

# 13

# Transcriptômica e Proteômica

Ricardo De Marco

Atualmente, o estudo dos conjuntos de moléculas de RNA e proteínas em uma célula é definido como transcriptômica e proteômica. Ambos os conjuntos são originários de um processo dinâmico nas células, derivado de programações de respostas ao ambiente, ou de processos de diferenciação celular. Neste capítulo, será abordada a importância do estudo de alterações celulares do ponto de vista da dinâmica de mudanças de transcriptomas e proteomas. São comentados métodos para identificar essas alterações do transcriptoma, desde análises de hibridação de ácidos nucleicos (p. ex., *Northern blot* e microarranjos) até técnicas modernas de sequenciamento de genes. Também são abordados métodos para determinação do conjunto de proteínas expressas em células e separação em gel 2D, incluindo análises em espectrômetros de massa para identificação de proteínas específicas e de interação proteica, através de duplos híbridos.

## O que são o transcriptoma e o proteoma?

---

As células dos seres vivos apresentam diversos conjuntos de biomoléculas com diferentes funções. As principais classes dessas biomoléculas são carboidratos, lipídeos, proteínas e ácidos nucleicos. As proteínas compõem um conjunto muito importante de biomoléculas, pois têm grande importância na efetivação de uma série de processos celulares. Ácidos nucleicos são as moléculas responsáveis pela transmissão da informação genética na forma de DNA e intermediário para produção de proteínas por esse repositório, na forma de RNA.

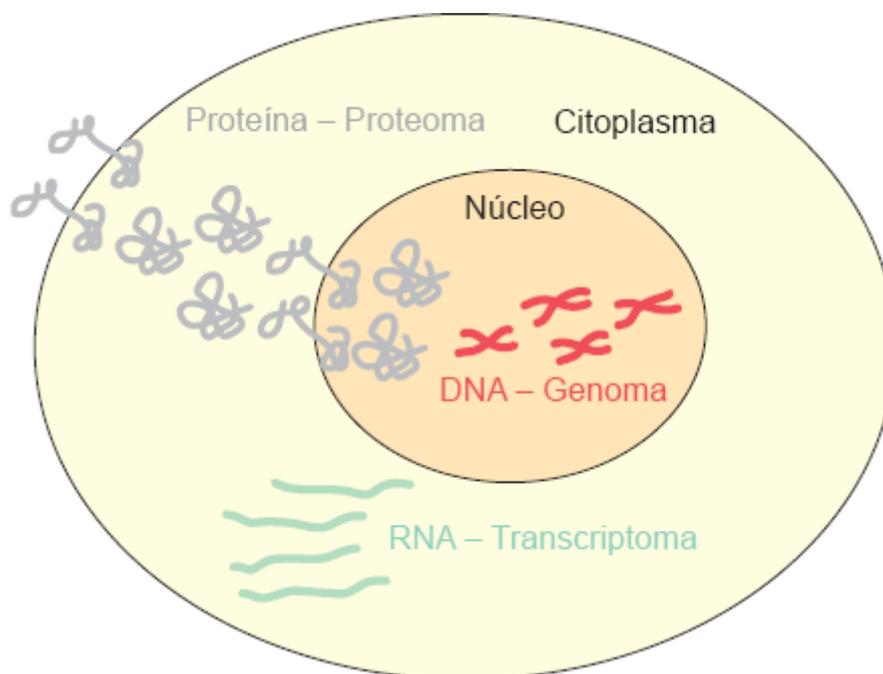
Com a finalidade de sistematizar o estudo dessas biomoléculas, passou-se a designar esses conjuntos com diferentes nomes. Por exemplo, o conjunto das moléculas de DNA que compõe o repositório genético de um indivíduo é designado genoma. Por sua vez, o conjunto de moléculas de RNA de uma célula é o seu transcriptoma, e o conjunto de proteínas dessa mesma célula é designado proteoma. Não é surpreendente que esses três conjuntos (genoma, transcriptoma e proteoma) estejam intimamente relacionados, pois, de acordo com o dogma central da biologia molecular, as moléculas de DNA servem como molde para a transcrição de moléculas de RNA que, por sua vez, serão traduzidas em proteínas (Figura 13.1).

Resumidamente, sem considerar as modificações pós-traducionais de proteínas, por exemplo, toda a informação do transcriptoma estaria contida no genoma, e o proteoma, por sua vez, poderia ser deduzido a partir do transcriptoma. No entanto, tal afirmação não leva em conta o caráter dinâmico do transcriptoma e do proteoma de um organismo. Desse modo, o estudo desses dois conjuntos de moléculas não é uma mera extensão do estudo de genomas, constituindo campos próprios, que irão prover diferentes tipos de informação e com diversos desafios na obtenção e interpretação dos dados.

O transcriptoma é composto de diferentes tipos de moléculas de RNA, mas o seu componente mais estudado é o RNA mensageiro (mRNA); e isso acontece a tal ponto que, muitas vezes, o termo transcriptoma é utilizado apenas para designar o conjunto de mRNA de uma célula, ignorando os outros tipos de RNA. Tal preferência é compreensível, visto que apenas o mRNA tem função de codificação de proteínas e a maioria dos outros tipos de RNA (rRNA, tRNA, snRNA) tem função estrutural. Esses outros tipos de RNA contêm um número limitado de moléculas distintas, existindo maior importância no seu estudo individual e entendimento da sua função em alguns eventos celulares, mas não na sua caracterização enquanto grupo de moléculas, como geralmente realizado em

estudos transcriptômicos. Uma exceção é o grupo de RNA não codificantes (ncRNA) – um conjunto de RNA regulatórios (cujas propriedades serão melhor exploradas no Capítulo 16), que também são comumente alvos de estudos transcriptômicos.

O proteoma, por sua vez, é formado pelo conjunto de proteínas de um organismo. Proteínas são moléculas complexas e que apresentam comportamentos bem distintos entre si. Dependendo dos aminoácidos que constituem uma proteína, ela pode ser solúvel ou estar inserida em uma membrana, por exemplo. Além disso, diversos processos realizados por proteínas dentro da célula eucariótica ocorrem em compartimentos específicos. Desse modo, não é de se estranhar que a distribuição de uma dada proteína dentro de uma célula seja heterogênea. Por exemplo, existem proteínas responsáveis pela transcrição do DNA que são encontradas somente dentro do núcleo, ou transportadores que são encontrados somente na membrana plasmática da célula (Figura 13.2). Em adição a sua cadeia proteica, muitas proteínas sofrem modificações pós-traducionais, nas quais moléculas (como carboidratos ou fosfatos) são adicionadas às cadeias laterais de alguns aminoácidos que compõem a proteína madura. Conseqüentemente, o proteoma de uma célula não é resultado da mera tradução dos transcritos que se encontram no seu citoplasma, mas o resultado de uma série de processos que irão gerar proteínas maduras e distribuídas de modo desigual dentro de uma célula.

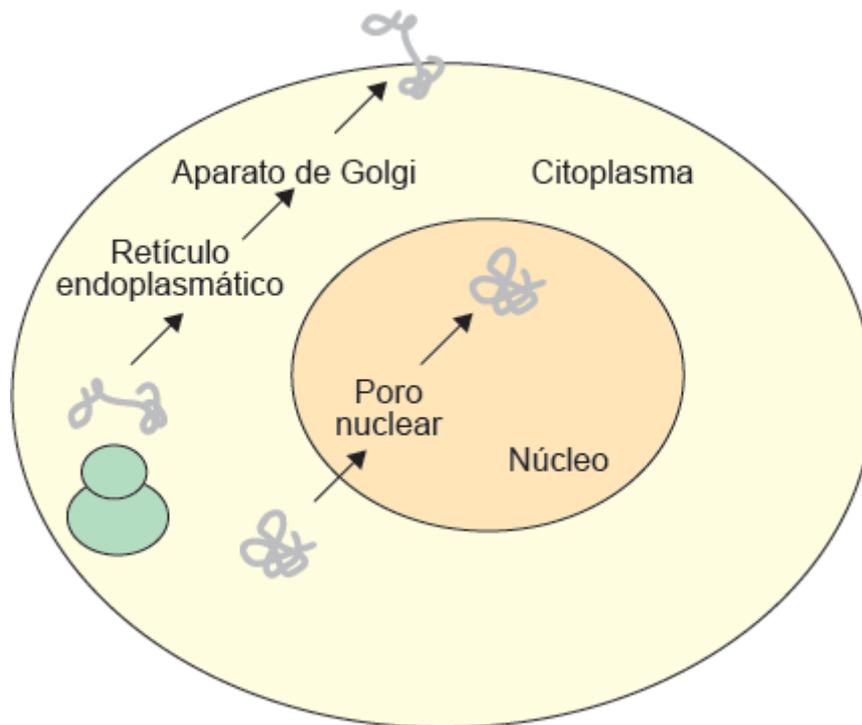


**Figura 13.1** Célula eucariótica e localização de alguns dos conjuntos de biomoléculas.

## **Dinamicidade do transcriptoma e complexidade das respostas celulares**

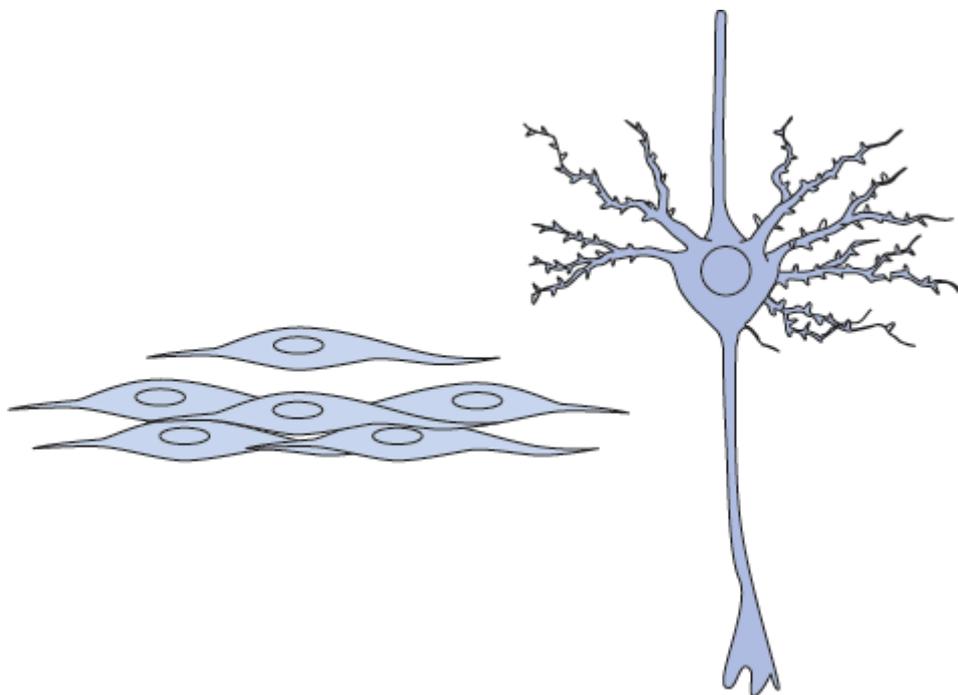
A princípio, toda a informação necessária para o desenvolvimento e funcionamento de uma célula encontra-se em seu genoma. No entanto, ao comparar diferentes células de um mesmo organismo, observa-se que elas apresentam características bem distintas. Por exemplo, comparando um neurônio e uma célula muscular (Figura 13.3), é possível notar que elas têm morfologias bem diversas, que são reflexo de sua função; no caso, a condução de impulsos nervosos para os neurônios e a realização de movimentos, respectivamente. Apesar de as informações contidas no genoma dessas duas células serem praticamente as mesmas, a utilização de tal informação ocorre de modo diferente. É possível notar, portanto, que cada célula do nosso corpo dispõe de uma programação genética distinta que irá determinar o seu fenótipo.

O que é, na prática, tal programação genética e como ela consegue produzir diferenças nas células? Primeiramente, é necessário lembrar-se de que o material genético da célula (o DNA) é relativamente inerte e que grande parte das funções celulares é promovida por proteínas. No entanto, um determinado tipo de proteína somente será produzido se o mRNA correspondente estiver disponível para tradução no citoplasma da célula; quanto maior a quantidade de um determinado tipo de mRNA, maior será a produção da proteína correspondente.

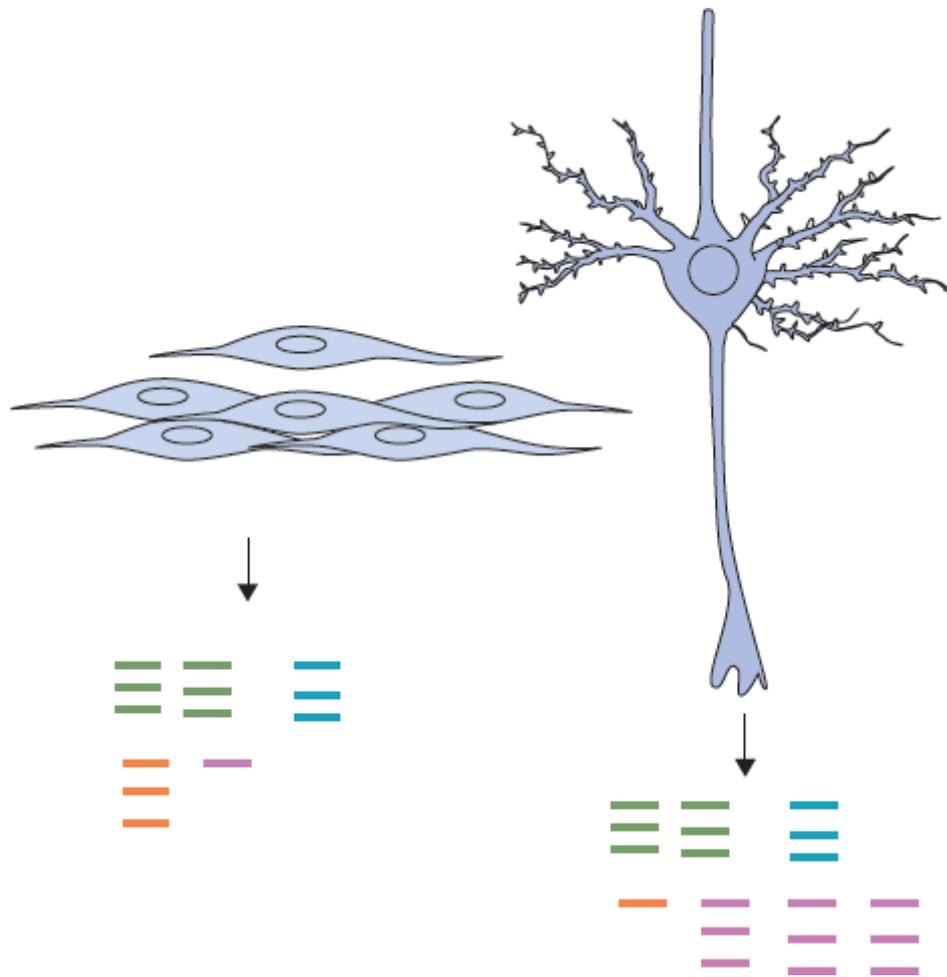


**Figura 13.2** Algumas das diferentes rotas que a proteína pode percorrer em uma célula eucariótica.

Existem alguns mRNA que codificam para proteínas que realizam funções básicas nas células (chamadas, em inglês, de *housekeeping genes*), para os quais geralmente não se verificam muitas mudanças em sua concentração no citoplasma entre os diferentes tipos celulares. No entanto, há uma série de outros mRNA que codificam para proteínas com funções mais especializadas, cuja quantidade pode variar significativamente entre dois tipos celulares diferentes. Como resultado, cada célula poderá apresentar um diferente repertório de moléculas de mRNA (Figura 13.4). Assim, dois tipos celulares distintos de um indivíduo devem apresentar um genoma praticamente idêntico, mas um transcriptoma bem diferente.



**Figura 13.3** Morfologia de células musculares (A) e nervosas (B).



**Figura 13.4** Representação dos conjuntos de transcritos distintos existentes em diferentes tipos celulares. As barras de diferentes cores representam mRNA originários de diferentes genes.

