UNIVERSIDADE ESTADUAL DE CAMPINAS

Marcus Bustamante Smolka

Análise Proteômica da *Xylela fastidiosa* e Desenvolvimento de Novo Método para Análise Quantitativa de Proteomas por Espectrometria de Massa

Tese apresentada ao Instituto de Biologia para obtenção do Título de Doutor em Biologia Funcional e Molecular, na área de Bioquímica

Orientador: Prof. Dr. José Camillo Novello

2002

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA - UNICAMP

Smolka, Marcus Bustamante

Sm78a Análise proteômica da *Xylela fastidiosa* e desenvolvimento de novo método para análise quantitativa de proteomas por espectrometria de massa/Marcus Bustamante Smolka. – Campinas, SP:[s.n], 2002.

Orientador: José Camillo Novello
Tese (Doutorado) – Universidade Estadual de Campinas. Instituto de Biologia

1.Espectrometria de massa. 2.Proteína. 3.Eletroforese. I. Novello, José Camillo. II. Universidade Estadual de Campinas. Instituto de Biologia. III.Título.

Data da Defesa: 05/12/2002

Banca Examinadora:

Prof. DrJosé Camillo Novello (Orientador) _____

Profa. Dra. Eliana Gertrudes Macedo Lemos _____

Prof. Dr. Marcos Antonio Machado _____

Prof. Dr. Marcos Nogueira Eberlin _____

Profa. Dra. Yoko Bomura Rosato _____

Prof. Dr. Sérgio Marangoni _____

Profa. Dra. Nilce Corrêia Meirelles _____

Aos meus pais,

os grandes incentivadores.

À minha esposa Beatriz,

pelo amor.

Agradecimentos

Ao Prof. Dr. José Camillo Novello, pela oportunidade, confiança e orientação. Ao Prof. Dr. Sérgio Marangoni, pelo apoio e confiança. Aos dois, pela amizade.

Ao Prof. Dr. Ruedi Aebersold, pela oportunidade nos EUA e grandes ensinamentos na área de análise de proteomas.

Aos Profs. Denise Macedo, Hiroshi Aoyama e Nilce Corrêia pela participação no exame de qualificação.

Aos colaboradores do projeto: Prof. Dr. Marcos Machado, Profa. Dra. Eliana Lemos, Prof. Dr. Silvio Lopes e Dra. Patricia Monteiro.

Aos grandes amigos e companheiros de trabalho Flávia Winck, Daniel Martins, Carlos Santoro, Paula Oblessuc, Itaraju Brum, Rafael Castellari e Fernanda Ferrari, pela amizade e pelo excelente trabalho em equipe.

Ao Paulo Baldasso pela amizade e pela sempre pronta ajuda.

À Bayki, Gláucia, Zeca, Adriana, Silvana, Luzia, Danielas, Bruno, Marinho, Marcos e demais amigos do LAQUIP.

À FAPESP por viabilizar os sonhos.

Este trabalho foi desenvolvido no Laboratório de Química de Proteínas (LAQUIP) do Departamento de Bioquímica da UNICAMP, Campinas, SP, Brasil, sob orientação do Prof. Dr. José Camillo Novello; e no *Institute for Systems Biology* em Seattle, EUA, sob supervisão do Prof. Dr. Ruedi Aebersold.

Índice

Resumo

A bactéria *Xylella fastidiosa* é responsável por várias doenças de plantas economicamente importantes, incluindo a clorose variegada do citros (CVC). Apesar de seu genoma já estar completamente sequenciado, os estudos das proteínas expressas pela bactéria são quase inexistentes.

