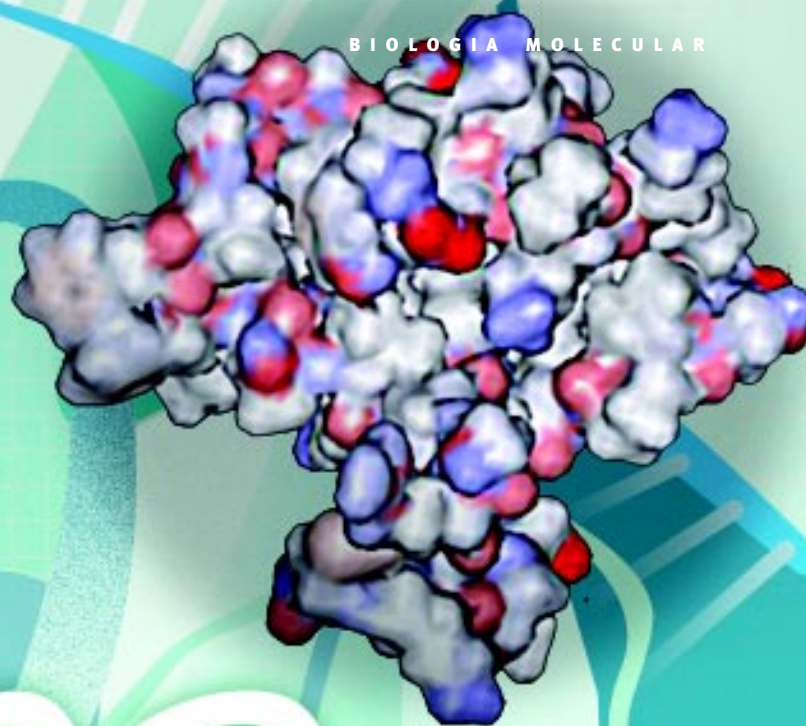
A inovação e consolidação das técnicas de análise genômica fizeram com que a biologia molecular ganhasse cada vez mais espaço também fora da área acadêmica, causando, de forma significativa e positiva, impacto para a vida cotidiana do cidadão comum. Nos últimos anos, temas relacionados às capacidades de aná-

do eoma

lise e de transformação de seres vivos, embutidas nas técnicas de biologia molecular, vieram à tona e a sociedade viu-se envolvida em discussões sobre testes de paternidade, alimentos transgênicos, terapia gênica e, principalmente, a seqüência genômica humana e a clonagem de seres vivos.

Muitos consideram o conhecimento sobre os genes apenas um primeiro passo, sendo necessária ainda uma compreensão mais aprofundada dos processos que ocorrem após o armazenamento, a duplicação e a tradução do material genético. A chamada 'era pós-genômica' começou com a constatação de que os dados trazidos pelo seqüenciamento das moléculas de DNA, embora relevantes, são limitados, sendo imprescindível investigar tanto os processos de transcrição das informações contidas no genoma quanto os seus produtos, as proteínas.

Como se sabe, os genes são compostos por quatro diferentes nucleotídeos, dispostos em uma seqüência ►



específica. Uma sequência de três nucleotídeos forma um códon. Cada códon é o código para um aminoácido, e vários aminoácidos ligados em sequência formam uma proteína. Por isso se diz que o DNA é um código genético: as informações contidas nesse código são usadas para expressar, ou seja, produzir as proteínas, as executoras das ordens do gene.

Contrariando as expectativas iniciais, o número total de genes no genoma humano é de menos de 40 mil. Aparentemente, esse número é apenas o dobro dos genomas de animais inferiores, como vermes ou moscas. Entretanto, há diferenças importantes entre esses genomas. Os genes humanos estão 'espalhados' em regiões maiores de DNA genômico e são usados para construir mais alternativas de transcrição, como é chamada a etapa inicial da produção de proteínas. Essas alternativas resultam, no homem, em talvez cinco vezes mais proteínas primárias (proteínas recém-produzidas e sem nenhum tipo de modificação química) que nos outros animais citados. O conhecimento dos proteomas, portanto, fornecerá oportunidades sem precedentes para estudos das funções dos genes.