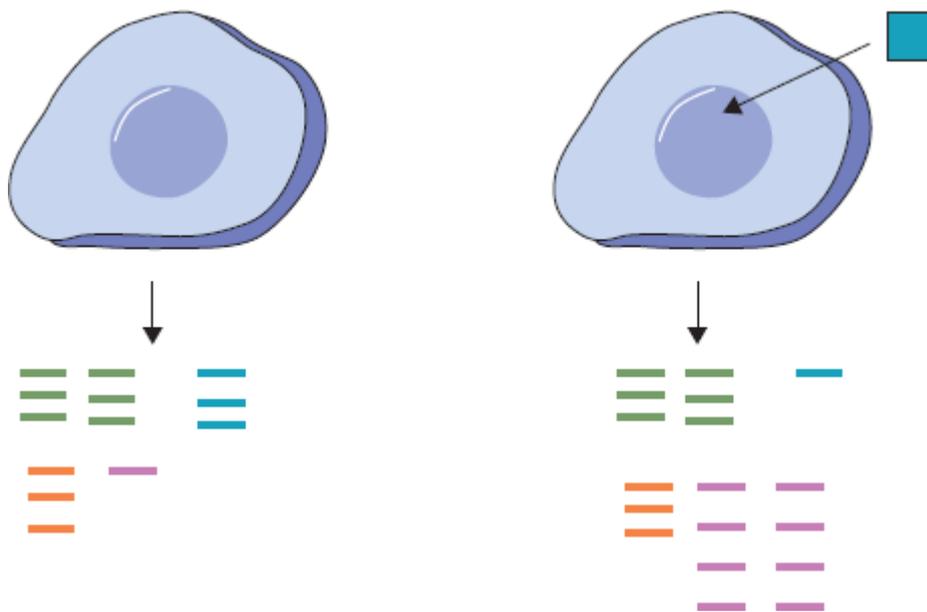
Ao analisar o transcriptoma de um dado tipo celular, estamos inferindo a programação dessa célula por meio dos níveis de mensagens de mRNA codificando para as diferentes proteínas. Por exemplo, ao analisar o transcriptoma de uma célula nervosa, certamente serão encontrados maiores níveis de mRNA codificando para proteínas de canais iônicos, responsáveis pela condução de impulsos nervosos, do que em células de epitélio, que não apresentam essa função. Embora haja alguns casos em que é óbvia a associação de função com a abundância de transcritos em uma célula, existem muitos nos quais a função da proteína é desconhecida ou o contexto celular no qual uma proteína atua não é bem entendido. Nesse caso, a verificação de maior abundância de um determinado transcrito em um determinado tipo celular pode servir como ponto de partida para tentar correlacionar a função da proteína codificada pelo transcrito com as peculiaridades do tipo celular.

Do mesmo modo, se os transcriptomas de tipos celulares idênticos de organismos diferentes forem comparados, eventuais diferenças entre eles podem refletir adaptações de cada espécie. Esse tipo de análise faz mais sentido quando comparamos duas espécies relativamente próximas evolutivamente, onde o conjunto de transcritos é praticamente o mesmo e, portanto, é relativamente fácil associar todos os transcritos de um organismo ao seu equivalente no outro organismo. Ao comparar os genomas dos seres humanos com os dos chimpanzés, encontra-se apenas 1,23% de diferença no conjunto de bases dos dois genomas. Apesar do reconhecimento que mutações que causem alterações na sequência proteica têm importância nas mudanças fenotípicas entre as duas espécies, é cada vez mais reconhecido que a alteração do padrão de transcrição das células representa um fator significativo no processo de diferenciação das duas espécies.

Tudo isso implica a necessidade de um sistema complexo de regulação das quantidades de todos os mRNA que uma célula produz. De fato, há diversas vias de sinalização cujo alvo final é uma classe de proteínas com a função de regular a produção diferencial de RNA, os chamados fatores de transcrição. Estes agem sobre a maquinaria de transcrição da célula, estimulando a transcrição a partir de certos *loci* específicos do genoma. Em razão desse complexo sistema de controle, é possível afirmar que a diversidade de transcriptomas observada entre diferentes tipos celulares ou de organismos não é fruto do acaso.

Em adição à diversidade intrínseca que o transcriptoma apresenta entre os diversos tipos celulares, o transcriptoma de uma célula também pode sofrer mudanças em decorrência de estímulos aos quais a célula é exposta. Por exemplo, a exposição de uma célula a moléculas sinalizadoras, como hormônios, pode levar a uma mudança no seu programa genético, o que pode se refletir na produção de diferentes transcritos, codificando para proteínas que exerçam funções críticas para os eventos sinalizados (Figura 13.5). Como esses sistemas de sinalização agem em cascata, é possível que a ligação de poucas moléculas de hormônios na superfície celular leve à estimulação de centenas de moléculas de fatores de transcrição que irão atuar em diferentes regiões do genoma, estimulando a produção de diversos transcritos distintos. Assim, é possível que pequenos estímulos levem a grandes mudanças no transcriptoma de uma célula. Por exemplo, determinado hormônio pode sinalizar para maior abundância de glicose, o que fará com que as células do fígado produzam maior quantidade de transcritos codificando proteínas responsáveis pela importação e armazenamento desse açúcar na forma de glicogênio.

Diversos pesquisadores se dedicam ao estudo do transcriptoma de células submetidas a diversas condições ou estímulos. São experimentos nos quais as células são expostas a diferentes tipos de estímulo e, depois, têm o seu mRNA extraído para possibilitar a sua comparação com o mRNA de células controle que não foram submetidas a este mesmo estímulo. Por meio de uma análise global dos RNA utilizando diferentes métodos, é possível comparar os dois transcriptomas e detectar transcritos cuja abundância tenha sido alterada pelo tratamento. A partir da análise desses dados, busca-se por padrões que possibilitem inferir uma programação genética específica de respostas a esses estímulos. Experimentos utilizando séries temporais, em que amostras de células submetidas a diferentes tempos de tratamento são analisadas, permitem acompanhar os diferentes passos da resposta celular ao estímulo. Esse tipo de estudo pode auxiliar no entendimento de mecanismos moleculares de adaptação das células às mais diversas condições.



**Figura 13.5** Mudança do conjunto de transcritos de uma célula, mediante estímulo de uma substância química (representada pelo quadrado).

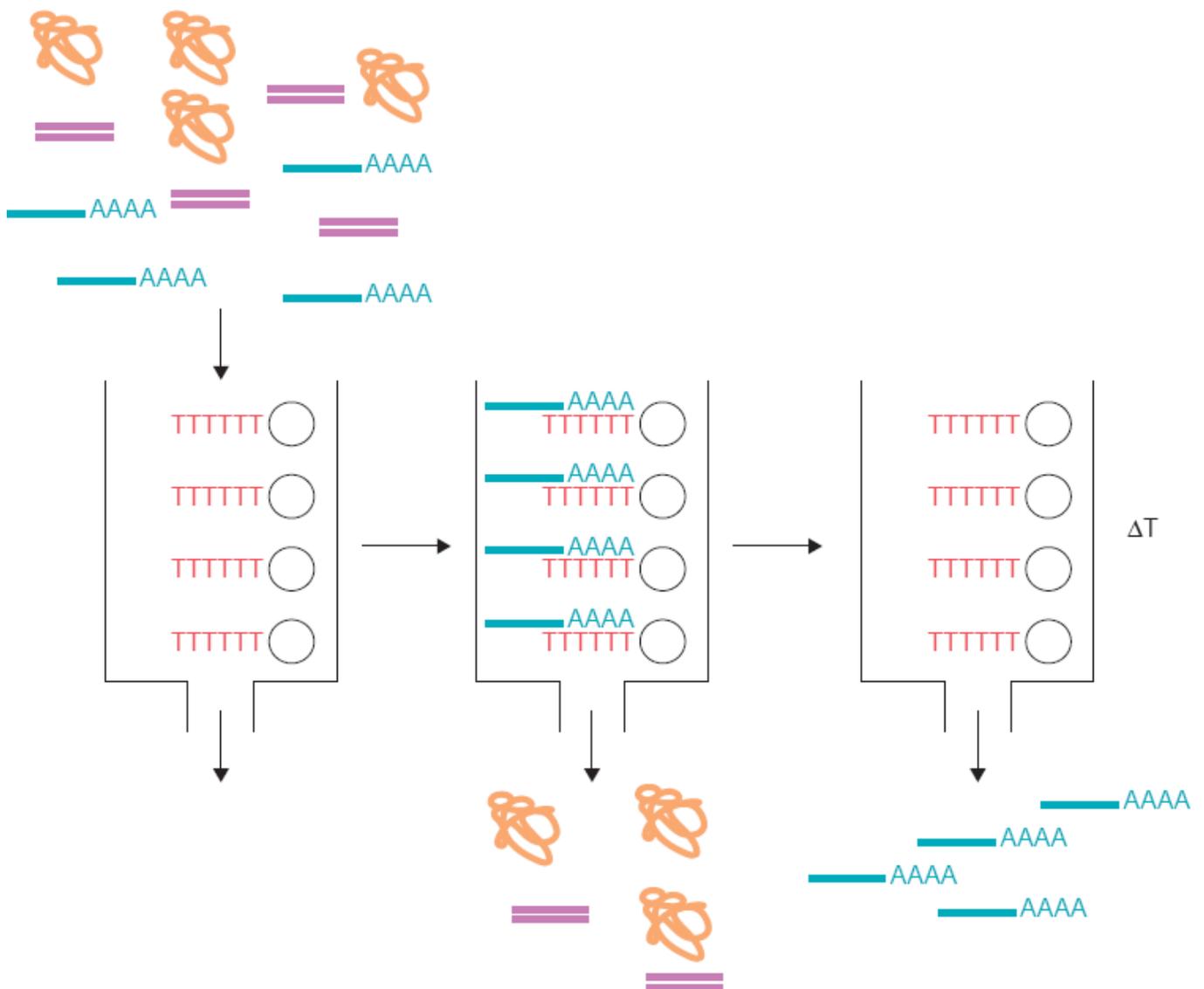
## Métodos para estudar o transcriptoma

### Isolamento de RNA mensageiro e síntese de DNA complementar

Conforme descrito na seção anterior, há grande interesse no estudo do transcriptoma, pois, por meio da descrição dos transcritos e sua abundância relativa nos diferentes tipos celulares ou organismos, é possível inferir uma série de mecanismos celulares subjacentes aos processos específicos de cada célula. Em geral, o primeiro passo para estudar o transcriptoma de uma célula é pelo isolamento do seu mRNA. Há diversos métodos utilizados para separar o mRNA dos outros componentes da célula; o mais utilizado é a passagem do material biológico por uma coluna que contém polímeros de desoxitimidina (poli-dT) imobilizados em uma resina. Como resultado, as moléculas de RNA de organismos eucariotos que têm uma cauda de adeninas em sua porção 3' serão hibridadas com o poli-dT e passarão a ficar imobilizadas nesta coluna, enquanto as outras moléculas biológicas continuarão livres, podendo ser retiradas facilmente da coluna (Figura 13.6). É muito importante retirar proteínas e DNA genômico existentes no

homogenato celular das quais são extraídas as moléculas de RNA, visto que, dentre as proteínas presentes, podem estar proteínas degradadoras do RNA, e o DNA genômico pode ser um contaminante que irá interferir nas análises em algumas das técnicas normalmente utilizadas para o estudo do mRNA.

Após o isolamento do mRNA, este costuma ser convertido em uma molécula de DNA por meio do uso de uma enzima chamada transcriptase reversa. Ela tem origem viral e é utilizada por esses vírus para transformar seu material genético na forma de RNA em moléculas de DNA que podem ser integradas no genoma da célula infectada. Como o RNA serve de molde para a síntese da molécula de DNA, toda a informação contida no RNA é preservada na molécula de DNA. O motivo pelo qual convertemos a molécula de mRNA em uma molécula de DNA deve-se ao fato de que moléculas de RNA são facilmente degradadas por enzimas chamadas RNases, produzidas em grandes quantidades por todos os organismos vivos e presentes no ambiente. Além disso, uma série de técnicas de manipulação foi desenvolvida para moléculas de DNA e, ao transformar moléculas de RNA em DNA, é possível fazer uso delas para o estudo de nossas moléculas. O DNA proveniente deste processo de transcrição reversa é denominado DNA complementar (cDNA).



**Figura 13.6** Purificação de mRNA com a utilização de colunas com oligonucleotídeos poli-dT imobilizados. Colunas contendo oligonucleotídeos poli-dT imobilizados (em vermelho) são carregadas com um homogenato celular contendo moléculas de mRNA (barras azuis), DNA genômico (barras duplas roxas) e proteínas (enovelados laranjas). Somente as moléculas de mRNA são retidas na coluna, devido à complementaridade do oligo-dT com as caudas poli-A presentes na região 3' do mRNA. Após lavagem da coluna para retirada de todo DNA e proteínas contaminantes, o mRNA puro é recuperado com o aumento da temperatura, que desestabiliza a ligação deste com o oligo-dT.

### Sequenciamento de bibliotecas de DNA complementar

O modo mais simples de explorar cDNA é realizar o seu sequenciamento. Técnicas de sequenciamento de DNA foram desenvolvidas, inicialmente, na década de 1970 e possibilitam que a ordem de nucleotídeos em uma molécula

de DNA seja deduzida. Com o intuito de permitir o sequenciamento sistemático de cDNA derivadas de um transcriptoma, foram desenvolvidas técnicas de montagem das chamadas bibliotecas de cDNA, que consistem na inserção das moléculas de cDNA geradas a partir da reação de transcrição reversa em um vetor de propagação (em geral, um plasmídeo), por meio de técnicas de clonagem. A clonagem é realizada utilizando as moléculas de cDNA geradas a partir de todas as moléculas de mRNA extraídas de um tecido ou célula. Em função disso, as moléculas de cDNA da biblioteca irão representar o conjunto de moléculas de mRNA daquele tecido ou célula no momento em que ocorreu a extração de seu mRNA. A inserção em um vetor de propagação torna possível que essas moléculas de cDNA possam ser utilizadas na transformação de células bacterianas, o que permitirá que elas sejam facilmente estocadas e amplificadas. O sequenciamento desses clones possibilita que se defina a ordenação dos nucleotídeos do polímero de ácido nucleico, e a técnica predominantemente utilizada é a técnica de Sanger.

O sequenciamento em larga escala de clones de bibliotecas de cDNA utilizando a técnica de Sanger foi popularizado na década de 1990, com o maior acesso de pesquisadores a sequenciadores automáticos. Esse tipo de equipamento possibilita que a realização de um grande número de sequenciamentos de moléculas distintas (cerca de 100 seqüências em máquinas comerciais modernas) ocorra de maneira paralela em um único experimento. Em geral, esse tipo de sequenciamento é realizado por escolha randômica de clones e realização de uma única reação de sequenciamento por clone. Isso faz com que cada clone geralmente não seja sequenciado em toda a sua extensão, pois uma reação de sequenciamento é capaz de determinar de 600 a 800 bases de um polímero e, em geral, cada molécula de cDNA formada a partir do mRNA contém milhares de bases.

Esse tipo de seqüência parcial gerada a partir de seqüências de cDNA foi denominado, em inglês, *expressed sequence tag* (EST), que significa “etiqueta de seqüência expressa”. A ideia associada à produção desse tipo de seqüência deriva do fato de que é extremamente trabalhoso gerar seqüências completas de todos os transcritos de uma célula ou tecido, pois requer a obtenção de clones de cDNA que representem o transcrito em toda sua extensão. Isso nem sempre é comum, devido à baixa processividade da transcriptase reversa. Além disso, mesmo quando obtemos um clone completo, é necessário realizar múltiplas reações de sequenciamento iniciadas em pontos diferentes da molécula para obter uma cobertura total de sequenciamento.

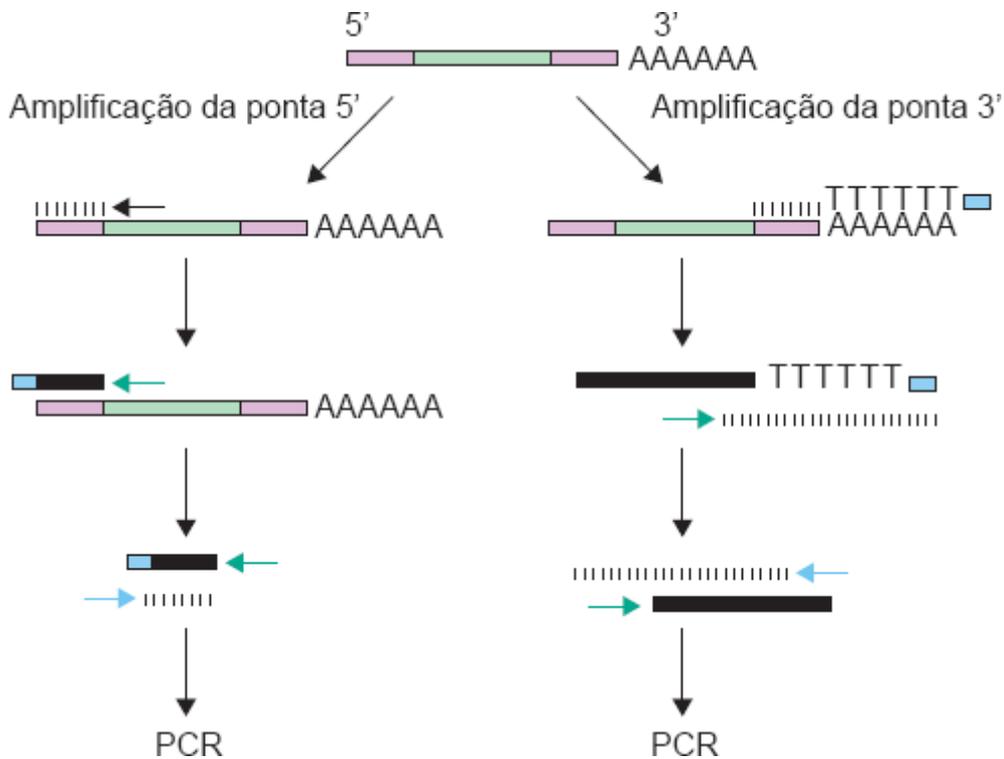
A geração de seqüências parciais de transcritos é relativamente mais direta, pois exige apenas que a molécula de cDNA seja clonada e a realização de uma única reação de sequenciamento iniciada em uma das pontas da molécula. Muitos pesquisadores preferem fazer um sequenciamento parcial, que possibilite a geração de seqüências representando uma grande amostragem dos transcritos de uma célula, em vez de produzir seqüências completas de poucos transcritos.

