Proteomas: a interface entre a biologia molecular e a bioquímica das proteínas

GABRIELA ALMEIDA¹, CARLA RODRIGUES¹, JORGE LAMPREIA¹

1. Introdução

A sequenciação dos genomas de diversos organismos está a decorrer a um ritmo consideravelmente rápido, gerando vastas quantidades de informação genética, cujas possibilidades de aplicação são diversas: virtualmente, todas as áreas da Biologia e da Biomedicina serão por elas influenciadas. [1] Contudo, tem sido fraco o impacto destes dados nas experiências bioquímicas, especialmente quando direccionadas para a caracterização de proteínas individuais e de complexos proteicos. [2] Esta situação deve-se à existência de certas lacunas na informação potencialmente contida nos Genomas. De facto, à medida que se tem vindo a identificar novos genes, tornou-se evidente que a mera inspecção, mesmo que pormenorizada, da sequência de ácidos desoxirribonucleicos (DNA), não é suficiente para elucidar a função biológica das proteínas por eles codificadas, nem tão pouco deduzir outras informações importantes, como o nível de expressão proteica e as modificações co- e pós-tradução, que por seu turno, podem influenciar algumas propriedades das proteínas, tais como a carga, a hidrofobicidade, a conformação e a estabilidade. [3-6]

Esta problemática tem incentivado o desenvolvimento de alguns projectos de análise descritiva, em larga escala, das PROTEÍNAS expressas pelo genOMA de determinados organismos, tecidos, células ou fluidos biológicos – o PROTEOMA. A técnica suporta à produção de

Proteomas é a electroforese bidimensional (2DE), uma técnica poderosa na resolução de misturas complexas de proteínas. No gel 2D resultante é possível observar inúmeras manchas, cada qual correspondendo a uma única espécie proteica (figura 1). Feita a identificação e caracterização e posterior anotação de todas as proteínas obtém-se o Proteoma, que não é mais do que um mapa de referência, pleno de informação sobre a amostra examinada. [8]

figura 1 Gel 2D das proteínas da fracção solúvel da bactéria redutora de sulfato *Desulfovibrio desulfuricans* ATCC 27774. [8]

Para além da possibilidade de catalogação extensiva das proteínas complementares de um dado Genoma, um outro aspecto relevante dos Proteomas advém do seu carácter dinâmico, dado reflectirem os estados dos sistemas biológicos, ao modificarem-se com o nível de desenvolvimento, o tipo de tecido ou com as condições ambientais e de

stress às quais o organismo está sujeito; em contrapartida, o Genoma permanece imutável durante a totalidade da vida de um organismo. Esta mais valia contribuiu fortemente para que a análise de Proteomas – a PROTEÓMICA – se tornasse numa área de investigação emergente, capaz de oferecer novas ferramentas para o estudo de problemas biológicos complexos, nomeadamente nos ramos biomédicos, farmacêuticos, biotecnológicos e na agricultura. [3,6,9]

Embora recentes, os projectos proteómicos são um assunto de extrema importância, tal como pode ser comprovado pelo gráfico da figura 2, onde se ilustra a pronunciada evolução do número de publicações nalgumas revistas de prestígio, incluindo as conceituadas *Science* e *Nature*, dedicadas à análise dos proteomas completos de alguns organismos relativamente simples, e à caracterização dos perfis de expressão proteica em determinadas células e tecidos de organismos superiores. [10]

É importante referir que a expansão vertiginosa da Proteómica, só foi possível mediante o simultâneo desenvolvimento, integração e automatização de uma plethora de técnicas, que permitem resolver (2DE), identificar (micro-sequenciação, composição dos aminoácidos, espectrometria de massa, etc.) quantificar e caracterizar proteínas, bem como armazenar (bases de dados de géis 2D) e interligar esta informação com a obtida a partir dos programas genómicos (Bioinformática). [4,11,12]

¹Departamento de Química, Centro de Química-fina e Biotecnologia, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

2. Do Genoma para o Proteoma

Tal como foi anteriormente referido, o aparecimento de projectos proteómicos é uma consequência directa do desenvolvimento de projectos de caracterização de diversos genomas. Na era *genómica*, o principal objectivo consistia em expor o genoma completo de todos os organismos Eucariotas e Procariotas existentes, de forma a compreender a complexidade e diversidade dos mesmos.[3]