Na presente tese, foi utilizada a eletroforese de duas dimensões em combinação com a espectrometria de massas para identificar os produtos de 142 diferentes genes da *X. fastidiosa*, incluindo proteínas secretadas e as relacionadas com adesão, que apresentam particular interesse para o estudo da patogenicidade. Análise comparativa do perfil protéico da bactéria crescida como colônias isoladas em meio sólido de cultura com o da bactéria formando biofilme em meio líquido de cultura indicou a expressão específica de uma cisteíno protease, uma lipase e uma isoforma da subunidade da fímbria do tipo IV na condição de formação de biofilme. Análise do índice de uso de códons específicos ("*codon bias*") das proteínas mais expressas no extrato celular total revelou uma distribuição de valores inesperadamente baixa, o que é proposto como sendo possível causa da natureza fastidiosa da *X. fastidiosa*. Um banco de dados disponibilizando as informações geradas foi construído e pode ser acessado pela internet: URL: www.proteome.ibi.unicamp.br. Os resultados representam um esforço inicial para a caracterização da expressão de proteínas na *X. fastidiosa* e devem ajudar a promover um maior entendimento dos mecanismos de sobrevivência, adesão e secreção, bem como viabilizar subsequentes estudos de análise diferencial do proteoma da *X. fastidiosa* em diversos modelos experimentais.

Na segunda parte da tese é mostrado um novo método para análise quantitativa de proteomas por espectrometria de massas. A estratégia proposta se baseia em marcar duas amostras a serem comparadas com diferentes formas isotópicas do reagente ICAT ("Isotope Coded Afinity Tags"), comigrá-las num único gel 2DE e identificar e quantificar as proteínas por espectrometria de massas. Os resultados mostram que proteínas rotuladas com as formas isotopicamente pesada e normal do reagente ICAT comigram na mesma posição em géis 2DE. Quando as proteínas são subsequentemente digeridas e analisadas por espectrometria de massas, a razão da abundância relativa das proteínas em cada amostra pode ser determinada precisamente, baseando-se nas intensidades relativas do sinal espectrométrico dos peptídios contendo as diferenças isotópicas. Utilizando esta estratégia, foi possível realizar quantificações com precisão melhor que 20 %.

Abstract

The bacteria *Xylella fastidiosa* is the causative agent of a number of economically important crop diseases, including citrus variegated chlorosis (CVC). Although its complete genome is already sequenced, *X. fastidiosa* is poorly characterized by biochemical approaches at the protein level.

In the present work it was used two dimensional electrophoresis in combination with mass spectrometry to identify the products of 142 different *X. fastidiosa* genes, including secreted and adhesion related proteins that show particular interest for the study of pathogenesis. Comparative proteome analysis of the bacteria grown as isolated colonies on a solid culture medium and the bacteria forming biofilm on a liquid culture medium indicated the specific expression of a cysteine protease, a lipase and an isoform of the type IV fimbriae subunit in the biofilm forming condition. A codon usage analysis of the most expressed proteins from the whole cell extract revealed a low biased distribution, which we propose to be related with the slow growing nature of *X. fastidiosa*. A database of the *X. fastidiosa* proteome was developed and can be accessed via the internet (URL: www.proteome.ibi.unicamp.br). The results represent an initial effort to characterize protein expression in *X. fastidiosa* and should contribute for a better understanding of the survival, adhesion and secretion mechanisms. It also provides the basis for subsequent differential proteome analysis on various experimental models.

The second part of the thesis presents a new method for quantitative proteome analysis using mass spectrometry. The proposed strategy is based on labelling two different samples to be compared with the different istopic forms of the ICAT (Isotope Coded Afinity Tags) reagent, combine and comigrate them in a single 2DE gel and identify and quantify the proteins by mass spectrometry. It is shown that proteins labeled with the isotopically heavy and normal forms of the reagent co-migrate in 2DE gels and that the ratio of protein abundance can be accurately determined from the relative mass spectrometric signal intensities of the heavy and normal forms of labeled, cysteine-containing peptides. Both identification and accurate quantification (better than 20 %) are therefore achieved in a simple single-step analysis.

Introdução Geral

1. *Xylella fastidiosa*

A *Xylella fastidiosa* (Figura 1) é uma bactéria Gram-negativa que habita o xilema de certas plantas. A sua ocorrência está associada a doenças economicamente importantes, como a clorose variegada do citros (CVC ou amarelinho) em laranjas doces[1,2], doença de Pierce em videiras[3], escaldadura de folhas da amendoeira e de outras espécies como café e ameixa [4]. Em todos os casos, *X. fastidiosa* infecta o xilema da planta e compromete a produção de frutas.