O termo proteoma apareceu em meados dos anos 90 e designa o conjunto de proteínas expressas por um determinado genoma. O tema foi prontamente absorvido pela comunidade acadêmica como uma das vertentes da 'era pós-genômica'. A utilização dos termos relacionados ao tema cresceu rapidamente no Medline, principal banco de dados em pesquisas biomédicas, desde sua criação até o final de 2002 (figura 1).

Enquanto o genoma de um organismo permanece relativamente estável ao longo da sua vida, o proteoma é extremamente dinâmico e variável. Todas as células de um organismo, exceto as reprodutivas (os gametas), têm rigorosamente o mesmo genoma, mas apresentam formas, funções e características totalmente diferentes. Então, o que faz uma célula pancreática pertencer ao pâncreas e secretar insulina, e não ser responsável pela transmissão de impulsos nervosos no cérebro, por exemplo? O que tor-

na uma célula do músculo afilada e com propriedades contráteis e não arredondada e cheia de gordura?

Não temos ainda respostas para várias dessas perguntas, mas sabemos que o conjunto de proteínas expresso em cada uma dessas células (o seu proteoma) é extremamente diferente. A análise proteômica permite saber se e quando um produto genético está sendo expresso, a concentração relativa desse produto e, por fim, as modificações que podem ocorrer nessas proteínas após a sua tradução. Tais informações não podem ser previstas com precisão a partir das sequências dos ácidos nucleicos.

A análise proteômica vai muito além da simples listagem de proteínas e pode fornecer indícios substanciais quanto à organização e à dinâmica dos processos metabólicos, regulatórios e de sinalização através dos quais a célula se desenvolve. Mais ainda, a análise proteômica pode mostrar como esses processos se tornam disfuncionais nos estados patológicos e como podem ser manipulados, mediante, por exemplo, a administração de medicamentos ou a terapia gênica.

Portanto, uma das grandes inovações prometidas pelo conhecimento dos proteomas, além do mapeamento de intrincados caminhos metabólicos celulares, é a possibilidade de identificar novos alvos farmacológicos, novas moléculas bioativas e marcadores biológicos que podem ser usados para diagnósticos clínicos.

Como é feita a análise proteômica?

A análise proteômica baseia-se em duas plataformas metodológicas complementares.

Uma vez determinado o complexo multiprotéico a ser estudado, a primeira plataforma realiza uma separação prévia dos componentes a serem analisados na segunda plataforma. A passagem da amostra pela primeira plataforma tem dois objetivos bem estabelecidos. O primeiro é aumentar, através de uma 'separação' de moléculas, o poder resolutivo da futura análise dos componentes da amostra. O segundo é fornecer uma figura estática (uma 'fotografia') dos compostos separados, que permite uma comparação não só qualitativa, mas também quantitativa entre duas ou mais amostras 'fotografadas'. As 'fotografias' retratam o proteoma estudado em determinadas condições fisiológicas ou patológicas.

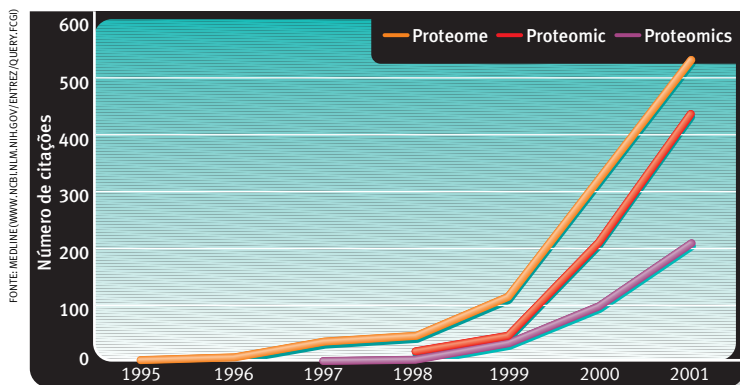


Figura 1. Evolução do uso dos termos relacionados ao estudo de proteomas nas publicações científicas integrantes do banco de dados Medline

A comparação qualitativa pode mostrar apenas se uma proteína está presente ou não na amostra estudada, enquanto a análise quantitativa revela as concentrações relativas de determinada proteína nas diferentes condições analisadas. Técnicas como a eletroforese bidimensional em gel de poliacrilamida ou a cromatografia líquida multidimensional, capazes de separar diferentes frações da amostra segundo suas propriedades físico-químicas, são usadas como primeiras plataformas.