Desde a publicação em 1995, da primeira sequência nucleotídica de um genoma microbiano (*Haemophilus influenzae* Rd), foram já depositadas nas bases de dados 56 sequências genómicas completas. Entre estas, 4 são de proeminentes organismos Eucariotas, nomeadamente a levedura *Saccharomyces cerevisiae* (1999) e os 3 únicos genomas de organismos multicelulares, o verme *Caenorhabditis elegans* (1998), a mosca da fruta *Drosophila melanogaster* (1999) e a planta *Arabidopsis thaliana* (2000). Dos restantes genomas, 10 pertencem ao reino Archae e 42 pertencem ao reino Bactéria, incluindo-se neste último grupo importantes organismos modelo, como as bactérias *Bacillus subtilis* (1997) e *Escherichia coli* (1998).[10]

Entre todos os projectos de sequenciação em curso ou já terminados sobressai, indiscutivelmente, o do genoma Humano, um projecto ambicioso, liderado pela empresa *Celera Genomics Group* e pelo consórcio público *Human Genome Sciences Inc.*, e que tem recebido uma cobertura invulgar por parte da comunicação social, dado a expectativa e a controvérsia em torno dele criadas. O esboço oficial do genoma Humano (32.000 genes, dos quais foram identificados 22.000) foi já publicado no início do corrente ano, quatro anos antes da data prevista, e as conclusões foram, em parte, surpreendentes. Quando comparado com os genomas conhecidos, o número total de genes é inferior ao esperado, demonstrando que a complexidade de um organismo não é directamente proporcional à dimensão do seu DNA. Por outro lado, ficou patente que o perfil genético de um indivíduo

figura 2 Evolução do número de publicações de artigos científicos sobre o tema "Proteoma", de acordo com uma pesquisa na base de dados da PUBMED.[10]

por si só é insuficiente para determinar a sua evolução. O próximo passo consistirá na análise descritiva das moléculas funcionais propriamente ditas – as proteínas.[1]

Com efeito, embora a sequência genómica possa ser usada, por um lado, para prever os correspondentes produtos genéticos e como tal, trazer inúmeras vantagens para a compreensão dos fenómenos biológicos, por outro lado, deixa em aberto, questões não menos importantes, como por exemplo, *quando i.e.*, em que nível de desenvolvimento ou diferenciação é que um certo gene é transcrito e traduzido, *a que velocidade* ocorrem estes processos e quais os *produtos finais* deste gene. Além disso, a descodificação da sequência de ácidos nucleicos tem-se mostrado claramente insuficiente para atribuir *funções* até cerca de 40 a 60% dos genes estru-

turais de um organismo em particular, e não permite sequer saber, em organismos multicelulares, *em que células* as proteínas serão expressas (figura 3). Ou seja, a complexidade funcional de um organismo excede largamente as indicações fornecidas isoladamente pela sequência genética. O desvendar desta complexidade, constitui pois, o grande desafio da era *pós-genómica* e a solução parece residir na análise directa dos produtos genéticos, nomeadamente do ácido ribonucleico (RNA) e/ou das proteínas.[3,13,14]

Os primeiros estudos relacionados com a expressão genética baseavam-se preferencialmente em análises e quantificações do RNA mensageiro (mRNA), o que não é de todo surpreendente dado ser mais fácil trabalhar com os ácidos nucleicos do que com as proteínas: a estrutura básica dos ácidos nucleicos é

figura 3 Proteómica: a tradução da genómica nos seus produtos.

tabela 1 listagem representativa da diversidade de Base de Dados de Géis 2D, disponíveis na Web