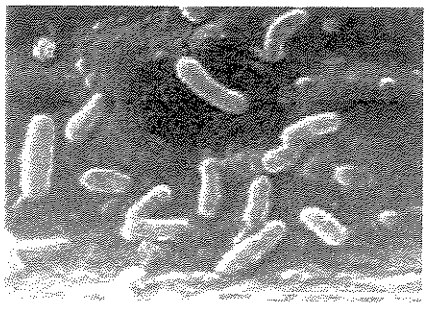


FIGURA 1: Micrografia eletrônica da bactéria *Xylella fastidiosa* (Fonte: Fundecitrus).

A célula bacteriana não é flagelada, apresenta forma de bastonete e dimensões de 0.25 a 0.35 µm de largura por 0.9 a 3.5 µm de comprimento. De natureza fastidiosa, a *X. fastidiosa* tem por característica marcante a exigência de meios complexos para seu desenvolvimento em meio de cultura e tempos longos de duplicação (acima de 9 horas)[5,6]. Ela é estritamente aeróbia, não fermentativa, apigmentada, com temperatura

[1] Rossetti, V., Garnier, M., Bové, J. M., Beretta, M. J. et al. Presence de bacteries dans le xyleme d'orangers atteints de chlorose variegee, une nouvelle maladie des agrumes au Bresil. *C. R. Acad. Sci. Paris* 1990, 310, 345-349.

[2] Chang, C. J., Garnier, M., Zreiki, L., Rossetti, V. et al. Culture and serological detection of the xylem-limited bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. *Curr. Microbiol.* 1993, 27, 137-142.

[3] Davis, M. J., Purcell, A. H., Thomson, S. V. Pierce's disease of grape-vines: isolation of the causal bacterium. *Science* 1978, 199, 75-77.

[4] Purcell, A. H., Hopkins, D. L. Fastidious xylem-limited bacterial plant pathogens. *Annu. Rev. Phytopathol.* 1996, 34, 131-151.

[5] Wells, J. M., Raju, B. C., Hung, H., Weisburg, W., G. et al. *Xylella fastidiosa* gen. nov. sp. nov.: Gram-negative, xylem limited, fastidious plant bacteria related to *Xanthomonas spp. Int. J. Syst. Bacteriol.* 1987, 37, 136-143.

[6] Holt, J. G. (ed.) Genus *Xylella*. Em: *Bergey's Manual of Determinative Bacteriology*, Wiliams & Wilkins, 1994, pp.100-115.

ótima de desenvolvimento de 26 a 28 °C e pH ótimo de 6.5 a 6.9. Tem capacidade de hidrolisar gelatina, produzir catalase e apresenta reação negativa para os testes de indol, H_2S, β-galactosidase, lipase, amilase, coagulase e fosfatase[5].

Até o momento, sua classificação taxonômica não é bem definida, sendo relacionada ao grupo das *Xanthomonas spp.* Apesar de terem sido classificadas em uma espécie única, as variedades de *Xylella fastidiosa* apresentam diferenças entre si que justificam uma separação taxonômica sub-específica (revisado por Hopkins, 1989[7]).

1.1. Clorose Variegada do Citros

Todas as variedades de laranja doce são suscetíveis à CVC, que causa, basicamente, diminuição do tamanho de folhas e frutos, endurecimento de frutos, redução na velocidade de crescimento das plantas, aparecimento de cloroses amareladas nas folhas e degeneração da copa das plantas (Figura 2). Na maioria dos casos, frutas de plantas infectadas não podem ser usadas para fins comerciais e industriais.

Os vetores são homópteros (cigarrinhas) que se alimentam do xilema, sendo que no Brasil, a bactéria foi encontrada associada com espécies como *Oncometopia facialis, Acrogonia terminalis* e *Dilobopterus costallimai.* Ao se alimentarem da seiva do xilema de árvores contaminadas, as cigarrinhas também sugam a bactéria e passam a transmití-la para outras plantas sadias.