A segunda plataforma realiza a análise e caracterização estrutural das moléculas separadas na primeira. A sensibilidade, precisão e poder resolutivo fazem dos aparelhos de espectrometria de massa a plataforma ideal para a análise estrutural e identificação das proteínas presentes na amostra estudada (ver 'Precisão e sensibilidade'). Além dessas vantagens, tais equipamentos são a 'galinha dos ovos de ouro' das grandes indústrias farmacêuticas e biotecnológicas, por sua fácil automatização e pela rapidez

de análise. Decidida a metodologia a ser empregada, a análise de uma amostra no espectrômetro de massa pode ser realizada em apenas alguns minutos.

O objetivo dessas análises é fornecer uma lista de massas moleculares (*mass fingerprinting* ou *peptide mapping*), usada para identificar sem ambigüidades a amostra ou os componentes dessa amostra. Embora haja uma superposição entre os termos, o conceito de *mass fingerprinting* – cuja tradução seria 'impressão digital de massas', dando a idéia de uma listagem única de massas moleculares – é usado principalmente para relacionar os componentes protéicos de um dado proteoma. Já o termo *peptide mapping* (mapeamento peptídico) é usado preferencialmente para descrever o padrão de clivagem (corte) por enzimas, ou fragmentação, de uma molécula protéica. Em ambos os casos, o objetivo é listar as massas dos componentes do proteoma ou dos fragmentos (peptídeos) de uma proteína. Essa 'impressão digital' é única para cada amostra analisada. ►

Precisão e sensibilidade

Desde seu surgimento, na primeira metade do século passado, a espectrometria de massa é uma peça importante na pesquisa em química de compostos orgânicos e inorgânicos. Apenas na última década, no entanto, tornou-se uma técnica analítica molecular valiosa em pesquisas biológicas, fornecendo de maneira rápida, precisa e sensível a massa de biomoléculas como proteínas, açúcares, ácidos nucleicos, lipídios e outros compostos orgânicos. Essa técnica mede a massa molecular (em daltons) a partir da razão entre a massa e a carga de moléculas ionizadas. A necessidade de ionizar as moléculas para obter um espectro de massa era um grande obstáculo no caso de moléculas biológicas. Nos métodos iniciais de ionização, as moléculas a serem analisadas tinham que estar em fase gasosa, sob alto vácuo e a altas temperaturas, condições incompatíveis com a maioria das moléculas biológicas.

Esse problema foi resolvido com a primeira grande inovação na espectrometria de massa: o surgimento da ionização branda, no fim dos anos 80. Na ionização do tipo ES (*electrospray ionization*), a amostra, solubilizada, é passada por um tubo capilar submetido a alta voltagem. Na saída do tubo forma-se uma nebulização, e as moléculas aprisionadas nas gotículas suspensas em vácuo são, então, ionizadas. Na ionização do tipo Maldi (*matrix-assisted laser desorption ionization*), a amostra é misturada a uma matriz inorgânica solubilizada em um solvente bifásico. A mistura é aplicada em uma placa de metal e, quando se promove a evaporação do solvente, as moléculas da amostra cristalizam-se junto com a matriz. Esses cristais são então bombardeados por um feixe de raio *laser* para a ionização das moléculas. Essas técnicas de ionização são as mais utilizadas hoje – para ilustrar sua importância para a ciência, dois dos ganhadores do Nobel de química de 2002, o norte-americano John B. Fenn e o japonês Koichi Tanaka, foram escolhidos pelo desenvolvimento de métodos de ionização branda para análise por espectrometria de massas de macromoléculas biológicas.