BASE DE DADOS	ORGANIZAÇÃO	ESPÉCIE(S)
SWISS-2DPAGE ⁽¹⁾	Swiss Institute of Bioinformatics, Geneva University Hospital (Suíça)	Homem, Rato, <i>Arabidopsis thaliana</i> , <i>Dictyostelium discoideum</i> , <i>Escherichia coli</i> , <i>Saccharomyces cerevisiae</i>
SIENA-2DPAGE ⁽¹⁾	Department of Molecular Biology, University of Siena (Itália)	Homem, <i>Chlamydia trachomatis</i> , <i>Caenorhabditis elegans</i>
2-DE database ⁽¹⁾	Max Planck Institute for Infection Biology, Berlin (Alemanha)	Homem, <i>Mycobacterium bovis</i> , <i>Mycobacterium tuberculosis</i> , <i>Helicobacter pylori</i> , <i>Borrelia garinii</i>
Argonne Protein Mapping Group	Center for Mechanistic Biology and Biotechnology, Argonne National Laboratory (EUA)	Homem, Rato, <i>Pyrococcus furiosus</i>
Large Scale Proteomics	Large Scale Biology Corp. ⁽²⁾ (EUA)	Homem, Rato, Milho, Trigo
HSC-2DPAGE ⁽¹⁾	Heart Science Centre, Harefield Hospital	Homem, Rato, Cão
RAT Heart-2DPAGE and HEART-2DPAGE ⁽¹⁾	German Heart Institute, Berlin (Alemanha)	Homem, Rato
Human and Mouse 2DPAGE Databases ⁽¹⁾	Danish Centre for Human Genome Research, (Dinamarca)	Homem, Rato
BALF 2D Database	Department of Biological Chemistry, University of MonsHainaut (Bélgica)	Homem, Rato
HP-2DPAGE ⁽¹⁾	MDC, Berlin (Alemanha)	Homem
Inner Ear Protein Database ⁽¹⁾	Washington University (EUA)	Homem, Porco da Guiné
PMMA-2DPAGE ⁽¹⁾	Purkyne Military Medical Academy (Rep. Checa)	Homem
2D-Database of Human Lysosomal Proteins	Joint Protein Structure Lab. (Austrália)	Homem
Toothprint (2DE gels of dental tissues) ⁽¹⁾	University of Otago, Dunedin (Nova Zelândia)	Rato
PHCI-2DPAGE ⁽¹⁾	Department of Medical Microbiology and Immunology, University of Aarhus	Interacção Parasita/Célula Hospedeira (Dinamarca)
ECO2DBASE	Department of Microbiology and Immunology, University of Michigan Medical School (EUA)	<i>Escherichia coli</i>
Cyano2Dbase	Kazuka Research Institute (Japão)	<i>Synechocystis sp</i>
Sub2D – 2D PROTEIN INDEX ⁽¹⁾	Department of BioSciences, University of Greifswald (Alemanha)	<i>Bacillus subtilis</i>
Human PSD TM YPD TM Yeast Protein Database	Proteome Inc. ⁽²⁾ (EUA)	Homem <i>Saccharomyces cerevisiae</i>

Notas: (1) Base Federada (2) Empresa privada: consulta mediante subscrição.

Para uma listagem mais completa, consultar o site [http:// www.lecb.ncicrf.gov/EP/table2Ddatabases.html](http://www.lecb.ncicrf.gov/EP/table2Ddatabases.html)

mais simples e mais homogênea, são mais estáveis e, através da técnica *Polymerase Chain Reaction* (PCR) é possível amplificá-los rapidamente. Porém, foi cedo demonstrado que a correlação entre a abundância do mRNA e a quantidade das proteínas num dado momento é bastante pobre, devido principalmente aos diferentes tempos de meia-vida biológica das duas famílias de moléculas. Por outro lado, o actual conhecimento da existência de rearranjos do material genético (*splicing* alternativo¹, *editing* do mRNA²) e das modificações pós-tradução (fosforilações, glicosilações, acetilações, etc.) mostrou haver forte probabilidade de ocorrência de variação molecular entre os genes e o seus correspondentes produtos activos, derrubando consequentemente o paradigma "um gene, uma proteína". Um simples segmento de DNA é, portanto, capaz de gerar inúmeros produtos proteicos complexos³. [1-3,5,6]

Pode-se assim afirmar que os estudos baseados nos ácidos nucleicos, embora altamente eficientes, têm necessariamente de ser acompanhadas por uma análise massiva das proteínas, quer ao nível qualitativo, quer ao nível quantitativo. Deste modo, sempre que se pretende uma visão global da complexidade de um organismo, deve proceder-se tanto à análise do Genoma como do Proteoma. [5,13]

Actualmente, estão já em pleno desenvolvimento projectos de análise do proteoma de diversos organismos, muitos dos quais têm o genoma completamente mapeado. A Tabela 1 descreve algumas das principais bases de dados disponíveis na Web, ilustrando a diversidade de organismos e materiais celulares em estudo, bem como das instituições públicas e privadas envolvidas na Proteómica.

3. Tecnologia Proteómica

Os estudos Proteómicos baseiam-se na caracterização sistemática de amostras complexas de proteínas, envolvendo para tal, uma série de técnicas de separação, identificação e quantificação sistemática e simultânea de muitas proteí-

nas, provenientes de uma única amostra.