Apesar de esse tipo de abordagem fornecer uma informação incompleta sobre o transcriptoma, ele é bastante útil, pois, mesmo com o sequenciamento parcial de um transcrito, geralmente é possível – por meio de análises bioinformáticas – sugerir a função da proteína codificada por ele. Isso é feito pelo alinhamento da seqüência parcial da proteína codificada pelo transcrito, com proteínas com função conhecida presentes em bancos públicos de dados sobre moléculas biológicas (esse alinhamento é realizado utilizando o programa BLAST). Caso haja um alto nível de similaridade detectado entre o fragmento da proteína codificada pelo transcrito e alguma proteína com função conhecida, pode-se atribuir uma possível função à proteína codificada por esse transcrito. Obviamente, a confirmação da função da proteína somente é possível a partir da obtenção da seqüência completa do transcrito e a realização de experimentos bioquímicos com a proteína derivada deste transcrito. A partir de uma seqüência parcial, pode-se recuperar o transcrito completo por meio do clone correspondente na biblioteca de cDNA, ou de técnicas de amplificação de pontas de cDNA (Figura 13.7).

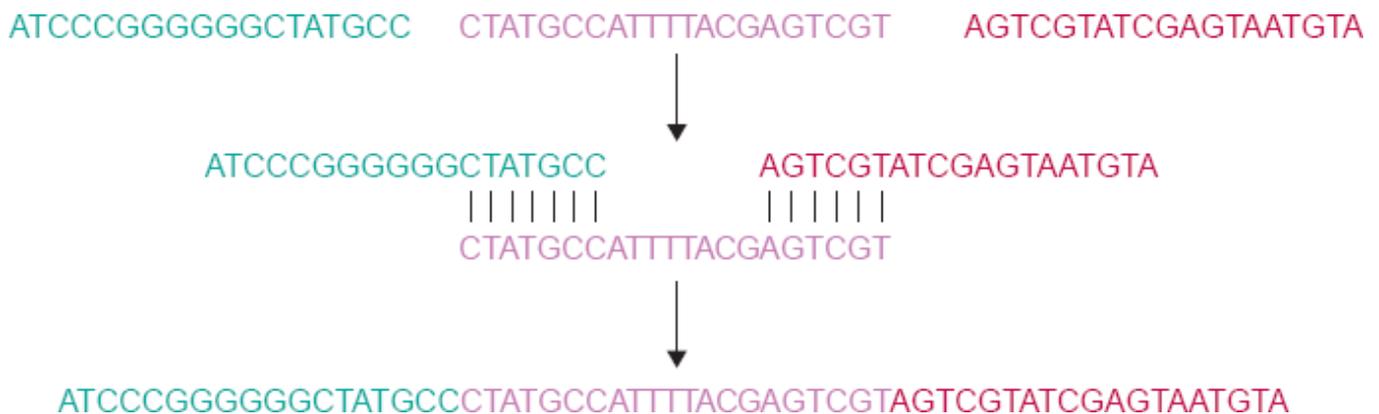
Com a produção de um grande número de seqüências parciais de transcritos de bibliotecas de cDNA, surgiu a necessidade de criar bancos de dados específicos para armazenar esse tipo de informação. Em alguns casos, o banco de dados gerado é um mero repositório onde estão disponíveis todas as seqüências parciais geradas. No entanto, há casos em que análises computacionais irão agregar todas as seqüências parciais produzidas para cada transcrito em uma única seqüência, esperando-se obter seqüências mais completas e diminuir a redundância desses bancos de dados (Figura 13.8).

Além disso, a contagem da frequência de sequenciamento de clones de um mesmo transcrito pode ser utilizada como uma medida indireta de sua abundância relativa. Isso ocorre porque a amostragem de clones da biblioteca tende a ser randômica; logo, se supormos que uma célula tem 100.000 transcritos e, dentre estes, existe um transcrito A de alta abundância com 1.000 cópias, a chance de este ser amostrado no sequenciamento de uma cópia

é de 1/100. Considerando um transcrito B de baixa abundância com apenas 10 cópias, a chance de esse transcrito ser amostrado é de apenas 1/10.000. Em um sequenciamento randômico de um número razoável de clones da biblioteca de cDNA desse exemplo hipotético, a frequência com que o transcrito A será amostrado será 100 vezes maior que o transcrito B. A partir desse princípio, é possível comparar a frequência de sequenciamento de clones em bibliotecas de cDNA originárias de diferentes amostras para inferir quais transcritos são mais abundantes em cada amostra, possibilitando levantar-se hipóteses sobre o papel desses transcritos nos diferentes processos celulares.



**Figura 13.7** Técnica de amplificação de pontas de mRNA. A barra no topo da figura representa uma molécula de mRNA, contendo uma região central, cuja sequência é conhecida (em verde), e pontas desconhecidas (em rosa). As setas representam oligonucleotídeos iniciadores sintéticos; e a região preta em pontilhado, uma fita de cDNA recém-sintetizada. A barra azul representa um adaptador de DNA cuja sequência é conhecida e que é ligado à ponta 5', ou encontra-se adjacente a um oligonucleotídeo poli-dT. Nota-se que, no passo final, a molécula é flanqueada por dois oligonucleotídeos iniciadores, o que irá possibilitar a sua amplificação pela reação em cadeia da polimerase (PCR).



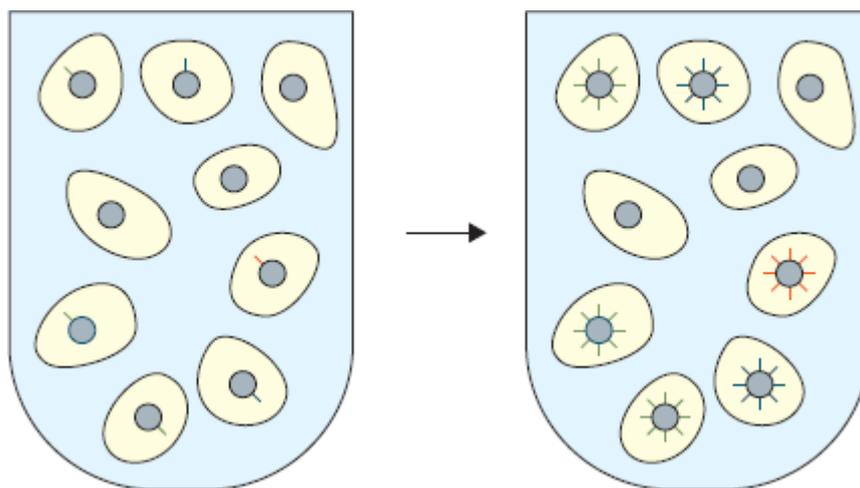
**Figura 13.8** Procedimentos utilizados para agregar sequências parciais de um transcrito para a dedução de uma sequência mais completa. Textos em verde, roxo e vermelho escuro representam diferentes sequências parciais obtidas de um transcrito. A partir da determinação de regiões de sobreposição entre as pontas dessas sequências (representada por barras horizontais), é possível deduzir uma sequência final que agregue as informações de todas juntas.

Nos últimos anos, os métodos de sequenciamento de DNA evoluíram muito com o surgimento de aparelhos de sequenciamento de segunda geração (*next generation sequencing*, em inglês). O sequenciamento tradicional utilizando a metodologia de Sanger exige que cada molécula de cDNA seja isolada por técnicas de clonagem e, depois disso, uma reação de sequenciamento separada é realizada para cada clone diferente. Esse tipo de

procedimento é extremamente trabalhoso quando se pretende sequenciar milhares de clones derivados de uma mesma biblioteca. Na tecnologia de segunda geração, há um processo de paralelização das reações de sequenciamento fazendo com que ocorram simultaneamente, dentro de um único ambiente, milhares de reações. A Figura 13.9 representa uma reação em cadeia da polimerase em emulsão, que exemplifica uma das técnicas que são utilizadas para possibilitar essa paralelização de processos. Isso faz com que haja um grande aumento do número de seqüências de cDNA que podem ser geradas a um custo relativamente baixo. Esse tipo de técnica está se tornando cada vez mais popular para a caracterização de transcriptomas, já que possibilita uma grande cobertura dos transcritos de uma célula ao custo de alguns poucos milhares de dólares.

### Análise do transcriptoma com técnicas de hibridação

Além das técnicas de sequenciamento, existem outros tipos de técnicas que possibilitam inferir a abundância dos mRNA em uma amostra. Várias técnicas de detecção de mRNA se baseiam na hibridação entre fitas de DNA e mRNA ou entre fitas de DNA e cDNA. É possível utilizar esse tipo de técnica para detectar uma molécula específica de DNA, pois ácidos nucleicos tendem a formar duplas-fitas somente quando há duas fitas complementares. Portanto, se houver um conjunto de moléculas de ácido nucleico na forma de fita simples, é possível utilizar uma molécula fita simples específica como sonda e verificar a quantidade de moléculas no conjunto inicial de ácidos nucleicos que são capazes de se hibridar com essa sonda. Com base nesse princípio, foram desenvolvidas técnicas como o *Northern blot* e de microarranjos de DNA, que possibilitam a detecção do nível de transcrição de um ou mais transcritos e diversas amostras. Ao contrário do sequenciamento, esse tipo de técnica exige conhecimento prévio da seqüência das moléculas de RNA para as quais se quer medir os níveis de transcrição, pois as sondas serão sintetizadas a partir das seqüências dos transcritos que se deseja monitorar. Assim, esse tipo de técnica não serve para a primeira exploração de um transcriptoma.



**Figura 13.9** Reação de amplificação de DNA em emulsão. Devido à emulsificação da solução, cada gotícula formada funcionará como um reator em separado. As fitas de DNA amplificados se ligam a partículas especiais colocadas na solução, de modo que, após a quebra da emulsão, seja possível separar as moléculas provenientes de cada microrreator.

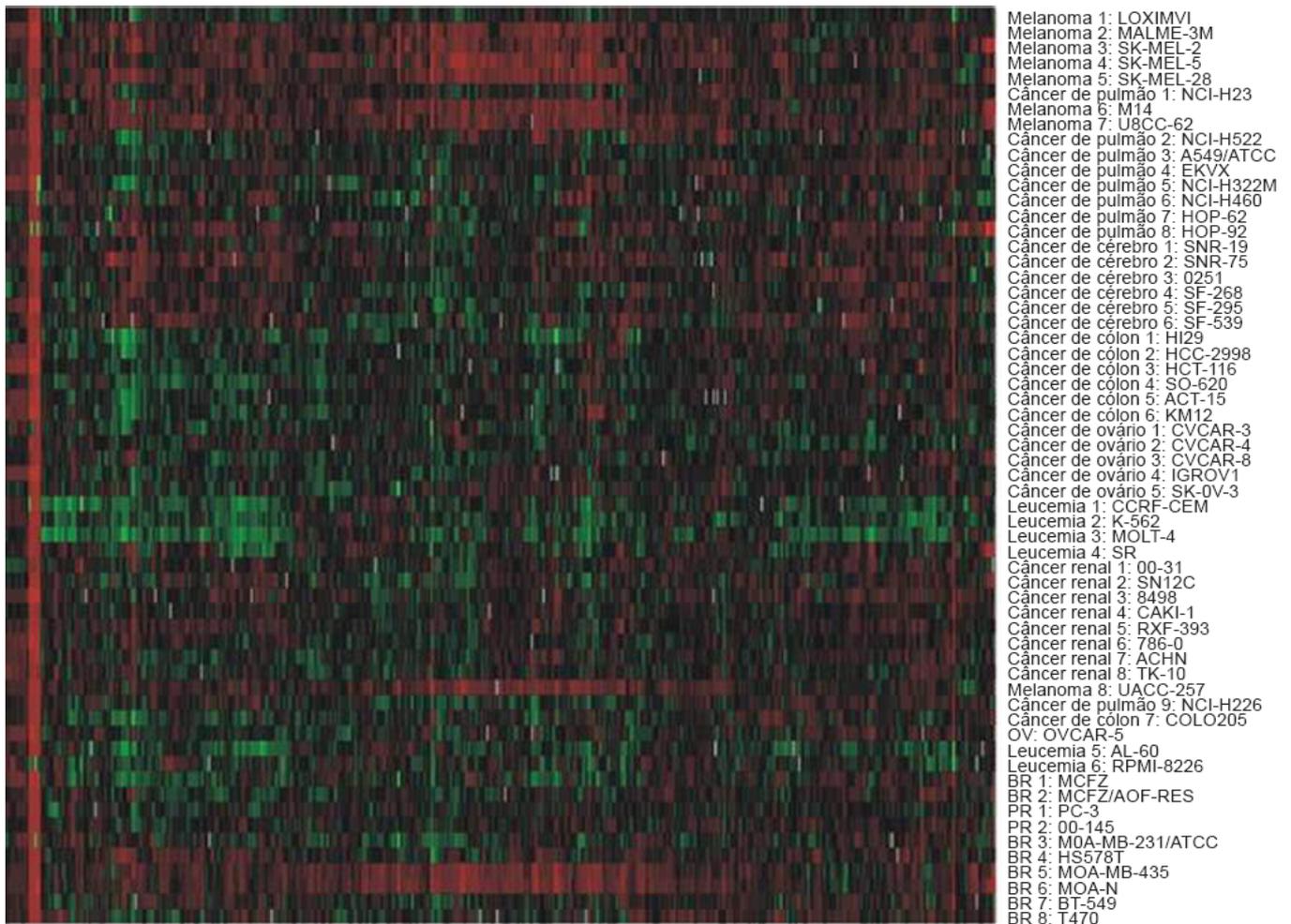
A técnica de *Northern blot* foi desenvolvida na década de 1970, e utiliza uma única sonda para realizar a detecção dos níveis de um único RNA em diversas amostras. Isso ocorre porque o mRNA é imobilizado em um suporte sólido que é incubado com uma solução contendo a sonda correspondente ao transcrito específico de interesse. Em geral, a sonda é marcada com isótopos radioativos ou moléculas fluorescentes, o que possibilita sua detecção. Por meio da medida da quantidade de sonda que se hibrida aos RNA imobilizados, é possível inferir a abundância de moléculas do transcrito específico na amostra. Esse tipo de experimento é muito útil quando se deseja estudar um RNA específico que esteja envolvido em um fenômeno biológico.

Uma grande vantagem dos microarranjos de DNA em relação ao *Northern blot* é que eles possibilitam monitorar ao mesmo tempo os níveis de milhares de moléculas de DNA a um custo relativamente baixo e com muito menos trabalho que a realização de múltiplos experimentos de *Northern blot*. Isso torna tal técnica popular para abordagem de problemas em que se busca propor um novo mecanismo específico subjacente a fenômenos celulares. Em um único microarranjo, é possível imobilizar dezenas de milhares de sondas, não existindo a necessidade de eleger um

número limitado de transcritos para estudo. Isso ocorre porque, com as tecnologias atuais, é possível imobilizar em uma única lâmina o número equivalente de sondas ou superior ao número de transcritos distintos de um organismo, sendo possível representar praticamente todo o transcriptoma. Uma vez sintetizadas as sondas, é possível construir centenas de lâminas que podem ser utilizados em diversos experimentos, possibilitando comparações entre diferentes amostras de modo rápido e prático. Um exemplo de comparação interessante que poderia ser realizada com esse tipo de plataforma é a comparação entre amostras de diversos pacientes em um estudo clínico, ou em um estudo temporal de efeito de uma substância em um conjunto de células. Em função da dinâmica dos transcriptomas, é esperado que cada amostra apresente um perfil diferente das outras, o que pode indicar efeitos de uma substância sobre a expressão de um determinado gene, ou até mesmo diferenças de respostas entre indivíduos (Figura 13.10). Uma dificuldade causada pela grande quantidade de dados gerados pelos experimentos de microarranjos é que o processamento e a interpretação dos dados não são triviais, exigindo testes estatísticos para verificar quais transcritos apresentam, de fato, expressão diferencial. Esse tipo de necessidade evidencia o papel cada vez mais importante que a bioinformática tem para a análise de conjunto de dados biológicos em larga escala.