Uma segunda inovação relevante foi a conjugação de analisadores de massa em *tandem*, ou seja, em seqüência. A espectrometria de massa em *tandem* aumentou muito o poder de resolução e a sensibilidade do método, tornando-se obrigatória nas análises estruturais de moléculas biológicas. Ela torna possível selecionar uma só molécula ionizada e fragmentá-la pela colisão com um gás inerte (em geral argônio ou mesmo ar). A análise das massas dos fragmentos obtidos fornece informações sobre a estrutura da molécula original, permitindo determinar, por exemplo, a seqüência de aminoácidos de um peptídeo ou uma alteração química específica em algum resíduo de aminoácido, por exemplo. Portanto, o grande e recente impulso no estudo dos proteomas – a idéia de decifrar o conjunto de proteínas de um genoma é relativamente antiga – deu-se a partir do desenvolvimento de técnicas de espectrometria de massa direcionadas para a análise de moléculas biológicas.

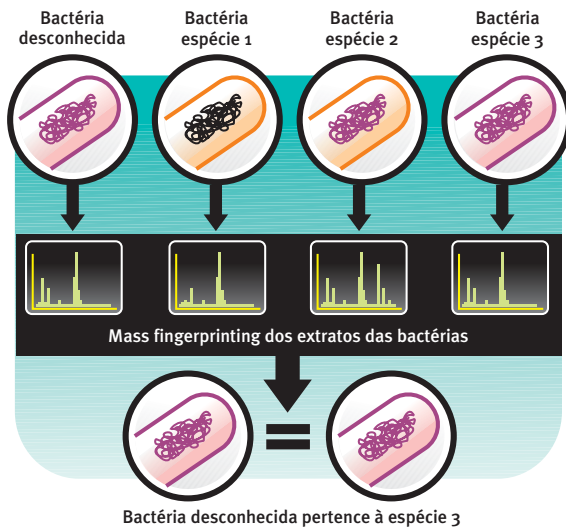


Figura 2. Identificação de células bacterianas por mass fingerprinting: a análise do extrato de uma bactéria desconhecida determina sua ‘impressão digital de massa’, o que permite identificá-la corretamente, por comparação com a mass fingerprinting de bactérias conhecidas – tipo de análise usada, por exemplo, para identificar bactérias causadoras de infecções