O primeiro passo em qualquer projecto e, por sinal, um dos mais críticos, é a **preparação da amostra**. A solubilização completa da totalidade das espécies proteicas é crucial, de molde a produzir resultados válidos e reprodutíveis. Contudo, a complexidade da análise é, por vezes, de tal ordem que se prefere proceder a uma purificação parcial do material de partida. [3,15]

O segundo passo consiste na separação das várias proteínas do extracto numa só experiência. A técnica bioquímica que presentemente oferece melhor capacidade de resolução dos componentes de misturas complexas de proteínas é a **electroforese 2D**. Outras técnicas, como a cromatografia líquida ou a electroforese capilar acoplada à detecção por espectrometria de massa, estão actualmente em fase experimental, não conseguindo ainda providenciar proteínas purificadas com elevada resolução.[3] Os princípios básicos da 2DE, estipulados há 25 anos atrás por O'Farrel e J. Klose, mantiveram-se até à data inalterados. Essencialmente, as proteínas de uma amostra biológica são separadas, numa 1.^a dimensão, de acordo com a sua carga ou, mais correctamente, com o seu ponto isoelectrico (pI) e, a seguir, na 2.^a dimensão, com base no tamanho molecular (M_r), resultando um mapa característico de manchas, cada qual correspondendo a uma única espécie proteica. Na prática, corresponde a efectuar consecutivamente, uma focagem isoelectrica (FIE) e uma electroforese desnaturante em gel de poliacrilamida (SDS-PAGE). Idealmente, podem ser separadas em simultâneo milhares de proteínas diferentes, das quais se pode obter informação directa sobre o pI, a M_r e a quantidade. Originalmente, a FIE era realizada em tubos de géis e o gradiente de pH era estabelecido através de anfólitos transportadores. Esta metodologia apresentava algumas limitações sérias, sobretudo no domínio da reprodutibilidade. Tais inconvenientes foram recentemente ultrapassados com a introdução no mercado de gradientes de pH imobilizados

(IPG) em tiras de géis de acrilamida. Os mapas 2D assim produzidos têm um elevado nível de resolução e reprodutibilidade.[4,16] O recurso às tiras de IPG também alargou o intervalo de pH de utilização e aumentou a capacidade de carga da amostra. Todavia, persistem alguns problemas ao nível da sensibilidade da 2DE: as proteínas hidrófobas não são convenientemente solubilizadas e/ou introduzidas na primeira dimensão e as proteínas existentes em pequena quantidade (*low copy number*), que frequentemente constituem as proteínas chave nos sistemas em estudo, não são detectadas ou são mascaradas pelas proteínas predominantes (*high copy number*).[4,8] Importa salientar que muitas vezes, as proteínas minoritárias constituem justamente os marcadores de resposta biológica.

A **deteção** das manchas num gel 2D poderá ser feita no próprio gel, sendo a visualização tipicamente feita por coloração com Azul de Coomassie ou com Nitrato de Prata. Actualmente, existe uma alternativa de coloração com moléculas fluorescentes (por exemplo: laranja e vermelho SYPRO), que combina as vantagens dos dois métodos anteriores: tem um limite de detecção comparável ao do Nitrato de Prata e, tal como o Azul de Coomassie, é simples de executar, fornece uma análise quantitativa e é reprodutível. É ainda compatível com as técnicas de análise de espectrometria de massa e de sequenciação de aminoácidos. Requer porém, equipamento especial para visualização e registo de imagem, de custo relativamente elevado. Recentemente, tem-se igualmente vindo a assistir à recuperação dos métodos de detecção mais sensíveis, como a Autoradiografia e a Fluorografia, usados na visualização e quantificação de proteínas marcadas *in vivo* com isótopos rádioactivos como o ³⁵S, ¹⁴C, ³H, ³²P ou ¹²⁵I. [4,8,9]

Qualquer um destes modos de revelação dos géis 2D não são específicos. Assim, a utilidade do sistema é bastante limitada, resumindo-se à informação resultante da comparação dos perfis proteicos em géis distintos. [14] O progresso experimental depende então da

identificação de cada uma das manchas detectadas. Para tal, recorre-se a técnicas analíticas capazes de fornecer informações de outro tipo, tais como a Microsequenciação Química, a Análise dos Aminoácidos e a Detecção Imunoquímica, agora viável devido à grande variedade de anticorpos disponíveis. Todas estas tecnologias são eficazes e têm a vantagem de necessitarem de quantidades mínimas de amostra. Deve-se porém sublinhar que são relativamente morosas, comprometendo o tempo de resposta da análise.[8,15] Por esse motivo, a técnica preferencial de análise e identificação das proteínas exibidas num gel 2D é, presentemente, a *Espectrometria de Massa*, uma estratégia ímpar, que conjuga precisão com elevada sensibilidade e rapidez de resposta. Nesta abordagem, as proteínas são retiradas directamente do gel 2D e, ou são imediatamente analisadas, ou são digeridas química ou enzimaticamente, para produzir um padrão de degradação único, que possa ser reconhecido por espectroscopia de massa (*peptide mass fingerprint*). É assim possível derivar informação sobre a sequência de aminoácidos, que poderá ser cruzada com a informação disponível nas bases de dados genómicas, permitindo a identificação da proteína original. [12,15]