### **Detecção e quantificação de RNA mensageiro com a utilização da reação em cadeia da polimerase (PCR)**

Outro tipo de técnica utilizada para o estudo do transcriptoma é o chamado PCR em tempo real. Ela se baseia na amplificação de cDNA pela reação em cadeia da polimerase, sendo a amplificação de cada transcrito monitorada por meio de moléculas fluorescentes que passam a emitir maior quantidade de luz à medida que uma maior quantidade de cDNA é amplificada. Como nos estágios iniciais da reação de PCR a amplificação é exponencial, é possível inferir a quantidade de cDNA original a partir do número de ciclos de PCR necessários para a amostra apresentar uma concentração preestabelecida de DNA. Quanto menos ciclos forem necessários para a amostra chegar a essa concentração, maior a concentração inicial de cDNA. Esse tipo de técnica tem a vantagem de exigir uma baixa quantidade de cDNA, pois se baseia em uma técnica de amplificação. Além disso, por utilizar oligonucleotídeos específicos como iniciadores da reação de PCR, a amplificação de cDNA é bastante específica, sendo geralmente amplificado somente o DNA correspondente ao transcrito de interesse. No entanto, pelo fato de podermos realizar somente a análise de um transcrito em cada tubo, apenas um pequeno conjunto de genes pode ser monitorado por experimento.



**Figura 13.10** Exemplo de visualização de uma análise de dados de microarranjos. No gráfico, cada linha representa os resultados do experimento de um paciente diferente, enquanto cada coluna corresponde ao resultado da medida do nível de expressão de uma sonda específica que corresponde a um transcrito. A coloração indica uma expressão aumentada (vermelho) ou diminuída (verde) para aquele gene com relação à média dos pacientes.

## Proteoma | Conjunto de moléculas efetoras da maioria dos processos celulares

As proteínas representam um grupo extremamente versátil de biomoléculas. Um dos motivos é que os 21 aminoácidos que podem compor uma proteína apresentam cadeias laterais com propriedades físico-químicas bastante distintas. Portanto, diferentes combinações de aminoácidos irão gerar proteínas com as mais diferentes conformações e propriedades. Essa característica das proteínas se traduz no fato de que elas são os biopolímeros mais versáteis dos organismos vivos, podendo atuar em diversos contextos celulares, como moléculas estruturais, enzimas, moléculas sinalizadoras, entre outras funções. Além disso, são encontradas nos diferentes ambientes celulares, podendo estar no citoplasma, associadas a membranas, no núcleo, associadas ao DNA, entre outros.

As proteínas são os grandes efetores dos diversos processos celulares e, por isso, o seu estudo é de grande interesse para os pesquisadores. Elas são essenciais para a criação e a manutenção das outras macromoléculas, sendo responsáveis pela replicação do DNA, a transcrição do RNA e a síntese de carboidratos e lipídeos. Assim, a descrição da síntese e degradação dessas diversas biomoléculas passa por uma verificação da abundância e regulação das proteínas que compõem essas rotas metabólicas.

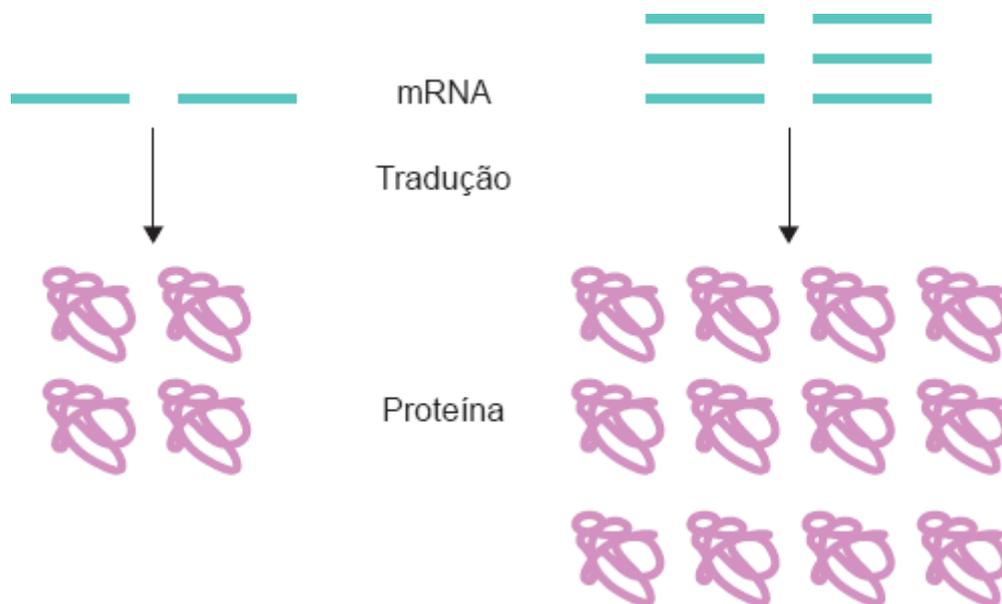
Além disso, são sintetizadas a partir da informação contida na molécula de mRNA, que é traduzido pelo ribossomo das células para a síntese de uma proteína específica. Cada mRNA que se encontra no citoplasma é traduzido por múltiplos ribossomos que migram em uma direção específica da molécula, transformando a informação do RNA em polímeros de aminoácidos. Continuamente, há a ligação de novos ribossomos à molécula de RNA, fazendo com que cada molécula esteja ligada a múltiplos ribossomos em diferentes estágios da tradução. Considerando esse cenário, quanto maior o número de moléculas de um mRNA codificado para uma proteína específica, maior é o número de ribossomos recrutados para a síntese dessa proteína e, portanto, resultará em sua

maior concentração final (Figura 13.11). Assim, pode-se estimar a abundância de proteínas utilizando a quantidade de mRNA como medida indireta.

No entanto, isso não leva em consideração o fato de que existem mecanismos que podem estimular a síntese proteica a partir de um mRNA, bem como mecanismos específicos de degradação de proteínas, que, certamente, podem alterar a correlação entre abundância de proteínas e correspondentes mRNA. Consequentemente, a medida direta dos níveis das diferentes proteínas em uma célula ou tecido é muito importante para avaliação completa dos sistemas moleculares atuantes.

Ademais, em muitos casos, a atividade de uma proteína pode ser alterada por modificações pós-traducionais. Tais modificações consistem na ligação de grupos químicos à cadeia de proteína após sua síntese, como a glicosilação e a fosforilação de proteínas. A glicosilação é um evento que geralmente ocorre em proteínas secretadas que são direcionadas ao retículo endoplasmático e complexo de Golgi. À proteína são adicionados diversos resíduos de açúcar que podem alterar significativamente seu peso molecular. A fosforilação, por outro lado, é uma modificação bem mais simples, que consiste na adição de um grupo fosfato em uma proteína. Apesar de ser um grupo pequeno e que, a princípio, não deveria alterar muito a estrutura da proteína, fosforilações são utilizadas como sinais para ativar/desativar uma série de proteínas envolvidas em catálise ou sinalização.

As análises do proteoma de uma célula ou tecido devem considerar todas as peculiaridades das proteínas. Logo, não se trata de um mero esforço de detectá-las, mas inclui também identificar compartimentos celulares nos quais se encontram essas proteínas e detectar eventuais alterações pós-traducionais em seus subconjuntos.



**Figura 13.11** Esquema que indica a variação de quantidade de mRNA (barras horizontais) em duas situações diferentes. Em razão desse aumento, ocorre um crescimento proporcional no número de proteínas produzidas a partir da tradução dessas mensagens.

## Métodos para estudar o proteoma

### Limitações no estudo de proteomas

O estudo do transcriptoma avançou muito nas últimas décadas, em função do avanço nas técnicas de sequenciamento e hibridação. No entanto, é necessário lembrar que a principal função do mRNA é codificar proteínas que serão produzidas com o auxílio do ribossomo. Assim, a presença de determinados níveis de mRNA no citoplasma não acarretaria diretamente em nenhum processo celular específico, e somente a partir do momento em que uma proteína é produzida com base neste mRNA é que, de fato, ocorrerão mudanças no ambiente celular.

Pode-se perguntar, portanto, por que existem tantos estudos buscando quantificar as moléculas de mRNA, para utilizá-las como uma medida indireta de seus níveis, em vez de realizar a medida direta de seus níveis. Isso ocorre porque as moléculas de mRNA são facilmente convertidas em cDNA e podem ser, então, facilmente amplificadas e manipuladas por meio de técnicas de biologia molecular. Isso possibilita que mesmo porções diminutas de mRNA possam ser amplificadas e depois analisadas. O sequenciamento de DNA é uma técnica consistente e facilmente

utilizável em larga escala, viabilizando a análise simultânea de milhares de moléculas de mRNA e a detecção até mesmo das moléculas mais raras.

Em contraste, técnicas de manipulação de proteínas são relativamente limitadas, não existindo, por exemplo, técnicas que possibilitem sua amplificação. Isso ocasiona uma grande dependência da quantidade e qualidade do material biológico a partir dos quais serão extraídas as proteínas. Como consequência, o detalhamento com o qual se consegue estudar proteomas é geralmente menor, muitas vezes restringindo-se às proteínas mais abundantes do sistema estudado, ou a um subconjunto de proteínas selecionadas com base em alguma característica desejada.

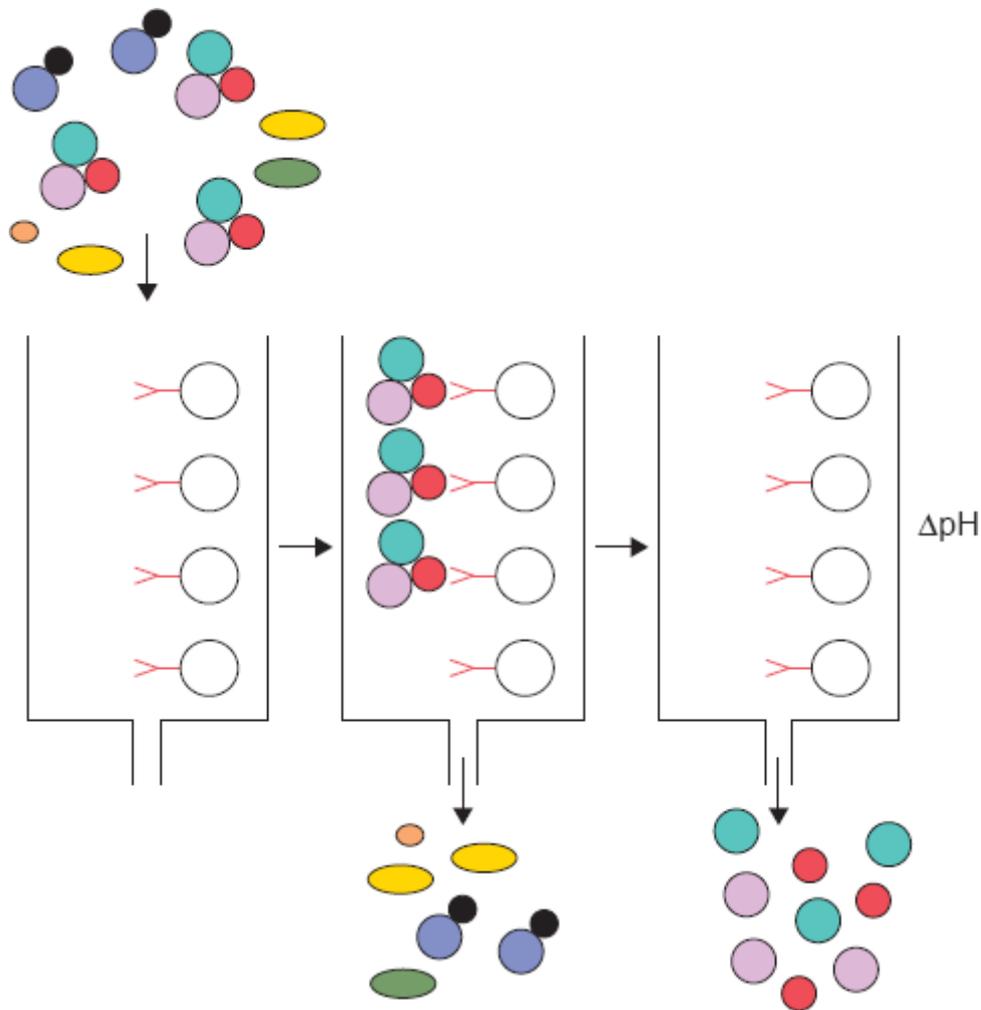
Além disso, devido à grande heterogeneidade do comportamento das proteínas, muitas vezes é difícil estabelecer protocolos universais para extração e processamento de proteomas. Isso faz com que seja necessário desenvolver protocolos bastante específicos para o estudo de proteomas de determinados organismos, ou para estudo de certos subconjuntos do proteoma. Tal fato é bastante contrastante com o estudo de transcriptomas, cujo maior desafio é a obtenção de RNA puro e íntegro, sendo os protocolos posteriores praticamente universais. Assim, o estudo de proteomas passa a exigir um preparo muito mais sofisticado dos experimentos, além de dificultar o estabelecimento de rotinas de preparo de amostras que possibilitem iniciativas em larga escala de exploração de proteomas de diversos organismos.

### **Fracionamento de amostras biológicas**

Em razão das limitações da técnica proteômica, em muitas ocasiões é necessário um processo inicial de fracionamento de amostras que possibilite um direcionamento a proteínas de interesse. Uma técnica bastante utilizada é o fracionamento celular por meio de ultracentrifugação. Essa separação se baseia no princípio de que, uma vez que a célula seja rompida e homogeneizada em procedimentos controlados, ela irá liberar suas organelas, que podem então, ser fracionadas. Tal fracionamento é realizado por centrifugação a altas velocidades (que geram forças da ordem de centenas de milhares de vezes à força da gravidade), aproveitando-se das diferentes taxas de sedimentação entre as organelas. Organelas com maior tamanho e densidade tendem a se sedimentar com maior facilidade em comparação àquelas de baixo tamanho e densidade. Assim, partículas maiores, como o núcleo, tendem a se sedimentar rapidamente quando submetidas a centrifugações a velocidades que geram forças de cerca de 1.000 vezes a gravidade, enquanto as outras organelas sedimentarão muito lentamente a essa velocidade e, na prática, continuarão em solução. Por meio de sucessivas centrifugações a diferentes velocidades, é possível separar diversos componentes celulares, viabilizando, assim, o estudo individualizado das proteínas de cada uma dessas frações.

### **Isolamento de complexos proteicos**

Existem outras técnicas que permitem restringir um subconjunto do proteoma para análises. Um exemplo é a técnica de coimunopurificação, que consiste em utilizar um anticorpo produzido contra uma proteína específica e imobilizá-lo em uma coluna ou resina. Anticorpos são proteínas produzidas pelo sistema imune que têm a capacidade de reconhecer de modo bastante específico regiões de proteínas. Uma vez que o anticorpo se liga a uma proteína, forma-se uma interação bastante estável. Quando o anticorpo é incubado com o extrato proteico, ele irá se ligar à proteína-alvo e, por consequência, irá imobilizá-lo também na resina. Juntamente com essa proteína, serão imobilizadas todas as outras que estejam interagindo diretamente ou façam parte do mesmo complexo proteico (Figura 13.12). Uma vez que o complexo é imobilizado na resina, é fácil retirá-lo da solução, separando-o, assim, das outras proteínas do organismo.



**Figura 13.12** Representação esquemática de imunopurificação. Anticorpos reagentes com a proteína representada em vermelho são imobilizados em uma coluna. Após a passagem de uma solução contendo diversas proteínas (representadas por círculos), somente a proteína representada em vermelho se ligará ao anticorpo e será retida na coluna. Outras proteínas interagindo com essa proteína também serão retidas. Posteriormente, uma solução com pH muito baixo ou alto é adicionada para desestabilizar a ligação proteína-anticorpo e possibilitar a recuperação das proteínas do complexo.