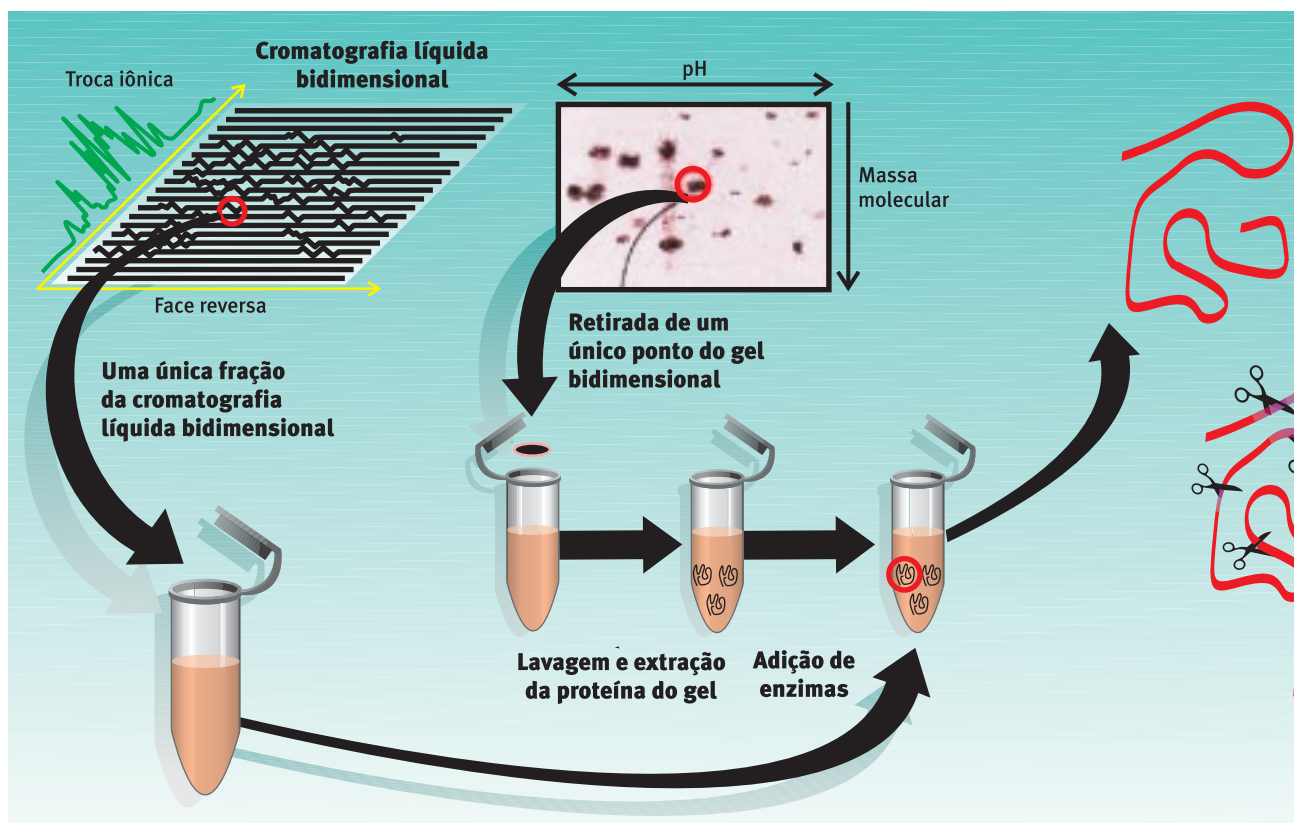
São, portanto, duas situações distintas. A primeira envolve apenas a detecção da ‘impressão digital’ das massas moleculares dos componentes de determinado extrato biológico. Nesse caso, a idéia é realizar uma classificação de diferentes extratos pela comparação das massas moleculares de seus componentes. Esse processo é extremamente útil, por exemplo, para identificar e classificar tipos de células cancerosas, diferentes microrganismos ou venenos de variadas espécies de serpentes (figura 2).

Na segunda situação, a proteína isolada na pri-

meira plataforma (por eletroforese bidimensional ou cromatografia líquida) é clivada por enzimas para a obtenção do mapeamento peptídico. A proteína intacta é clivada pela enzima em vários pontos específicos, que dependem da sequência dos aminoácidos e da própria enzima usada, resultando em vários fragmentos menores de massas moleculares diferentes. O padrão de clivagem fornecido pelas diversas massas moleculares dos fragmentos da proteína é então comparado com um padrão obtido *in silico*, ou seja, um padrão teórico de clivagem de uma sequência protéica extraída de um banco de dados de sequências genômicas ou protéicas. A comparação das massas moleculares obtidas com as do banco de dados permite identificar as proteínas (figura 3).

Nesse segundo caso, é possível ainda verificar modificações pós-traducionais, ou seja, alterações químicas ocorridas após a produção da proteína e determinantes para sua função.

Em ambas as situações acima, é possível determinar a expressão de proteínas e suas quantidades relativas em extratos biológicos. A quantificação



relativa dos componentes é extremamente importante para verificar se houve aumento ou redução da quantidade de proteína produzida. Essa regulação da expressão gênica, resultando em maior ou menor quantidade de proteína, pode estar relacionada ao aparecimento ou desenvolvimento de uma doença (como um tipo de câncer) – e medir o processo pode ajudar no diagnóstico ou no tratamento da mesma.