Em articulação com estas técnicas, encontra-se a **Bio-informática**, uma área de investigação na qual a biologia, a ciência computacional e a tecnologia de informação se fundem numa única disciplina. No âmbito da proteómica, a metodologia Bio-informática põe à disposição *hardware* e *software* muito potentes e eficientes para avaliações computadorizadas rotineiras dos perfis obtidos nos géis 2D com especial destaque para o *software* de imagem contendo as ferramentas indispensáveis para analisar, quantificar e caracterizar as proteínas nos mapa 2D. Permite ainda processar, comparar e catalogar um conjunto de informações complexas em bases de dados próprias, e interligá-las, quando requerido, com as bases de dados genómicas. [3,9] A possibilidade de relacionar os dados proteómicos textuais ou de imagem com as diferentes

bases de dados existentes, abre as portas para um futuro promissor, onde questões como a identificação de uma proteína pertencente a qualquer espécie viva, a comparação automática inter-espécies e a possível postulação sobre as suas estrutura e função biológica, podem ser respondidas em questão de poucos minutos! [3]

4. A Tecnologia Proteómica na Análise de Expressão Proteica

A compreensão, ao nível molecular, de como as células e tecidos operam em situação saudável e patológica (ou quando alterados pela acção de estímulos externos), requer o conhecimento das proteínas e dos outros componentes celulares presentes, saber como estas entidades interactivam e qual o resultado dessas interacções. A tecnologia proteómica permite obter uma descrição global e instantânea da expressão proteica, numa dada fase de desenvolvimento. [1,8] A abordagem experimental segue, genericamente, a estratégia a seguir descrita (figura 4). Em primeiro lugar, é feita a caracterização do proteoma do sistema biológico, em condições consideradas normais: o proteoma constitutivo. Depois, submete-se o sistema a determinados estímulos que correspondem frequentemente a situações de *stress* ambiental ou patológico. Seguidamente, procuram-se alterações no perfil proteómico, quer ao nível da posição quer ao nível da abundância das proteínas que, para finalizar, deverão ser devidamente identificadas e caracterizadas. Poderá assim, ser possível providenciar o reconhecimento de marcadores de resposta biológica e estabelecer uma correlação entre factores exógenos e a expressão proteica que, por sua vez, poderão contribuir para o esclarecimento do mecanismo de funcionamento de processos celulares.

Tal como poderemos seguidamente apurar, esta metodologia poderá ser utilizada na descoberta de novos marcadores de doença, na validação de alguns alvos de drogas, contribuir para o desenvolvimento de novos antibióticos e

vacinas e, inclusivamente, facilitar as análises toxicológicas. [3,17]

4.1 Aplicações Biomédicas e Farmacêuticas

O esforço desenvolvido em torno da compreensão do Genoma Humano, quer ao nível da análise das sequências dos genes, quer ao nível da sua regulação, tem contribuído decisivamente para o progresso da **Medicina Clínica**, especialmente pela capacidade demonstrada em revelar determinados processos inerentes ao desenvolvimento de doenças de origem genética. Todavia, o impacto destes conhecimentos na medicina laboratorial, nomeadamente na concepção de novas ferramentas de diagnóstico clínico e de novas terapias, não foi tão profundo. Uma das explicações plausíveis é o facto de apenas 2% do total das doenças humanas ser de causa monogénica. Além disso, as condições de vida do indivíduo modulam a predisposição hereditária para o desenvolvimento de patologias. Assim, torna-se particularmente útil efectuar o estudo dos marcadores biológicos de doenças (proteínas que aparecem ou desaparecem ao longo do seu decurso) em certos fluidos corporais ou em biópsias de tecidos. Como tal, não será surpreendente constatar que a aproximação proteómica tem vindo a assumir um papel de destaque na investigação sobre patologias diversas, como é o caso da diabetes, do cancro e das doenças cardiovasculares e cerebrais (ver Tabela 1). Através da análise de mapas 2D de amostras clínicas, é possível monitorizar terapias individualizadas, descobrir novos marcadores de doenças, analisar fenótipos proteicos e identificar a origem das amostras. [3,5,6]