[7] Hopkins, D. L. *Xylella fastidiosa*: xylem limited bacterial pathogen of plants. *Ann. Rev. Phytopathol.* 1989, 27, 271-290.

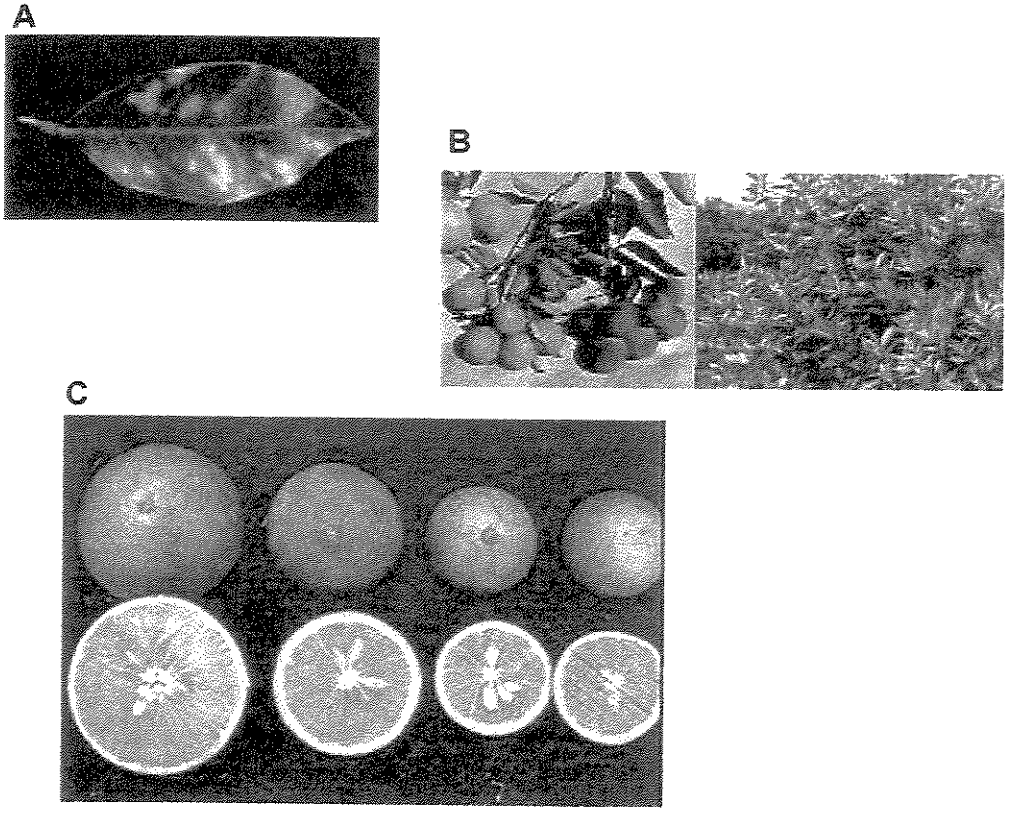


FIGURA 2: Sintomas da CVC. (A) A clorose aparece nas folhas maduras da copa. (B) No estágio avançado da doença, nota-se a presença de ramos secos com frutos pequenos. (C) Os frutos vão ficando queimados pelo sol, com tamanho reduzido, endurecidos e com maturação precoce. Estão imprestáveis para o comércio (Fonte: Fundecitrus).

A CVC foi primeiramente detectada no Brasil em 1987[1,2,8] e hoje representa uma grande ameça à indústria brasileira de citros, estando presente nas principais áreas de cultivo e sendo responsável por perdas significativas na produção de laranja. Estudos apontam alta incidência da CVC no estado de São Paulo, principalmente nas regiões Norte, Nordeste e Central do estado. Em 1996, foram encontradas plantas com sintomas severos de CVC em 88% das plantações[9]. Além do Brasil, a Argentina também registra ocorrência da doença[10].