Esses tipos de análises proteômicas, no entanto, apresentam duas limitações importantes. A primeira diz respeito ao fato de que, se uma proteína é identificada por comparação com padrões obtidos de proteínas já analisadas, a identificação de uma proteína nova, cuja seqüência de aminoácidos ainda não é conhecida, pode ser um sério problema. A segunda limitação é que identificar a seqüência de aminoácidos das proteínas é apenas a ponta do *iceberg*. Mesmo sabendo-se que proteína é, não se pode saber sobre sua atividade, a menos que sua função já tenha sido determinada, ou que a proteína pertença a uma família com funções já conhecidas. Em última análise, uma nova proteína precisará ser purificada, ter sua estrutura e arranjos espaciais determinados e sua função testada experimentalmente *in vitro* ou *in vivo*.

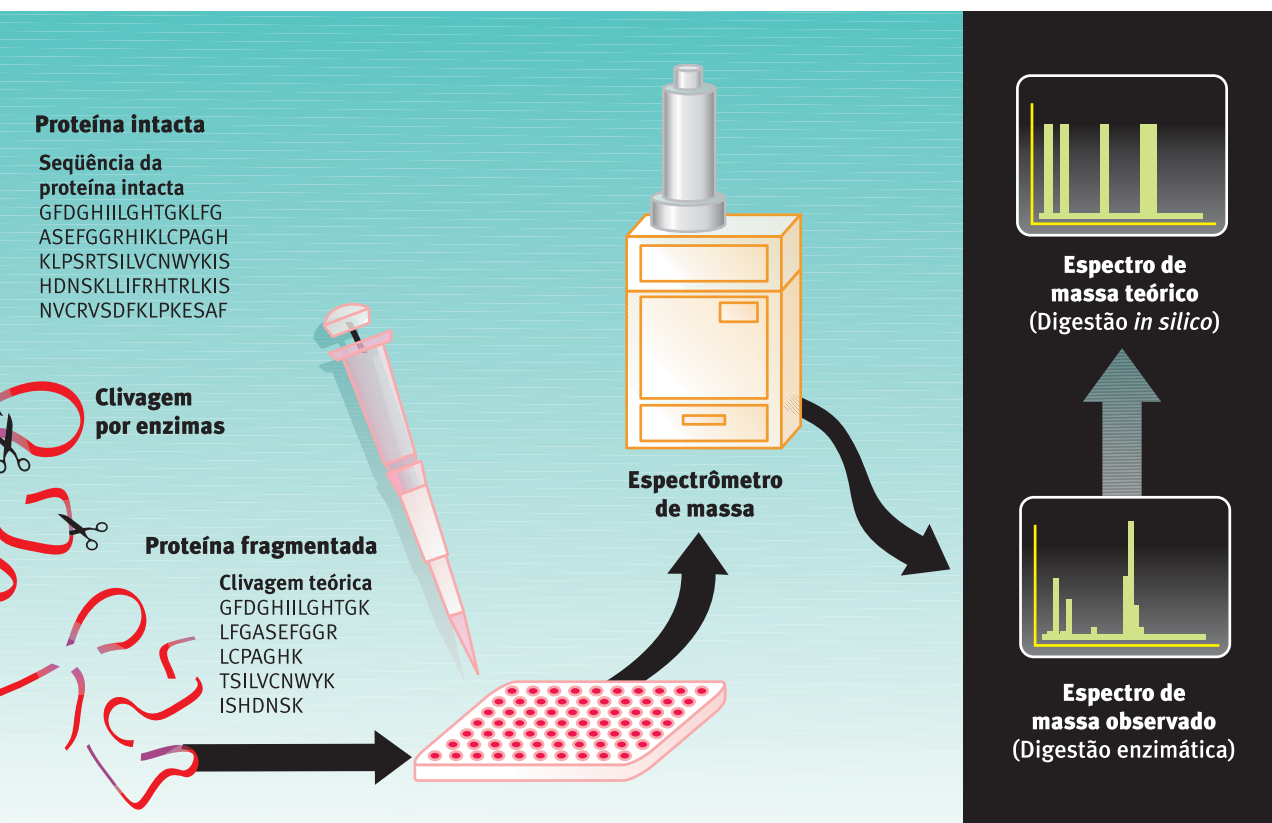
Muitas vezes, a quantidade de proteína é pequena demais e é preciso recorrer à síntese artificial ou à expressão da mesma por meio da engenharia genética de vírus ou bactérias. Tais mecanismos, porém,

não são garantidos e as proteínas assim produzidas podem não ter os efeitos biológicos esperados. Todo esse processo pode levar anos de estudo. Nesse aspecto, o estudo dos genomas apresenta uma enorme vantagem, pois as moléculas de DNA e RNA têm estrutura relativamente simples e podem ser copiadas pela técnica da ‘reação em cadeia da polimerase’ (conhecida como PCR, do inglês *polimerase chain reaction*).