A abordagem proteómica tem igualmente aplicabilidade na área da **Imunologia**, já que pela comparação do proteoma do soro de indivíduos infectados com o proteoma do agente infeccioso, é possível identificar as proteínas imunogénicas que, como se sabe, são potenciais candidatas a vacinas contra organismos patogénicos. É pois provável que a aproximação Proteómica se torne no protocolo base para o estudo de doenças infecciosas, bem como de alergias, uma

figura 4 Esquema genérico ilustrando a estratégia experimental usada na análise de Proteomas. As células (ou tecidos) A e B foram recolhidas em duas condições distintas (fase de diferenciação ou desenvolvimento, estímulos externos, etc.). As alterações na expressão proteica são manifestadas nos mapas 2D. Feita a identificação das proteínas marcadoras, é possível relacioná-las com o mecanismo de resposta biológico.

área de investigação onde o conhecimento da imunogenicidade das proteínas é também crucial. [3]

Outra das principais aplicações da análise de proteomas é na área da **Investigação Farmacêutica** e da **Toxicologia**. Alguns dos principais vectores que governam a Indústria Farmacêutica são a eficácia e a rapidez na proposta de novos medicamentos. Ora, a química combinatória é capaz de desenhar e produzir muito rapidamente centenas de compostos químicos que poderão eventualmente ser usados para esse fim. A comparação entre proteomas de tecidos saudáveis com a de tecidos alterados pela acção de uma determinada droga pode acelerar a selecção, entre vários candidatos, do químico mais específico e eficaz. Paralelamente, providencia informação sobre os seus efeitos colaterais e a sua toxicidade; em várias situações, os efeitos farmacológicos e metabólicos das drogas ocorrem ao nível das modificações pós-tradução, pelo que expô-las através de uma mapa proteómico, pode ser extremamente útil

para a análise dos mecanismos bioquímicos envolvidos, incluindo os que induzem efeitos secundários. [3,17]

Quer a Medicina quer a Farmacologia irão, num futuro relativamente próximo, beneficiar dos resultados daquele que se pode considerar, hoje em dia, um projecto central na investigação proteómica. Trata-se do "Índice de Proteínas Humano" (HPI, do inglês *Human Protein Index*), que tem por finalidade mapear e identificar na íntegra, o proteoma humano. Pretende, por conseguinte, compilar as proteínas expressas na espécie Humana na forma de um inventário compreensivo, específico para cada tecido. Esta base de dados caracterizará as diferenças entre os cerca de 252 tipos de células diferentes existentes no Homem e providenciará os alicerces para a descoberta sistemática das alterações moleculares decorridas ao longo de um processo de doença, descobrir o efeito real das drogas no corpo humano e a identificação de novos alvos para agentes terapêuticos.[18] É, indiscutivelmente, uma tarefa enormíssima,

ainda mais audaciosa do que a sequenciação do Genoma Humano mas, provavelmente, mais profícua para o desenvolvimento das Ciências da Saúde.

4.2 Aplicações Biológicas e Biotecnológicas

Tal como já foi referido, a Proteómica pode servir para investigar a influência de condições de *stress* sobre a regulação e expressão proteica dos organismos. Todas as espécies vivas – desde as bactérias ao Homem, passando pelas plantas – respondem a agressões produzindo "proteínas anti-*stress*". É o caso das proteínas de choque térmico (HSP, do inglês *Heat Shock Proteins*), que têm por função defender o organismo de alterações bruscas de temperatura. Tomemos, por exemplo, as bactérias hipertermófilas habituadas a um meio ambiente de 100°C que sofrem um choque térmico quando a temperatura externa sobe para 105°C, ou o caso das algas de águas glaciares do Antártico que estão adaptadas a temperaturas próximas dos 0°C e entram em *stress* se a temperatura subir para 5°C. Quer

numa situação quer noutra, a resposta é a produção de proteínas HSP. A análise proteómica facilitou a pesquisa sistemática destas proteínas, uma vez que certas alterações na resposta celular são visíveis num mapa 2D. [19]