[8] Hartung, J. S., Beretta, J., Brlansky, R. H., Spisso, J. et al. Citrus variegated chlorosis bacterium: axenic culture, pathogenicity, and serological relationships with other strains of *Xylella fastidiosa*. *Phytopathology* 1994, 84, 591-597.

[9] Machado, M. A. Citrus variegated chlorosis (CVC), a new destructive citrus disease in Brazil, and the xylem-limited bacteria, *Xylella fastidiosa*. URL: www.watson.fapesp.br/Xylella.htm. 1997.

[10] Brlansky, R. H., Davis, C. L., Timmer, L. W., Howd, D. S. et al. Xylem-limited bacteria in citrus from Argentina with symptoms of citrus variegated chlorosis. *Phytopatholgy* 1991, 81, 1210-1214.

Dentre os métodos atualmente utilizados para manejo da doença estão a poda de galhos infectados, controle químico dos vetores e utilização de mudas sadias[9,11].

1.2. Mecanismos de Colonização e Patogenicidade

Muitos dos mecanismos de colonização e patogenicidade atualmente discutidos já vêm sido propostos desde a década de 80, e são baseados principalmente em estudos com a doença de Pierce em videiras (revisado por Hopkins, 1989[7]). Consideráveis desavenças existem entre os cientistas sobre quais seriam os fatores preponderantes desencadeadores dos sintomas observados, porém, várias evidências apontam para uma disfunção do sistema condutor de água. Nesse mecanismo proposto, agregados de bactérias, goma e tiloses presentes no xilema obstruiriam a passagem de água. A formação de grandes agregados de *X. fastidiosa* já foi demonstrada [4,7] (Figura 3), mas não se sabe ao certo se ocorre oclusão suficiente para gerar estresse hídrico. Outros fatores como a liberação de fitotoxinas e desbalanço da regulação do crescimento poderiam estar também envolvidos[7,9].

O xilema das plantas hospedeiras é um ambiente com características nutricionais e físicas singulares para o crescimento bacteriano. Para sobreviver no meio hiponutricional do xilema, multiplicar-se no ponto de infecção e se espalhar para outras regiões do sistema condutor, a *X. fastidiosa* parece se utilizar de mecanismos específicos, ainda pouco conhecidos em detalhe. A bactéria se agrega nos vasos do xilema através de goma polissacarídica e provavelmente também por estruturas protéicas como fímbrias. É possível que essas fibras de polissacarídios, na maioria negativamente carregadas, funcionem como uma resina trocadora de cátions, ligando nutrientes iônicos e disponibilizando-os de forma concentrada para a bactéria, o que seria um importante mecanismo de sobrevivência em um ambiente onde os nutrientes estão muito diluídos. Além disso, a goma serviria como proteção contra defesas do hospedeiro[7].

[11] Manual de convivência com a CVC. *Fundecitrus*, 1997.

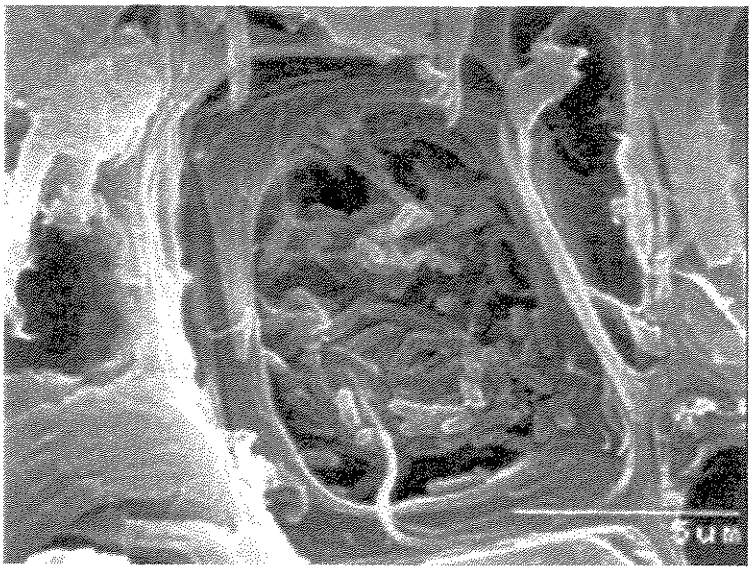


FIGURA 3: Micrografia eletrônica de vaso do xilema de citros obstruído por agregados de *X. fastidiosa* (Fonte: Fundecitrus).