Por que estudar o proteoma?

O estudo do proteoma é importante porque pode levar a três vertentes básicas com implicações diretas em vários campos da biologia e da biotecnologia.

Figura 3. Identificação de proteínas por *peptide mapping*. As proteínas de um complexo multiprotéico são separadas por cromatografia líquida ou eletroforese e em seguida uma proteína isolada (um ponto do gel) é clivada (cortada) por enzimas. Os fragmentos obtidos (peptídeos) são analisados no espectrômetro de massa, obtendo-se uma lista das massas moleculares dos mesmos, que permite a identificação da proteína original por comparação com a lista teórica existente em um banco de dados de seqüências proteicas



Sugestões
para leitura

ANDERSON, N. L.; MATHESON, A. D. & STEINER, S. 'Proteomics: applications in basic and applied biology', in *Current Opinion in Biotechnology*, v. 11 (4), p. 408, 2000.

JONSSON, A. P. 'Mass spectrometry for protein and peptide characterization', in *Cellular and Molecular Life Sciences*, v. 58, p. 868, 2001.

PIMENTA, A. M. C.; STOCKLIN, R.; FAVREAU, P.; BOUGIS, P. E. & MARTIN-EAUCLAIRE, M. F. 'Moving pieces in a proteomic puzzle: mass fingerprinting of toxic fractions from the venom of *Tityus serrulatus* (Scorpiones, Buthidae)', in *Rapid Communications in Mass Spectrometry*, v. 15 (17), p. 1.562, 2001.

POON, T. C. & JOHNSON, P. J. 'Proteome analysis and its impact on the discovery of serological tumor markers', in *Clinica Chimica Acta*, v. 313 (1-2), p. 231, 2001.

BANKS, R. E.; DUNN, M. J.; HOCHSTRASSER, D. F.; SANCHEZ, J. C.; BLACKSTOCK, W.; PAPPIN, D. J. & SELBY, P. J. 'Proteomics: new perspectives, new biomedical opportunities', in *The Lancet*, v. 356, p. 1.749, 2000.