Na mesma lógica do exemplo anterior, a abordagem proteómica tem sido aplicada ao estudo da influência da composição do meio de crescimento na indução e/ou repressão da síntese proteica de alguns microorganismos. É o caso do projecto em curso sobre a bactéria redutora de sulfato *Desulfovibrio desulfuricans* ATCC 27774, capaz de crescer em nitrato como via alternativa ao sulfato. Um dos principais objectivos é investigar a expressão genética induzida por alterações nos nutrientes (meio rico em nitrato ou sulfato, com e sem ferro), com especial ênfase nas enzimas envolvidas na respiração do nitrato.[8] Outro exemplo comparável prende-se com o estudo sobre o efeito da concentração salina na síntese de proteínas no periplasma da cianobactéria *Synechocystis sp.* O proteoma de células adaptadas a concentrações elevadas de sal exibiu três novas proteínas e mostrou um aumento substancial na abundância de outras seis. Pensa-se que estas proteínas estejam envolvidas na alteração da estrutura da parede celular. [20]

Um exemplo com implicações importantes no domínio da Biotecnologia é o estudo levado a cabo sobre a bactéria hipertermófila *Pyrococcus furiosus* (ver Tabela 1). O facto deste microorganismo ter eventuais aplicações industriais que requerem actividade biológica a temperaturas elevadas incentivou o estudo da sua biologia molecular e bioquímica, com particular interesse na termoestabilidade das suas enzimas. O registo celular proteico num mapa 2D tornou possível a caracterização da estrutura, função e regulação de algumas das proteínas desta bactéria, providenciando as bases para futuras manipulações bioquímicas, de molde a otimizar a produção de enzimas de utilidade. [15]

Outro exemplo relevante prende-se com a investigação realizada sobre relações taxonómicas entre diferentes espécies de leveduras usadas no fabrico indus-

trial de cervejas (as mais modernas estirpes são híbridas, consistindo em pelo menos dois genomas). A caracterização das estirpes foi feita por comparação com os proteomas de várias espécies de *Saccharomyces* isoladas nos fermentadores de fábricas de cerveja *lager*. [21]

A proteómica pode também ser usada no estudo das interações patogénio-hospedeiro ou parasita-hospedeiro. É o caso da fixação de azoto em legumes, através da formação de nódulos por associação com bactérias do género *Rhizobium*. Cada tipo de bactéria possui diferente capacidade de formar nódulos e de fixar o azoto, pelo que a análise proteómica de cada estirpe poderá dar uma visão das moléculas que são importantes neste processo.[3] Como é obvio, os conhecimentos adquiridos através deste tipo de experiências poderão trazer alguns benefícios no domínio da **Agricultura**. Aliás, esta última sofreu recentemente uma revolução biotecnológica, em consequência do aparecimento da tecnologia transgénica que permitiu, por exemplo, a introdução de resistência a patogénios / parasitas (virais, bacterianos, fúngicos ou insectos) na própria planta; muitos dos mecanismos de resistência envolvem a produção de proteínas tóxicas. Ora, não seria surpreendente se, muito brevemente, se comesse a recorrer à Proteómica para descobrir novas proteínas tóxicas ou para identificar plantas que, supostamente, parecem ser resistentes a um patogénio, ao exprimirem elevados níveis de uma proteína relevante imediatamente após a transformação. [3]

Um outro campo de aplicação a considerar futuramente será certamente o do **Controlo de Qualidade**. Um produto bem especificado evita mal entendidos na relação entre produtor e consumidor e, permite estabelecer bases legais sobre a responsabilidade de um determinado produto. Através de um mapa 2D é possível precisar, definir e especificar ao mais alto nível todos os produtos de origem proteica. [3]

5. Considerações Finais

A simples descodificação dos genomas mostrou ser nitidamente insuficiente para decifrar a complexidade dos sistemas biológicos. Ela representa apenas a primeira etapa na busca de respostas para questões relacionadas com o desenvolvimento e o funcionamento dos seres vivos. Decifrar a complexidade dos sistemas biológicos requer não só o entendimento do genoma como também das proteínas por ele expressas. A comparação dos mapas proteómicos em diferentes situações ou fases de desenvolvimento pode fornecer informações únicas e essenciais para a compreensão dos eventos biológicos.

Apesar de residirem algumas dificuldades técnicas associadas à 2DE, nomeadamente ao nível da reprodutibilidade, da sensibilidade e da incapacidade de analisar determinadas proteínas "difíceis", o contínuo melhoramento da plataforma tecnológica para a análise da expressão proteica, especialmente ao nível da caracterização das proteínas, conferem-lhe alguma automatização e grande capacidade de resposta, o que leva a crer que, em combinação com a sequência genómica completa, a aproximação proteómica possa ser usada, a curto prazo, com eficácia e rapidez na identificação de cada uma das proteínas observadas num mapa 2D. Parece pois que a proteómica e a genómica são dois campos de investigação que se complementam de uma forma sinérgica, esforçando-se por, num futuro próximo, caracterizar com pormenor os seres vivos, desde os organismos mais simples aos mais complexos, como o Homem, e providenciar à próxima geração de investigadores um "inventário contendo alguns dos ingredientes fundamentais que compõem a vida". [3,9,22]

Agradecimentos

Os autores agradecem aos Professores Doutores Isabel Moura e José J.G. Moura pelo incentivo e pelas facilidades de trabalho concedidas nos seus grupos de investigação. A Gabriela Almeida

agradece ainda ao programa Praxis XXI, pelo apoio financeiro.