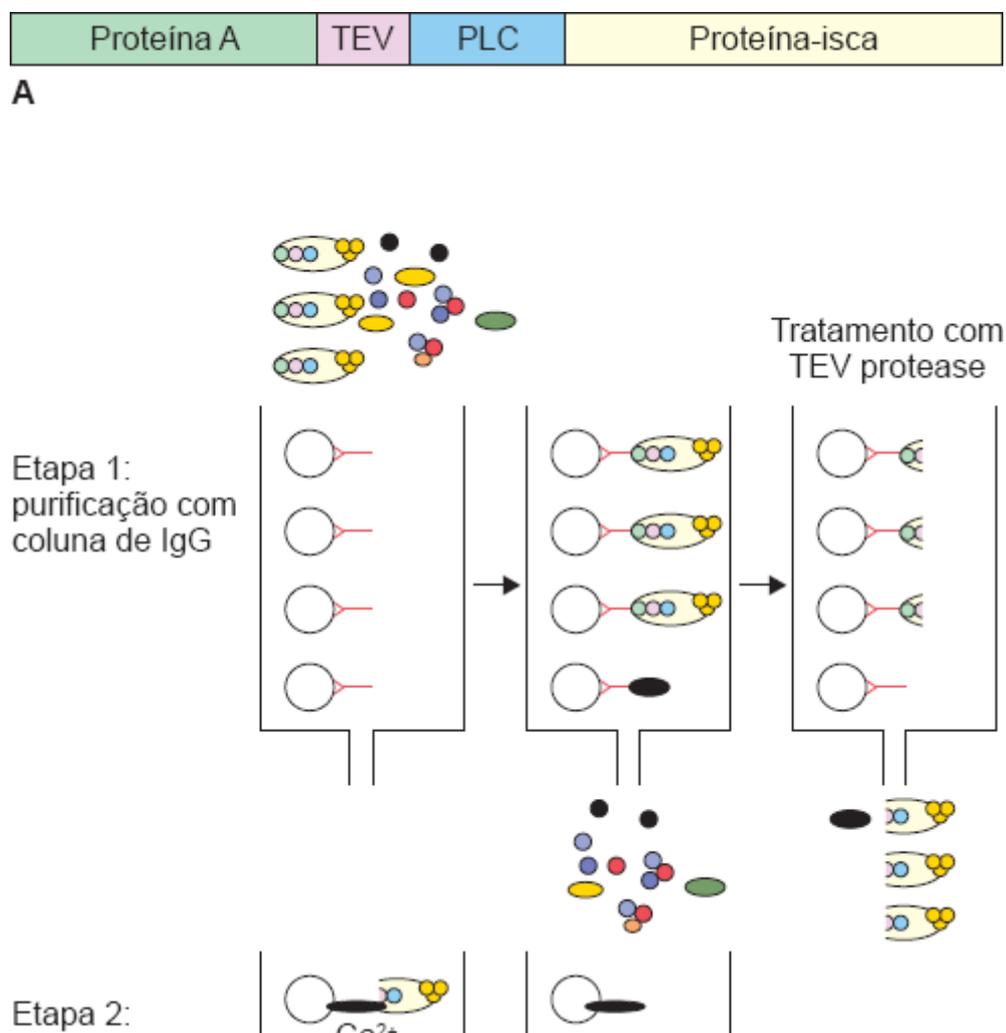
Uma técnica alternativa à utilização de anticorpos para estudo de complexos proteicos é a de purificação por afinidade em tandem (TAP), que purifica o complexo proteico a partir de duas sucessivas purificações com base na afinidade de peptídeos que são adicionados à cadeia de uma das proteínas do complexo. Nesse tipo de técnica, é selecionada uma proteína “isca”, com a qual se supõe que outras proteínas interajam. Uma cópia do gene codificando para essa isca é preparada contendo, em fusão, duas regiões codificando duas porções derivadas de diferentes proteínas. Em geral, essas duas porções codificam um peptídeo ligante de calmodulina e uma porção ligante de IgG da proteína A de *Staphylococcus aureus*. A calmodulina é uma proteína que se liga a outras em um processo dependente de íons cálcio. A proteína A tem a capacidade de se ligar à região invariável de anticorpos do tipo IgG. Em ambos os casos, a interação dessas proteínas com seus ligantes é bastante específica e, por isso, elas são utilizadas para a purificação a partir de afinidade. Entre as duas porções codificantes para esses peptídeos existe ainda um sítio de clivagem da TEV protease (Figura 13.13 A).

Assim, nesses estudos, o gene em análise é adicionado ao organismo de interesse, o qual passará a expressar uma cópia adicional da proteína do complexo proteico. A proteína codificada por essa nova cópia terá uma porção adicional contendo os peptídeos em fusão. Após a expressão desse gene, a proteína resultante passará a formar o complexo com as outras proteínas, do mesmo modo que a proteína nativa. A purificação dos complexos proteicos é realizada pela passagem do produto de lise celular em uma coluna contendo IgG imobilizado, na qual a porção da proteína A irá se ligar e, com isso, imobilizar o complexo proteico na coluna. Após a lavagem para retirada das proteínas que não interagem com a coluna, é realizada uma digestão com a TEV protease, que separará a porção contendo a porção da proteína A do restante da proteína, e fazendo com que o complexo proteico se desligue da coluna. Um segundo passo de purificação é realizado utilizando uma coluna contendo calmodulina imobilizada. Na

presença de cálcio, ocorre a ligação entre a calmodulina imobilizada e o peptídeo ligante de calmodulina existente na proteína do complexo. Isso possibilita um segundo passo que garantirá melhor purificação do complexo, o qual, após a lavagem, pode ser desligado da coluna com a retirada do cálcio tampão (Figura 13.13 B). A realização dessas duas purificações permite que o produto final seja o complexo proteico com alto grau de pureza, o que viabiliza análises posteriores para identificação de suas proteínas componentes. Existem casos em que os fragmentos de proteína A e peptídeo ligador de calmodulina são substituídos por outras proteínas que apresentam afinidade por outros compostos, mas o princípio de purificação continua o mesmo.

### Técnicas de separação de proteínas e peptídeos

Diversas estratégias empregadas para separação de proteínas e peptídeos em experimentos de proteômica utilizam técnicas eletroforéticas ou cromatográficas. É necessário notar que, ao trabalhar com identificação de proteínas, em algum ponto do processo deve ocorrer a sua digestão com o auxílio de proteases para a formação de peptídeos. Isso é necessário porque a análise de peptídeos por meio da técnica de espectrometria de massa é mais informativa que a análise de proteínas inteiras. Portanto, dentro de uma estratégia para caracterização de proteomas, podemos encontrar, muitas vezes, passos de separação de proteínas, seguidas de uma proteólise e passos seguintes de separação e purificação de peptídeos.

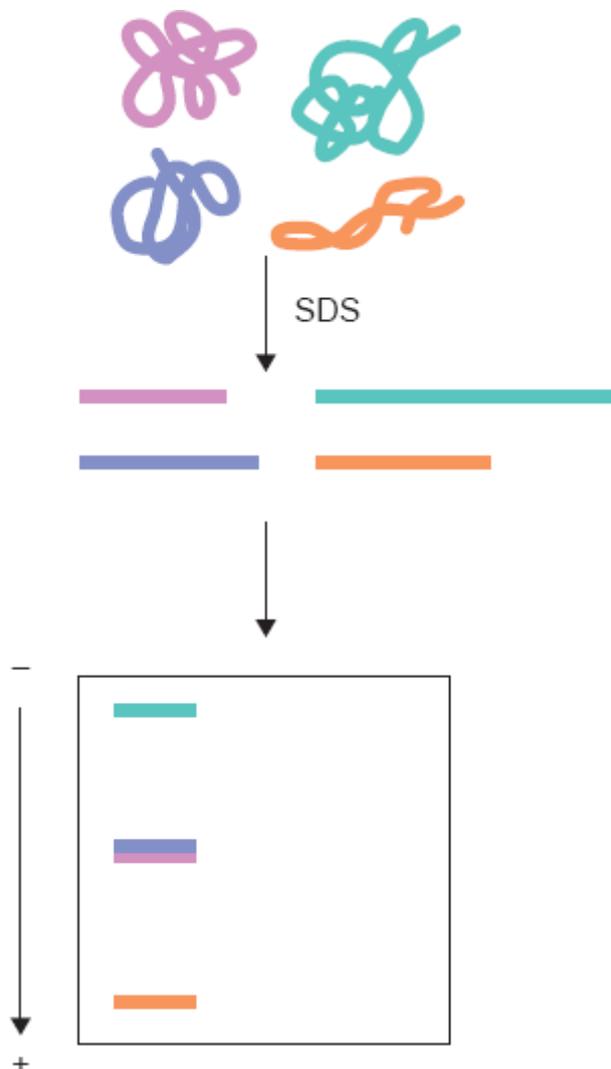


**Figura 13.13** Representação esquemática de uma purificação por afinidade em tandem (TAP). **A.** Representação das diferentes regiões da proteína de fusão incluindo regiões codificando para proteína A, sítio de clivagem da TEV protease (TEV) e peptídeo ligante de calmodulina (PLC). **B.** Representação dos diferentes passos de purificação do complexo de interação envolvendo a proteína-isca.

A separação de proteínas geralmente é realizada por meio de técnicas eletroforéticas. Um tipo de técnica clássica para separação de proteínas é através de eletroforese de géis de poliacrilamida contendo o detergente sódio dodecilsulfato (SDS). Esse tipo de técnica possibilita separar proteínas com base em seu peso molecular, pois o SDS é um detergente aniônico que irá se intercalar a uma proteína linearizando-a e dando uma carga negativa

relativamente uniforme a todas as proteínas. Assim, todas as moléculas de proteínas passam a ter aproximadamente a mesma carga e formato, e, quando submetidas a um campo elétrico, o principal parâmetro que irá determinar a sua velocidade de migração é a sua massa molecular (Figura 13.14). Apesar de ser uma técnica bastante consistente, esse tipo de separação não é muito eficiente quando temos uma mistura complexa de proteínas. Isso ocorre porque, nesse tipo de mistura, são esperadas várias proteínas com massa molecular semelhante e que irão migrar até praticamente a mesma região do gel.

Outra técnica interessante para a separação de proteínas é a chamada focalização isoelétrica. As cadeias laterais das proteínas apresentam grupos capazes de serem ionizados, e a carga desses grupos irá depender do pH onde eles se encontram. A carga de uma proteína em certo pH será a somatória das cargas apresentadas pelas suas cadeias laterais. Esse tipo de técnica aproveita-se das diferentes cargas apresentadas por proteínas em um pH, e consiste em submeter as proteínas a um campo elétrico em um gradiente imobilizado de pH. Quando realizada uma focalização, proteínas com uma carga total negativa serão atraídas para o ânodo; no entanto, à medida que migram, elas irão para regiões com menor pH e, portanto, tenderão a adquirir prótons que irão neutralizar as suas cargas negativas. A migração continua até o ponto em que a protonação da proteína neutraliza as cargas negativas e a proteína passa a ser neutra, não respondendo mais ao campo elétrico. O oposto ocorre com uma proteína com carga positiva, que irá migrar para o cátodo até que seja suficientemente desprotonada para se tornar neutra; a região do gradiente de pH em que cada proteína irá se concentrar é correspondente ao seu ponto isoelétrico (pI).



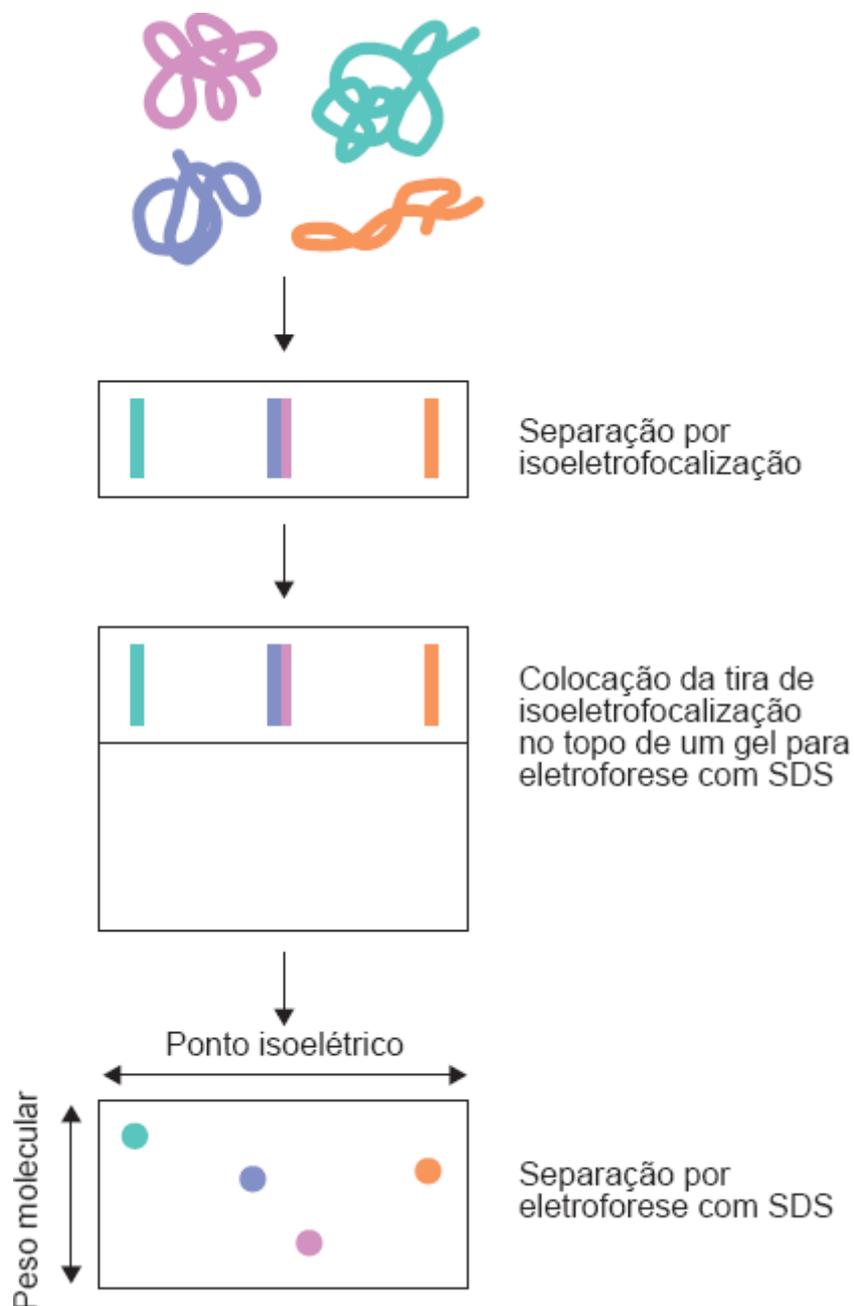
**Figura 13.14** Representação esquemática de separação de proteínas com gel de poliacrilamida com SDS. Primeiramente, as proteínas enoveladas são linearizadas com a exposição ao SDS e, depois, são submetidas à eletroforese em gel de acrilamida. Durante a separação, as proteínas migrarão do polo positivo para o negativo, e as proteínas de menor peso molecular migrarão mais rapidamente. Nota-se que as proteínas roxas e azuis têm peso molecular semelhante e não são bem separadas por esse tipo de técnica.

A combinação das metodologias de isoeletrofocalização e eletroforese em gel de acrilamida com SDS define a técnica de eletroforese em bidimensional. Tal técnica foi desenvolvida na década de 1970 pelo pesquisador Patrick

O'Farrell, cujo objetivo era estabelecer um sistema de separação que caracterizasse mudanças no padrão de expressão de proteínas causada por mutações que afetavam o desenvolvimento de algas do gênero *Volvox*. Em primeiro lugar, as proteínas são separadas por isoeletrofocalização em uma fita contendo um gradiente de pH e, posteriormente, essa fita é colocada no topo de um gel de poliacrilamida contendo SDS e submetida à eletroforese (Figura 13.15). Isso favorece que as proteínas se distribuam nesse gel de acordo com o seu pI no eixo horizontal e pelo seu peso molecular no eixo vertical. Como resultado, elas serão representadas por pontos no gel após a sua coloração com reagentes específicos para detecção de proteínas, sendo o posicionamento desses pontos reflexo de tais aspectos (Figura 13.15).

As proteínas nativas são submetidas a um processo de isoeletrofocalização em um gradiente imobilizado de pH. As diferentes proteínas irão se concentrar em faixas de pH correspondentes ao seu pI. Após a focalização isoeletrica, o suporte sólido contendo as proteínas é colocado no topo de um gel de poliacrilamida com SDS, e as proteínas são desenvolvidas e passam a migrar de acordo com o seu peso molecular. No gel final, elas serão separadas pelo seu ponto isoeletrico no eixo horizontal e pelo seu peso molecular no eixo vertical. Proteínas com pontos isoeletricos semelhantes (vermelha e amarela) não são bem separadas na primeira etapa; no entanto, pela diferença de peso molecular, acabam sendo separadas na segunda. As proteínas rosa e azul que não eram bem separadas na Figura 13.14 também passam a ter uma boa separação.