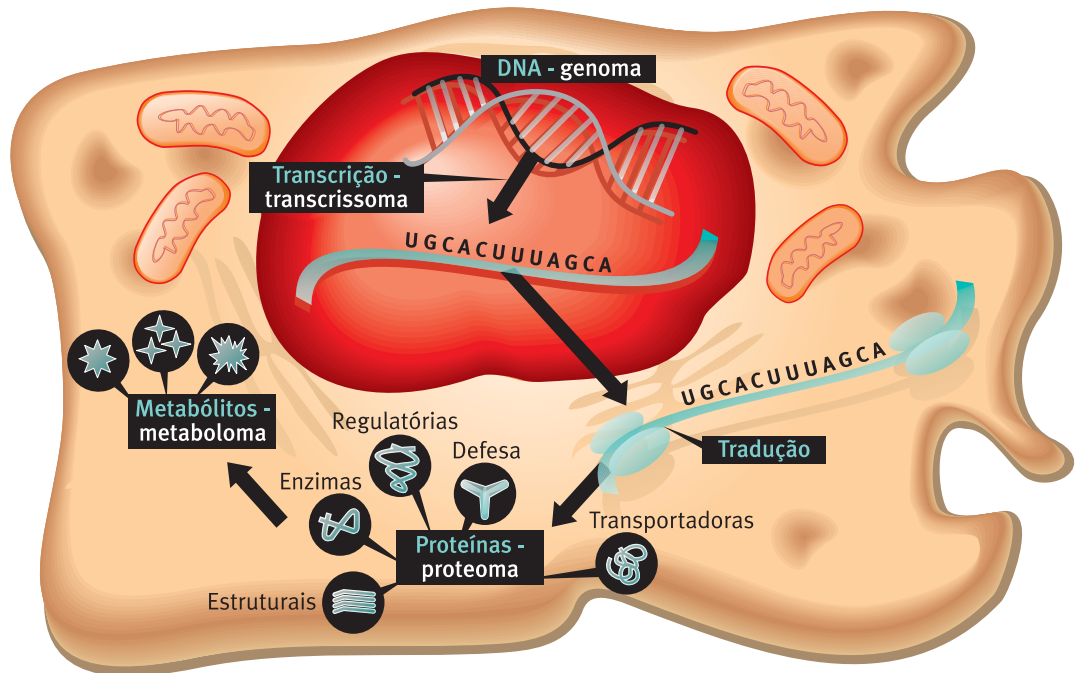


Figura 4. Fluxo das informações genéticas em uma célula: à medida que a informação genética é repassada, aumenta a complexidade molecular da célula – o sufixo ‘oma’ tornou-se sinônimo de ‘conjunto de’, e quando aplicado a uma palavra significa o estudo estrutural e dinâmico do conjunto de moléculas ou processos designados por aquela palavra

Primeiro, a caracterização proteômica é a chance de desvendar os intrincados caminhos metabólicos nas diversas etapas celulares, gerando um conhecimento sem precedentes na biologia celular. Tal conhecimento tornará possível identificar novos alvos farmacológicos específicos, relacionados, por exemplo, a determinada etapa de desenvolvimento de uma célula cancerosa.

Segundo, ela viabiliza a identificação de novas moléculas bioativas em extratos biológicos naturais, levando ao desenvolvimento de novos medicamentos.

Terceiro, permite a identificação e caracterização de marcadores biológicos, ou seja, moléculas endógenas ou exógenas específicas de um determinado estado patológico. A capacidade de identificar essas moléculas é extremamente útil no diagnóstico precoce de doenças e no acompanhamento da evolução do tratamento.

Apesar da novidade do tema, vários grupos de pesquisa, no mundo inteiro, usam as abordagens proteômicas em seus estudos. Na medicina, por exemplo, muitas aplicações já foram descritas. Inúmeros estudos proteômicos foram aplicados na caracterização do câncer, principalmente, mas também de doenças neurológicas (como o mal de Alzheimer), infecciosas e cardíacas, ou na caracterização de agentes infecciosos, como *Bacillus anthracis*

(antraz) e *Mycobacterium tuberculosis* (tuberculose). Além disso, estudos proteômicos têm sido utilizados na busca de novas moléculas bioativas em venenos de animais peçonhentos e extratos naturais e no desenvolvimento de novas drogas terapêuticas.

Outros membros da família ‘oma’

Seria ingenuidade imaginar que as respostas fornecidas apenas pelos estudos de genomas e proteomas bastariam para responder às complexas questões sobre a biologia celular, mesmo no caso dos organismos mais simples.

Um esquema do fluxograma (figura 4) dos componentes de uma célula permite ver uma complexidade química e estrutural cada vez maior. Para cada uma das etapas desse fluxograma estão sendo desenvolvidas abordagens metodológicas que visam desvendar os segredos da célula.

Assim, o ‘genoma’ estuda a molécula de DNA e a informação nela armazenada sob a forma de genes. A transcrição do DNA para o RNA, o primeiro passo do fluxo da informação genética, é investigada pelo estudo do ‘transcriptoma’, que dá uma idéia da funcionalidade do genoma daquela célula. As proteínas expressas são analisadas e identificadas pelo ‘proteoma’, enquanto o ‘metaboloma’ visa determinar os metabólitos, os produtos finais dos diversos processos celulares e que podem englobar, além dos nucleotídeos e aminoácidos, os açúcares, lipídios, esteróides e mais uma infinidade de outras moléculas importantes para a manutenção da atividade biológica. ■