Bibliografia

- 1 – <http://www.nature.com>
- 2 – Shevchenko A. *et al.* (1996), *Proc. Natl. Acad. Sci.* 93, 14440-5.
- 3 – Wilkins M.R., Williams K.L., Appel, R.D., Hochstrasser D.F. (Ed.s) *Proteome Research: New frontiers in functional genomics – Principles and Practice*, 1997, Springer, Alemanha.
- 4 – Dove A. (1999) *Nature Biotech.* 17, 233-236.
- 5 – Kahn P. (1995) *Science* 270, 369-370.
- 6 – Blackstock W., Weir M. (1999) *Trends Biotech.* 17(3), 121-127.
- 7 – Almeida, G., Rodrigues, C., Lino, A.R., Lampreia, I. Moura, J.J.G. Moura, "Nutrients Influence on Protein Expression of a Sulfate Reducing Bacteria *Desulfovibrio desulfuricans* ATCC 27774", poster comm. on 27th FEBS Meeting, Lisbon 2001.
- 8 – Berkelman T., Stenstedt T. *2D Electrophoresis Using Immobilized pH gradients, Principles & Methods*, Amersham Pharmacia Biotech (1999), 80-6429-60.
- 9 – Rabilloud T. (ed.) *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods*, 2000, Springer, Alemanha.
- 10 – <http://www.ncbi.nlm.nih.gov>
- 11 – Celis, J.E., Ostergaard, M., Jensen, N.A., Gromova, I., Rasmussen, H.H., Gromov, P. (1998) *FEBS Lett* 430, 64-72.
- 12 – James P. (1997) *Bioch Biophys Res Comm* 231, 1-6.
- 13 – Hochstrasser D.F. (1998) *Clin Chem Lab Med* 36, 825-836.
- 14 – Pennington S.R., Wilkins M.R., Hochstrasser D. (1997) *Trends in Cell Biology* 7, 168-173.
- 15 – <http://www.expasy.ch/>
- 16 – <http://www.anl.gov/BIO/PMG/text.html>
- 17 – Steiner S, Anderson NL (2000) *Ann N Y Acad Sci* 919, 48-51.
- 18 – Anderson, NG, Matheson, A, Anderson, NL (2001) *Proteomics* 1, 3-12.
- 19 – Groß M. (1999) *La Recherche* 317, 82-86.
- 20 – Fulda S, Huang F, Nilsson F, Hagemann M, Norling B (2000) *Eur. J. Biochem* 267, 5900-7.
- 21 – Joubert J., Brignon P., Lehman C., Mouribot C., Gendré F., Boucherie H. (2000) *Yeast* 16, 511-22.
- 22 – S.R. Pennington and M.J.Dunn (Ed.s) "Proteomics: from protein sequence to function", 2001, Bios Scientific Pub. Ltd.

Notas

- ¹ Splicing – remoção das partes não codificadas do mRNA. Ocorre antes da tradução da proteína.
- ² Editing – alteração do mRNA por inserção de novos nucleótidos ou por deleção de alguns existentes.
- ³ Estudos preliminares mostraram que, em média, nas bactérias, um simples gene codifica uma ou duas manchas num gel 2D; nas leveduras, um gene codifica cerca de 3 manchas e nos seres humanos, um gene pode codificar entre 3 a mais de 6 manchas. [1,12]

Equipamento de Laboratório
 Balanças - Centrífugas - Aparelhos de pH - Tituladores
 Condutímetros - Agitadores - Espectrofotómetros
 Microscópios - etc.

Vidros e Plásticos de Laboratório
 Distribuidores NORMAX

Material Didático
 Ensino Secundário e Superior
 Representantes exclusivos SISTEDUC - Sistemas Educativos S.A.

Rua Soeiro Pereira Gomes, 15 r/c Frente
 Bom Sucesso - 2615 Alverca
 Telef. (01) 957 04 20/1/2 - Fax (351-1-957 04 23) - Portugal