Aquelas separadas pela técnica de eletroforese bi-dimensional podem então ser individualmente analisadas para estabelecer as suas identidades. Com essa finalidade, as regiões correspondentes a cada ponto do gel são recortadas e passam por um processo que irá resultar na digestão de proteína em peptídeos. Essa digestão é realizada por proteases que contêm regiões específicas de corte, o que possibilita prever os peptídeos que serão formados a partir da digestão de cada proteína do organismo. A protease mais utilizada em experimentos de proteômica é a tripsina, que realiza cortes nas ligações peptídicas do grupo carboxila dos aminoácidos lisina ou arginina com o grupamento amino de qualquer outro aminoácido. Essa digestão é realizada nos próprios fragmentos de gel recortados, que são desidratados e depois reidratados com uma solução contendo a protease. Uma vez que a digestão é completada, os peptídeos formados não são mais retidos pela malha do gel, devido ao seu pequeno tamanho, e se difundem na solução na qual o gel se encontra mergulhado.



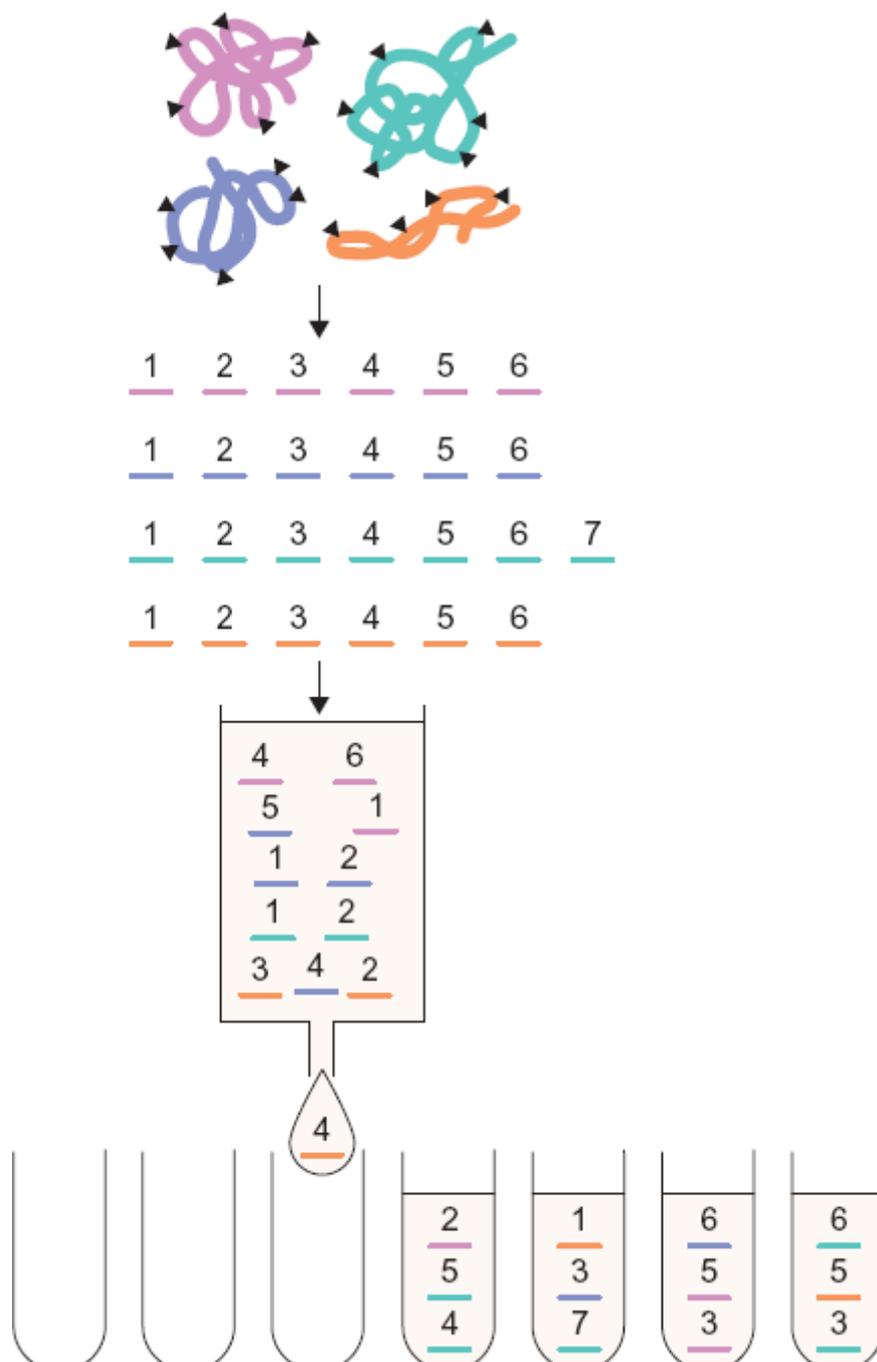
**Figura 13.15** Representação esquemática da separação de proteínas por meio da técnica de eletroforese bidimensional.

Além disso, todas as ligações dissulfeto realizadas pela proteína por meio de seus resíduos de cisteína são reduzidas, e grupos químicos são ligados aos grupos SH livres formados para impedir o restabelecimento das ligações dissulfeto. Esse tipo de tratamento é realizado porque, após a tripsinização, os peptídeos formados que contenham pontes dissulfeto entre si estarão ligados covalentemente. Logo, para posterior análise e identificação desses peptídeos, seria necessário prever essas ligações por pontes dissulfeto, o que não é trivial, complicando a análise dos peptídeos.

Em adição aos métodos de separação de proteínas por eletroforese bidimensional, é possível realizar estudos de proteômica através de métodos de separação cromatográfica. Esse tipo de abordagem geralmente tem enfoque na separação de peptídeos após a digestão de proteínas. Inicialmente, uma solução contendo proteínas é digerida com tripsina ou outra protease, e será formada uma mistura complexa contendo peptídeos provenientes de todas as proteínas que se encontravam em solução. Essa mistura de peptídeos é então separada a partir de técnicas cromatográficas, com o objetivo de criar frações com uma complexidade menor que a mistura inicial (Figura 13.16). Proteínas de uma mistura são digeridas com tripsina (pontos de digestão indicado por setas). Os diferentes peptídeos resultantes são representados por barras com coloração idêntica à proteína de origem e numeradas para possibilitar a distinção entre os peptídeos gerados. Uma separação desses peptídeos em uma coluna cromatográfica é representada

com peptídeos separados em diferentes frações, de acordo com suas propriedades de interação com a coluna e o solvente, e alguns peptídeos ainda retidos na coluna neste momento do processo.

No caso de amostras bastante complexas, como a fração total de proteínas solúveis de uma célula, mais de um tipo de separação cromatográfica é normalmente utilizado no processo. A abordagem mais comum é realizar uma primeira separação em uma coluna de troca catiônica, onde os peptídeos serão separados com base em sua afinidade por uma coluna negativamente carregada, e as frações formadas a partir dessa coluna serão, em seguida, separadas por uma coluna hidrofóbica, que irá separar peptídeos com base na sua afinidade por uma coluna que tem uma cadeia hidrocarbônica alifática. Cada fração da coluna original irá ser subfracionada, criando um grande número de amostras, cada uma contendo um número limitado de peptídeos que apresentam características semelhantes em termos de interação com grupos carregados negativamente e com grupos hidrofóbicos. A divisão da amostra em várias frações é importante, pois, ao analisar uma mistura complexa em um espectrômetro de massa, somente um número limitado de peptídeos será detectado por amostra, já que eles competem no passo inicial de ionização. Ao subdividir-se a amostra, ocorre a detecção de um número muito maior de peptídeos, pois cada fração será menos complexa e, portanto, o número de peptídeos competindo pela ionização será menor.



**Figura 13.16** Esquema de separação de peptídeos por meio de técnicas cromatográficas.

Em vários casos, utiliza-se uma combinação de técnicas eletroforéticas e cromatográficas, a fim de obter melhor separação de peptídeos para análise com espectrometria de massa. Um exemplo é a metodologia chamada Gel-LC, que se inicia com uma separação de proteínas por meio de eletroforese em gel de poliacrilamida com SDS, que irá separar as proteínas por peso molecular. Após isso, o gel de poliacrilamida é recortado, separando-se as diversas faixas de tamanho, e cada um dos recortes é processado separadamente para obtenção dos fragmentos trípticos das proteínas contidas em cada fração. Cada uma das soluções contendo os peptídeos resultantes é submetida a um processo cromatográfico que irá separar os peptídeos componentes em frações que serão analisadas por espectrometria de massa.

## Utilização de espectrometria de massa para estudo de proteomas

A análise de peptídeos em experimentos de proteômica é realizada utilizando a técnica de espectrometria de massa (ver o box “Desenvolvimento da espectrometria de massa”). Esse tipo de técnica possibilita medidas muito precisas da massa molecular de moléculas e pode ser utilizada para determinar a composição de polímeros biológicos, incluindo peptídeos e proteínas. Trata-se de uma técnica com alta sensibilidade, o que é muito importante quando trabalha-se com moléculas biológicas, pois a quantidade de uma proteína extraída de material biológico costuma ser limitada.

### Desenvolvimento da espectrometria de massa

Espectrômetros de massa são aparelhos que utilizam campos eletromagnéticos para determinar de modo extremamente preciso a massa de íons produzidos a partir da ionização de uma amostra. É reconhecido que o desenvolvimento desse tipo de aparelho iniciou-se com os experimentos de J. J. Thomson, no início do século 20, que realizou a deflexão de gases ionizados em câmeras de vácuo, utilizando campos eletromagnéticos, e demonstrou que essa deflexão dependia da relação massa/carga do íon, estabelecendo, assim, os fundamentos para o desenvolvimento desse tipo de aparelho. Durante as três primeiras décadas do século 20, Francis Ashton (que havia sido aluno de Thomson) e outros cientistas aperfeiçoaram esse aparato inicial, melhorando a sua resolução, e forneceram evidência da existência de isótopos para uma grande parte dos elementos químicos. A partir da década de 1940, os espectrômetros de massa passaram a se popularizar, e aparelhos comerciais ficaram disponíveis e foram utilizados não somente por físicos, mas por químicos interessados em determinar a composição química de amostras. Somente na década de 1980, com o desenvolvimento de técnicas suaves de ionização, foi possível utilizar a espectrometria de massa para estudo de biomoléculas.

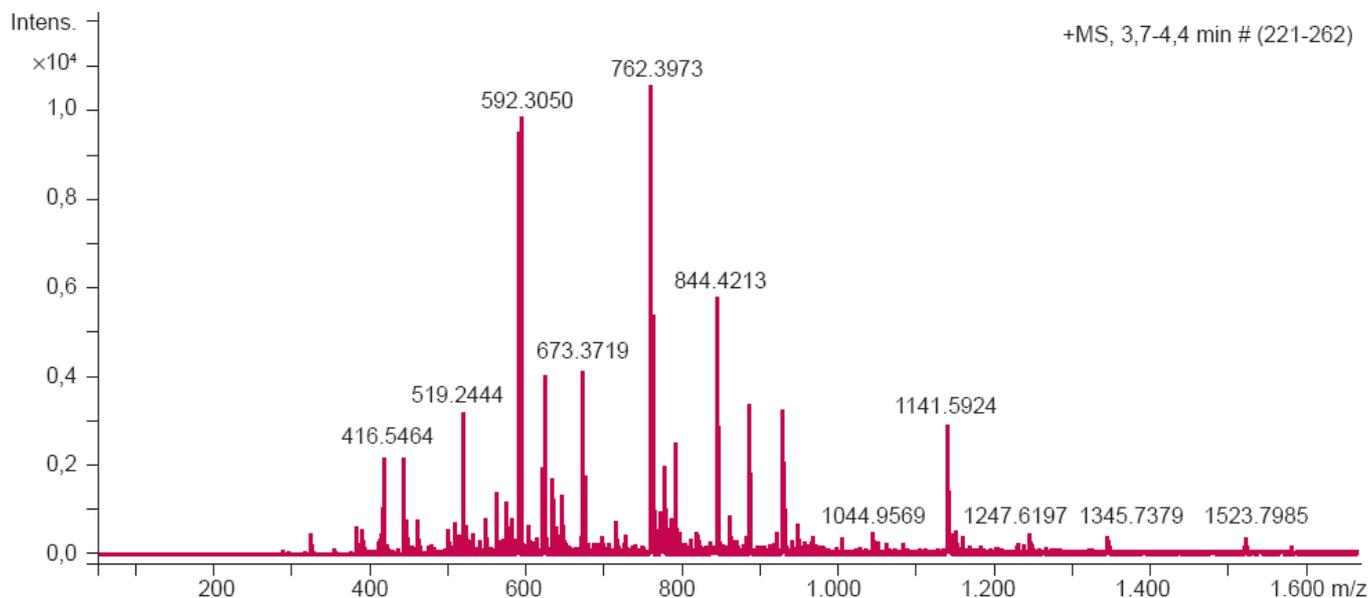
Pelo menos cinco cientistas que desenvolveram técnicas críticas para o desenvolvimento e aperfeiçoamento da espectrometria de massa foram ganhadores do Prêmio Nobel: Joseph John Thomson, Prêmio Nobel de Física de 1906; Francis William Aston, Prêmio Nobel de Química de 1922; Wolfgang Paul, Prêmio Nobel de Física de 1989; John Bennet Fenn e Koichi Tanaka, Prêmio Nobel de Química de 2002.

O primeiro passo necessário para a análise de peptídeos é a sua ionização. Isso é necessário porque o espectrômetro de massa trabalha com campos elétricos e, portanto, moléculas sem carga são invisíveis para esses aparelhos. Além disso, as moléculas devem passar a fase gasosa para que possam ser analisadas pelo aparelho. Existem diversas técnicas para ionizar moléculas, mas as mais utilizadas para análise de peptídeos são o *electrospray* e MALDI (ionização/dessorção de matriz assistida por *laser*), que, por serem técnicas suaves de ionização, preservam a estrutura de biomoléculas.

A técnica de ionização por *electrospray* se baseia na passagem de líquido por uma agulha que esteja carregada eletricamente. Isso produz uma série de gotículas com cargas. Tais gotículas passarão por uma região com gás quente que levará à evaporação do solvente e, como resultado, teremos os peptídeos de interesse ionizados e em fase gasosa. Já a técnica de ionização por MALDI consiste na mistura da substância a ser analisada a uma matriz que tem a propriedade de absorver energia luminosa. Quando um *laser* é aplicado a essa matriz, ela rapidamente absorve energia, ionizando a amostra e ejetando-a para a fase gasosa.

As amostras de peptídeos ionizadas serão objeto de análise por aparatos que possibilitarão medir a sua relação de massa molecular/carga com grande exatidão. Existem diversos tipos de analisadores que farão essa medida, cada um com suas características específicas. Independentemente do analisador, o resultado final obtido será um gráfico contendo, no eixo X, as diferentes relações massa molecular/carga e, no eixo Y, a intensidade de detecção (Figura 13.17). Há uma série de metodologias que possibilitam que a carga do peptídeo detectado seja determinada, sendo possível converter esse gráfico em um que represente apenas as massas moleculares dos peptídeos detectados em seu eixo X.

Esse tipo de dado pode ajudar na identificação de uma proteína, pois cada pico irá representar um peptídeo diferente. Se a amostra injetada contiver peptídeos de uma única proteína (como no caso de um ponto recortado de um gel resultante de uma eletroforese bidimensional), pode-se considerar que esse espectro é uma espécie de impressão digital da proteína. A partir dos bancos de sequências de proteínas, são realizadas digestões virtuais com a enzima proteolítica utilizada no experimento, o que torna possível prever os peptídeos formados por cada proteína. São então deduzidos os espectros teóricos para cada proteína do banco de dados, e estes são comparados com o espectro experimental, para determinação do espectro com maior taxa de coincidência (Figura 13.18).



**Figura 13.17** Espectro de massa de uma amostra contendo peptídeos. O eixo  $x$  indica a relação massa/carga das substâncias detectadas ( $m/z$ ); o eixo  $y$  indica a intensidade de cada pico. Os números acima dos picos indicam a medida exata de  $m/z$  de alguns dos picos detectados.

Diversas sequências de um banco de dados são analisadas, possibilitando a construção de um espectro de massa teórico para cada um deles. A comparação entre o espectro experimental e os teóricos (linha pontilhada) torna possível verificar coincidências e, a partir delas, encontrar o espectro que forneça uma identificação confiável.

Não se espera que haja 100% de coincidência entre os picos do espectro teórico e o experimental de uma proteína, pois, durante o passo de ionização, nem todos os peptídeos são efetivamente ionizados e, portanto, alguns serão perdidos durante o processo. Logo, o processo de comparação entre espectros teóricos e experimentais deve ser auxiliado por uma análise estatística que indique o grau de confiança da identificação de uma proteína com base no número de picos coincidentes entre os dois espectros.

Esse tipo de abordagem apresenta algumas limitações, visto que, dependendo do número de picos de peptídeos detectados e do tamanho do banco de dados de proteínas utilizado, podem ocorrer diversas coincidências entre espectros teóricos de proteínas não relacionadas e o espectro experimental, de modo que o resultado não será confiável. Além disso, é necessário que cada amostra injetada no espectrômetro de massa contenha apenas peptídeos derivados de uma única proteína, o que exclui certas abordagens em larga escala que realizam a clivagem de várias proteínas simultaneamente e posterior separação dos peptídeos resultantes.

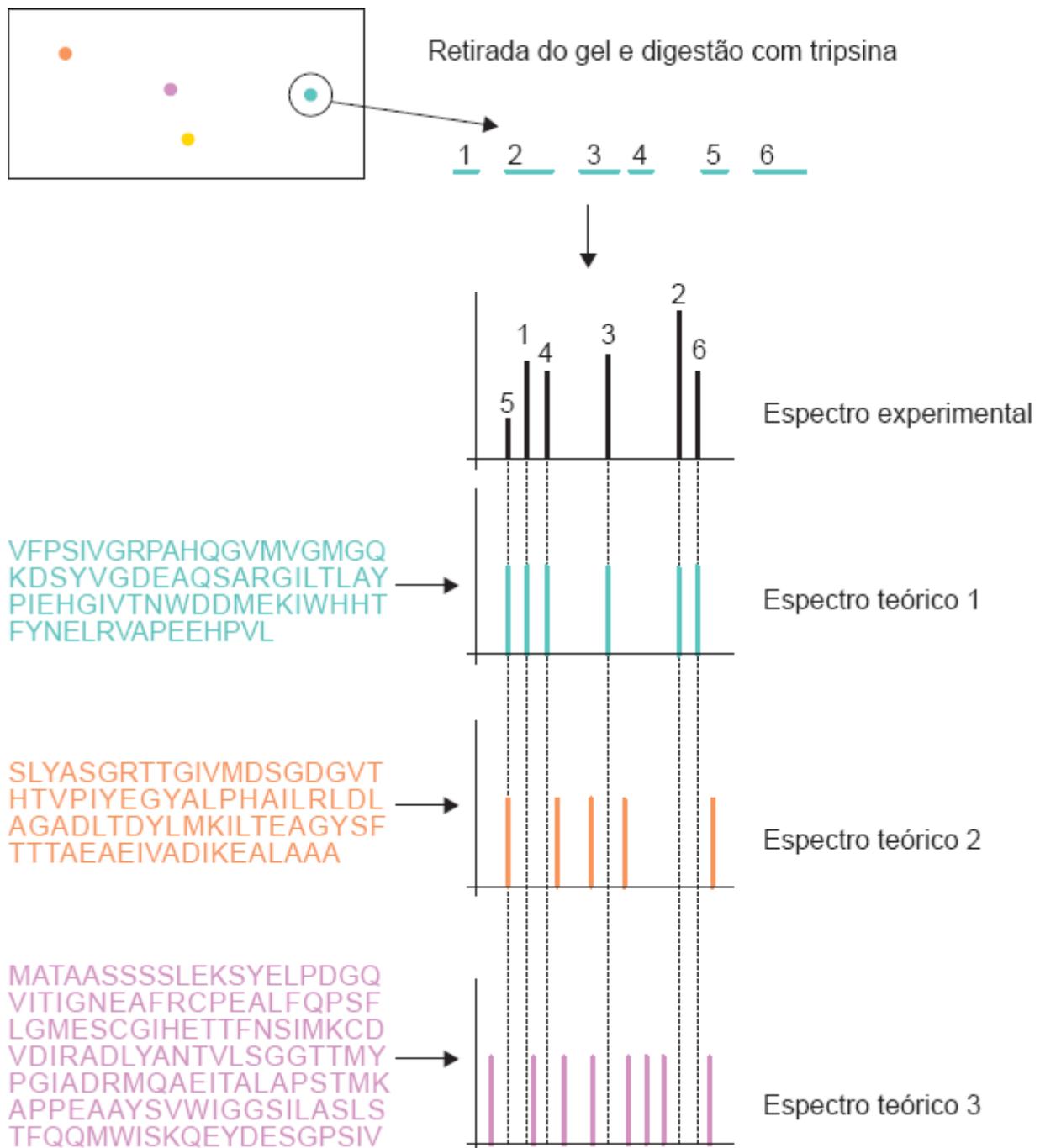
Devido a essas limitações, em geral, a análise de mistura de peptídeos produzidos a partir da digestão de múltiplas proteínas precisa ser avaliada por uma metodologia que forneça mais detalhes do que a simples espectrometria de massa. Isso é alcançado com a utilização da espectrometria de massa em tandem, também referenciada como MS/MS. Nessa técnica, cada uma das moléculas representadas pelos picos gerados pela espectrometria de massa tradicional passa por um processo adicional de fragmentação. Os fragmentos gerados também terão a sua relação massa/carga medida pelo espectrômetro, formando um espectro de fragmentação para cada um dos picos do espectro original.

Em geral, a fragmentação dos peptídeos no processo de MS/MS é obtida pelo choque destes com um gás inerte, geralmente o argônio. Nessas condições, os peptídeos tendem a sofrer quebras principalmente em suas ligações peptídicas (Figura 13.19). É possível selecionar um único peptídeo de uma mistura para análise de fragmentação por vez, pois os aparelhos que realizam espectrometria de massa em tandem têm uma espécie de filtro de massas

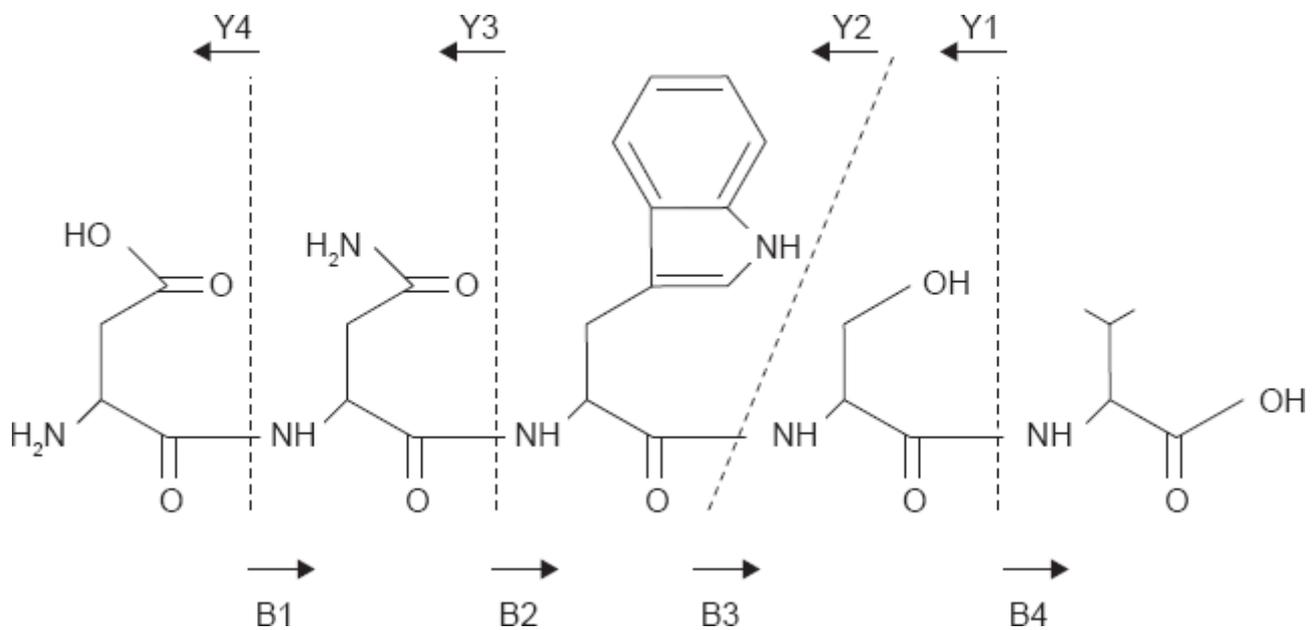
(chamado de quadrupolo), que permite selecionar íons dentro de uma faixa estreita de massa molecular/carga. Os fragmentos resultantes passam então para o analisador do aparelho, que irá registrar o espectro de massa correspondente aos novos fragmentos gerados. Na prática, isso possibilita que, ao injetar no aparelho uma mistura de peptídeos, ele seja capaz de separar cada um deles e gerar um espectro de fragmentação individual.

Nesse tipo de experimento, além da medida da massa de cada peptídeo, será obtido um registro dos fragmentos gerados a partir de quebras na sua cadeia principal. Assim, espera-se que seja formada uma série de fragmentos que tenham diferenças de massa de um aminoácido (Figura 13.20).

A medida da massa exata de cada um dos fragmentos permite determinar sequencialmente cada um dos aminoácidos de uma cadeia polipeptídica. Os únicos aminoácidos para os quais não é possível realizar uma distinção por espectrometria de massa são a leucina e isoleucina, que apresentam massas idênticas entre si. No entanto, para realizar esse tipo de sequenciamento, é necessário um espectro de fragmentação de ótima qualidade, que garanta a definição de pelo menos um fragmento para cada ligação peptídica existente no peptídeo. Infelizmente, a obtenção de espectros com a qualidade necessária não é comum, pois nem sempre os fragmentos representando a quebra de cada ligação peptídica estão presentes no espectro produzido; além disso, picos representando quebras em ligações em regiões diferentes da molécula podem complicar a análise do espectro. Assim, não é possível realizar o sequenciamento de peptídeos de modo sistemático, como acontece para ácidos nucleicos. Isso impede um sequenciamento em larga escala do proteoma com base nos métodos atuais de espectrometria de massa.

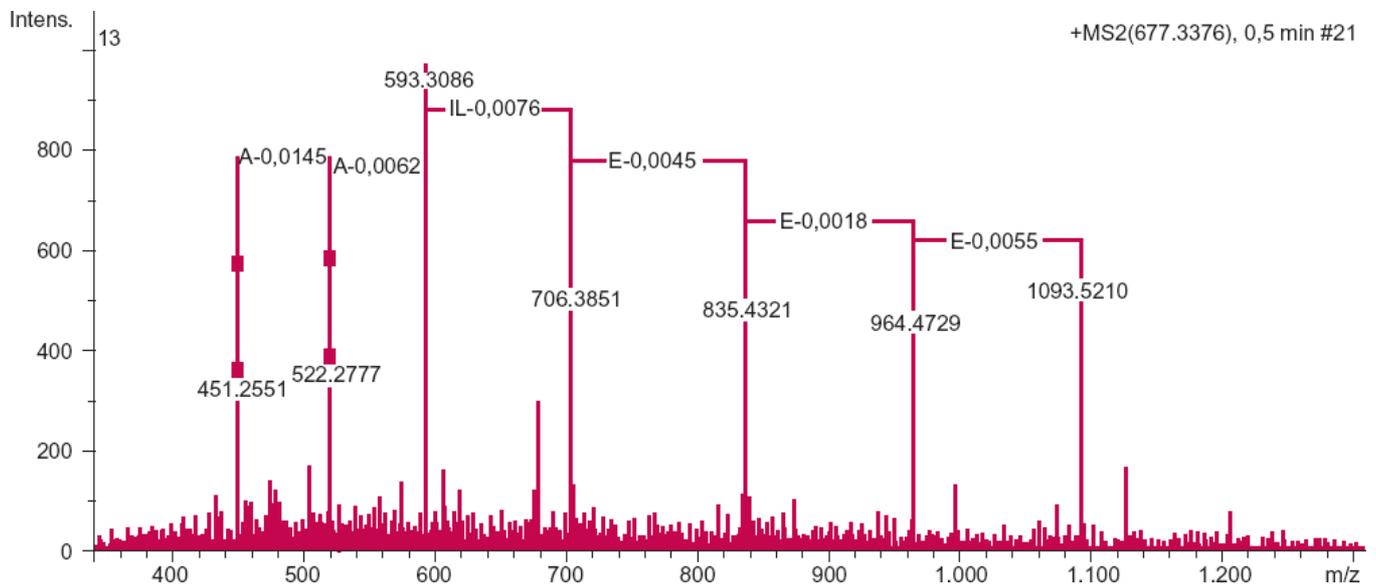


**Figura 13.18** Identificação de proteínas com a utilização de espectrometria de massa. O retângulo representa um gel proveniente de uma eletroforese bidimensional de proteínas. Os pontos representando as diferentes proteínas são exibidos em diferentes cores. O ponto representando a proteína verde é retirado para digestão, resultando em seis diferentes peptídeos que, ao serem analisados por um espectrômetro de massa, resultam em um espectro com sua relação massa/carga (eixo x) e intensidade de sinal (eixo y).



**Figura 13.19** Pontos de fragmentação de um peptídeo que geram fragmentos da série Y e B, normalmente detectados em espectrometria de massa em tandem.

Apesar dessa impossibilidade, é comum realizar identificações em larga escala de proteínas por meio da comparação dos espectros de fragmentação gerados para cada peptídeo com os espectros de fragmentação teóricos calculados a partir de sequências de proteínas em um banco de dados. Essas sequências de proteínas costumam ser deduzidas a partir do quadro de leitura aberto de sequências de RNA ou DNA, permitindo que tenhamos grandes bancos de sequências para diversos organismos. Com esse banco de proteínas é realizada uma simulação computacional do processo de digestão por tripsina, resultando em um banco de peptídeos associados às proteínas existentes no banco. Tais peptídeos terão sua fragmentação simulada computacionalmente, resultando em espectros teóricos de fragmentação para cada um deles.



**Figura 13.20** Espectro de massa da fragmentação de um peptídeo pela qual é possível determinar a sequência de aminoácidos a partir da diferença de massa entre os picos. Junto à linha, ligando os picos, existe uma letra representando um aminoácido: E: ácido glutâmico; I/L: isoleucina/leucina; A: alanina. Os números ao lado de cada letra indicam o desvio entre a massa experimental do aminoácido e a massa teórica. Nota-se que esse desvio representa cerca de 1/10.000 da massa do peptídeo, indicando que as medidas são extremamente precisas.

A análise comparativa dos espectros de fragmentação experimentais e teóricos é muito semelhante à análise de impressão digital de uma proteína com base no espectro de massa de seus peptídeos, que foi explicada anteriormente. A diferença é que, neste caso, se buscam coincidências entre espectros de fragmentação experimental e teóricos de um único peptídeo, e não mais do espectro de um conjunto de peptídeos derivados de uma mesma

proteína. Assim, cada espectro de fragmentação individual possibilitará a identificação de um peptídeo derivado de uma proteína, viabilizando, assim, identificações de proteínas em misturas complexas, sem que haja necessidade de uma separação prévia de cada proteína (Figura 13.21).

### Estudos de interações proteicas e técnica de duplo híbrido

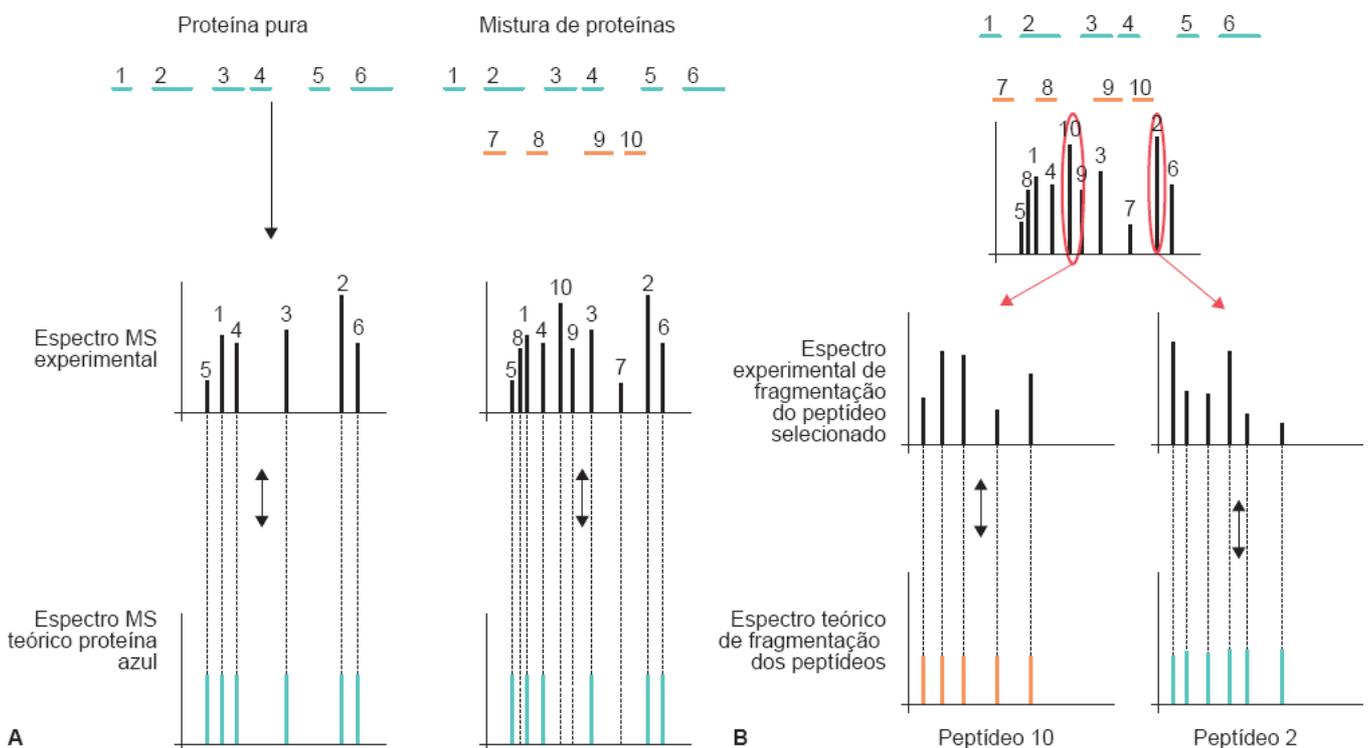
Outra promissora técnica para o estudo de interações do proteoma é a de duplo híbrido. Tal método utiliza um sistema de expressão em levedura contendo um domínio de ligação a DNA, retirado de um ativador transcricional fundido a uma proteína de outro organismo escolhida pelo pesquisador, e para qual se deseja verificar interações (chamada de isca). Essa construção é introduzida em uma célula de levedura, juntamente com diversos clones de uma biblioteca de genes do organismo em estudo, que se encontram fundidos a um domínio de ativação transcricional da levedura (chamados de presa). Quando a isca encontra-se na célula com a presa, caso ocorra interação entre as duas proteínas, haverá um efeito de aproximação entre o domínio de ligação de DNA e o domínio ativador da transcrição, fazendo com que aconteça uma ativação da transcrição de um gene repórter, que fará com que a levedura na qual ocorreu essa interação seja distinguível das outras leveduras (Figura 13.22). Um exemplo de gene repórter utilizado é o LacZ, que codifica para uma proteína metabolizadora de açúcares, a qual, quando na presença de um substrato chamado X-gal, produz coloração azulada.

Uma vez detectada uma levedura onde ocorre interação, o clone derivado da biblioteca de presas é sequenciado para determinação de sua identidade. Deve-se notar que a técnica de duplo híbrido permite uma sondagem inicial de possíveis interações, mas é preciso realizar experimentos mais específicos para confirmação da interação detectada.

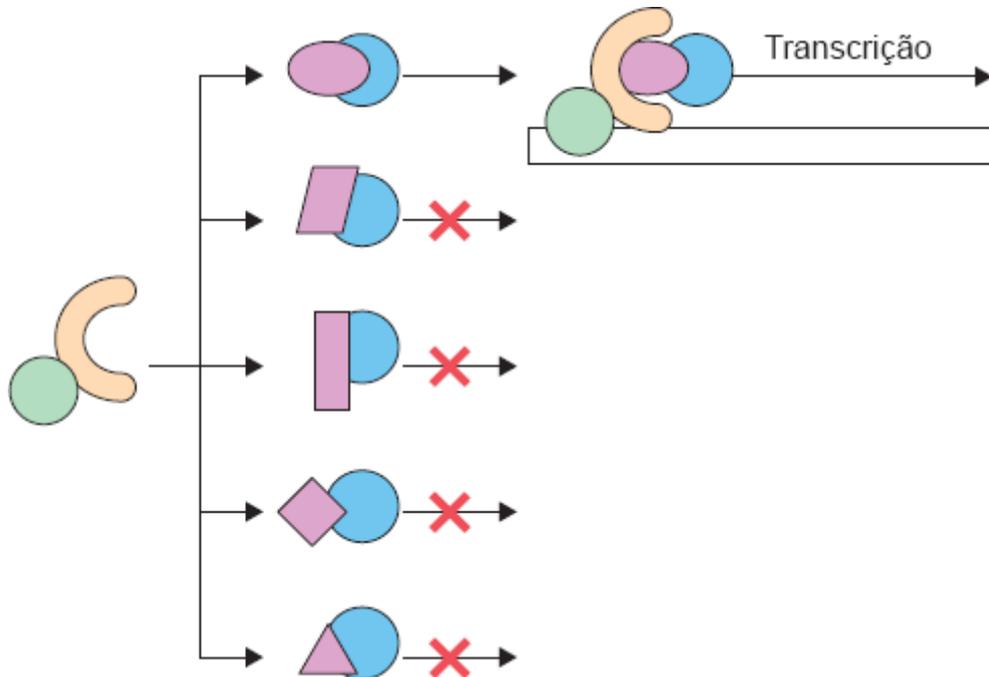
## Estudo do transcriptoma e do proteoma | Entendimento da biologia celular

### Dinâmica do transcriptoma e complexidade em organismos

Durante o sequenciamento do genoma humano, um dos desafios que instigava os cientistas era a definição de quantos genes existiriam neste genoma. Tal era a expectativa que, no ano 2000, vários respeitados profissionais realizaram uma aposta para adivinhar o número de genes que iria surgir a partir do sequenciamento completo do genoma humano, a ser publicado dentro de 1 ano. As projeções para o número de genes humanos foram de 27.000 até 153.000 genes. Tais números eram justificados por algumas análises preliminares, além do fato de genomas de organismos bem mais simples – como a mosca *Drosophila melanogaster* e o verme *Caenorhabditis elegans* – apresentarem aproximadamente 14.000 e 19.000 genes, respectivamente. Portanto, devido à grande complexidade do ser humano comparada a esses organismos mais simples, era natural imaginar que ele teria pelo menos o dobro do número de genes. Não foi sem grande surpresa que, após o término do genoma humano, foi anunciado que o ser humano tinha apenas cerca de 30.000 genes, sendo esse número revisto posteriormente para apenas 24.000.



**Figura 13.21** Comparação entre a identificação por impressão digital utilizando diferentes modos de espectrometria de massa. **A.** A utilização da espectrometria de massa de uma mistura de peptídeos derivados de uma única proteína (verde) possibilita a identificação a partir da comparação entre os picos dos espectros experimental e teórico. No entanto, a adição de peptídeos de uma segunda proteína (laranja) irá gerar novos picos no espectro e, portanto, haverá picos sem coincidência entre os espectros, não sendo mais possível identificar a proteína. **B.** Utilizando a espectrometria de massa em tandem, é possível selecionar picos representando peptídeos individuais (no caso da figura, os peptídeos 2 e 10) e é realizada a fragmentação destes, gerando um espectro individual de fragmentação de cada peptídeo. Comparando o espectro de fragmentação experimental com o teórico, é possível identificar os peptídeos números 2 e 10 e, conseqüentemente, inferir a presença das proteínas verde e laranja na amostra.



**Figura 13.22** Representação esquemática da técnica de duplo híbrido. O esquema mais à direita representa a proteína-isca produzida a partir da fusão do gene de interesse (cujo produto é a porção laranja) com uma porção codificando um domínio de ligação de DNA (cujo produto é representado pelo círculo verde). Setas indicam instâncias distintas nas quais esse produto é exposto a diferentes proteínas presas, representadas em roxo e fundidas a um domínio ativador transcrricional, em azul. Nos casos em que ocorre interação, o domínio ativador transcrricional é recrutado à proximidade do sítio promotor reconhecido pela proteína ligante de DNA, o que possibilita a transcrição de um gene repórter.

Tais dados pareciam indicar que, em termos moleculares, não haveria uma grande diferença entre seres humanos e outros organismos multicelulares mais simples. No entanto, genes de mamíferos, quando comparados com aqueles de organismos mais simples, tendem a ser subdivididos em um maior número de exons. Isso sugere que, apesar de não haver um ganho tão considerável no número de genes, eles seriam mais complexos em mamíferos.

De fato, a partir do estudo em larga escala do transcriptoma humano, pesquisadores passaram a notar que genes de humanos tendiam a realizar com maior frequência o *splicing* alternativo de seus transcritos do que os organismos mais simples. A presença de genes com maior número de exons facilita o processo de retirada/introdução de módulos funcionais em uma proteína por meio do *splicing* alternativo, criando variantes que podem realizar funções diferentes da proteína original. Esses dados parecem apontar que a evolução de organismos mais complexos não é realizada somente com a aquisição de novos genes, mas também com a utilização criativa daqueles que já existem no organismo.

Estudos recentes comparando genomas e transcriptomas de primatas com o de humanos sugerem que grande parte das alterações fenotípicas entre as espécies é causada por variações no padrão de expressão dos genes, em vez de mutações nas suas regiões codificantes. Mutações em regiões não codificantes, como em promotores, tendem a produzir mudanças mais sutis que mutações em regiões codificantes, que levam a modificações em sequências proteicas. Tais dados indicam que a coordenação da expressão dos genes nos diversos tecidos possui um papel bastante relevante na evolução de organismos.

A visão mais tradicional do transcriptoma tinha como base o fato de que a maioria dos sítios de transcrição em nosso genoma estava associada à produção de mRNA para a produção de proteínas. Em adição a estes, existiriam os

*loci* para RNA estruturais, como o rRNA e o tRNA. No entanto, nos últimos anos, a exploração do transcriptoma de organismos utilizando tecnologia de microarranjos e sequenciamento em larga escala possibilitou a verificação de transcritos derivados de *loci* do genoma que não correspondiam aos exons geradores de mRNA. Utilizando esses dados, foi possível definir uma série de *loci* no genoma localizados em introns de genes ou em regiões intergênicas, onde ocorre a transcrição de RNA não codificantes. Essa classe de RNA não tem o potencial de produzir proteínas, pois não dispões de um quadro de leitura aberto em sua sequência que possibilite sua decodificação pelos ribossomos. Apesar de ainda não haver um consenso a respeito da função de tais RNA, há hipóteses de que eles realizariam a regulação da transcrição dos mRNA.

Esses dados sugerem que a dinâmica do transcriptomas de organismos eucariotos é mais complexa que o previamente imaginado. A existência de RNA não codificadores regulatórios adicionaria outro nível de regulação da abundância de transcritos, aumentando ainda mais a importância da caracterização de transcriptomas para o entendimento dos mecanismos celulares.

## **Eucariotos simples contêm proteomas extremamente intrincados**

Com o desenvolvimento de técnicas proteômicas, foi possível realizar estudos em larga escala com o objetivo de caracterizar o conjunto de proteínas expressas por um organismo ou tecido. Foi realizado um mapeamento completo do proteoma da levedura *Saccharomyces cerevisiae*, um eucarioto unicelular, durante o seu período de crescimento em fase exponencial. Nas células estudadas, foram detectadas proteínas representando 80% dos genes desse organismo (cerca de 4.200 proteínas). A abundância individual das proteínas detectadas variava de 50 a  $10^6$  moléculas por célula.

Ao estudar diretamente proteomas mais complexos, como os tecidos humanos, a lista de identificação de proteínas dificilmente chega ao número de proteínas identificadas em *S. cerevisiae*. Esse aparente paradoxo foi explicado pelo fato de que, nos proteomas de organismos complexos, existe uma diferença muito maior entre as quantidades das proteínas mais abundantes em comparação com as menos abundantes (diferença na ordem de  $10^{11}$ ) do que as diferenças verificadas em células de organismos unicelulares simples (diferença da ordem de  $10^4$ ). Tais diferenças, em associação às limitações técnicas dos presentes métodos para estudo do proteoma, fazem com que, nesses organismos mais complexos, somente a fração mais abundante do proteoma possa ser detectada. Em adição, organismos mais complexos têm maior fração de proteínas que apresentam isoformas, devido a fenômenos como o *splicing* alternativo e modificações pós-traducionais, o que dificulta sobretudo a detecção dessas proteínas. Esses dados demonstram que a diferença entre proteomas de organismos de diferente complexidade não se limita ao número de proteínas presentes. Estudos proteômicos em células mais complexas devem utilizar técnicas de fracionamento celular e outras abordagens para fracionar amostras de proteínas, com o objetivo de obter uma cobertura maior desses proteomas.

Estudos proteômicos em leveduras utilizando inibidores da síntese proteica possibilitaram analisar o processo de degradação de proteínas. Uma vez que a síntese de proteína era bloqueada, a abundância das diversas proteínas da levedura era monitorada. Isso possibilitava a análise de sua taxa de degradação, pois novas proteínas não eram sintetizadas, e qualquer variação da quantidade de proteína é resultado somente da degradação das proteínas já existentes no citoplasma. Foi possível notar que as proteínas tinham, em média, uma meia-vida (tempo necessário para atingir metade da concentração original) de 43 min. No entanto, foi possível dividir as proteínas em duas populações distintas: a primeira consiste em proteínas produzidas em grandes quantidades e com grande estabilidade (esta classe é enriquecida com enzimas envolvidas na síntese de proteínas); a segunda é formada por proteínas de menor abundância que são degradadas rapidamente (esta classe é enriquecida com proteínas envolvidas no controle do ciclo celular). Aparentemente, o desenvolvimento dessas duas populações é uma adaptação das células para impedir gastos energéticos desnecessários ao evitar a degradação rápida de proteínas essenciais, mas, ao mesmo tempo, tendo uma população de proteínas de regulação com menor estabilidade, que possibilite respostas celulares rápidas através da variação da abundância dessas proteínas.

Estudos de larga escala utilizando a técnica de purificação por afinidade em tandem com proteínas da levedura *S.cerevisiae* tornaram possível caracterizar centenas de complexos proteicos, que resultam de milhares de interações em células que se apresentavam em fase de crescimento em meio de cultura. Estudos de duplo híbrido com proteínas de *S.cerevisiae* também permitiram mapear milhares de interações entre proteínas desse organismo, confirmando o alto número de interações entre as suas proteínas. Esses dados demonstram que a grande parte das

proteínas não atua de modo isolado, mas sim como parte desses complexos multiproteicos. Devido à alta complexidade de células de organismos eucarióticos multicelulares, é difícil realizar uma catalogação de todas as interações de uma célula, como foi feito para levedura. No entanto, estudos mais restritos parecem confirmar que um grande número de interações também existe nas células mais complexas.

Todos esses dados demonstram que até mesmo o proteoma de organismo mais simples apresenta uma grande heterogeneidade em relação à abundância e estabilidade de proteínas. Além disso, o entendimento atual das redes de interação das proteínas presentes no proteoma continua bastante preliminar, e ainda são necessários muitos estudos para compreender de modo satisfatório os diversos mecanismos celulares mediados pelo nosso proteoma.

## Considerações finais

---

O avanço das tecnologias no estudo de biomoléculas vem possibilitando uma nova abordagem mais abrangente, que permite interrogar o comportamento de grandes conjuntos de moléculas, em contraste com técnicas voltadas para o estudo caso a caso. Isso torna possível o desenvolvimento de uma percepção mais global do funcionamento celular e dos desafios em seu estudo. Ainda restam enormes lacunas em nosso entendimento completo da biologia molecular da célula e, certamente, são necessárias maior acumulação e integração de dados biológicos. No entanto, se considerarmos que essas técnicas de estudo em larga escala são relativamente recentes, é possível perceber seu grande potencial, e esperar que elas favoreçam grandes avanços no entendimento da biologia celular.

## Bibliografia

---

- Beck M, Claassen M, Aebersold R. Comprehensive proteomics. *Curr Opin Biotechnol.* 2011;22(1):3-8.
- Brady SM, Long TA, Benfey PN. Unraveling the dynamic transcriptome. *Plant Cell.* 2006;18(9):2101-11.
- Cox J, Mann M. Is proteomics the new genomics? *Cell.* 2007;130(3):395-8.
- Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, et al. Global analysis of protein expression in yeast. *Nature.* 2003;425(6959):737-41.
- Griffiths J. A brief history of mass spectrometry. *Anal Chem.* 2008;80:5678-83.
- Gustincich S, Sandelin A, Plessy C, Katayama S, Simone R, Lazarevic D, et al. The complexity of the mammalian transcriptome. *J Physiol.* 2006;575(Pt 2):321-32.
- O'Farrell PH. The pre-omics era: the early days of two-dimensional gels. *Proteomics.* 2008;8(23-24):4842-52.
- Larance M, Lamond AI. Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol.* 2015;16(5):269-80.
- Völkel P, Le Faou P, Angrand PO. Interaction proteomics: characterization of protein complexes using tandem affinity purification-mass spectrometry. *Biochem Soc Trans.* 2010;38(4):883-7.
- Wilkins MR. Hares and tortoises: the high- versus low-throughput proteomic race. *Electrophoresis.* 2009;30(Suppl 1):S150-5.