Foi recentemente demonstrado que a *X. fastidiosa* tem a capacidade de formar biofilme[12]. O biofilme pode ser definido como comunidades estruturadas de agregados microbianos cobertas por uma matriz polimérica e aderidas a uma determinada superfície[13]. Esta forma de organização parece ser um mecanismo bastante importante de sobrevivência da bactéria na planta e no inseto. Foi proposto que a formação de biofilme pela *X. fastidiosa* é um fator crítico para sua virulência em plantas, tendo um papel decisivo na oclusão de vasos do xilema[12].

Outro aspecto importante é o movimento sistêmico, isto é, como a bactéria se espalha para colonizar o hospedeiro. O xilema é constituído de um conjunto de vasos, que são divididos entre si por membranas com poros que não são grandes o suficiente para passagem de bactéria. Com isso, a *X. fastidiosa* teria que passar de alguma maneira por esta membrana para se movimentar de vaso para vaso, o que poderia envolver rompimento enzimático da membrana, que é formada basicamente por celulose, hemicelulose, pectina, lignina e proteina. Dessa forma, proteínas com atividade de degradação destes compostos desempenhariam papel fundamental no processo de colonização. Os agregados de bactéria poderiam conservar e concentrar

[12] Marques, L. R., Ceri, H., Manfio, G. P., Reid, D. M. Characterization of biofilm formation by *Xylella fastidiosa*. *Plant Dis.* 2002, 86, 633-638.

[13] Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. et al. Microbial biofilms. *Annu. Rev. Microbiol.* 1995, 49, 711-745.

enzimas digestivas produzidas pela bactéria e direcioná-las contra o tecido hospedeiro. Muito pouco, porém, é conhecido sobre isso.

1.3. Genoma

Em 2000, um consórcio de laboratórios paulistas financiado pela FAPESP (Fundação de Apoio à Pesquisa do Estado de São Paulo) publicou a sequência do genoma da estirpe 9a5c da *X. fastidiosa* causadora da CVC[14]. Além do cromossomo principal (com 2.679.305 pares de base), outros dois plasmídios (um com 51.158 e outro com 1.285 pares de base) foram sequenciados. Ao todo, foram identificados 2905 genes, sendo que aproximadamente metade apresentou significante similaridade com proteínas de função conhecida (Tabela 1).

O conhecimento da sequência completa do genoma representou um grande passo para compreensão das características metabólicas e replicativas, e para a determinação dos potenciais mecanismos de patogenicidade. Vários artigos têm explorado as informações geradas pelo sequenciamento genômico, destacando uma série de hipóteses para o funcionamento do metabolismo energético, transporte de nutrientes, adesão, agregação, toxicidade, secreção de fatores de patogenicidade, interações intercelulares, homeostase de ferro, resposta antioxidante e outros mecanismos potencialmente importantes para a patogenicidade[14,15,16,17,18,19,20].

[14] Simpson, A. J. G., Reinach, F. C., Arruda, P., Abreu, F. A. et al. The genome sequence of the plant pathogen Xylella fastidiosa. *Nature* 2000, 406, 151-158.

[15] Keen, N. T., Dumenyo, C. K., Yang, C. H., Cooksey, D. A. From rags to riches: insights from the first genomic sequence of a plant pathogenic bacterium. *Genome Biol.* 2000, 1, 10191-10194.

[16] Dow, J. M., Daniels, M. J. Xylella genomics and bacterial pathogenicity to plants. *Yeast* 2000, 17, 263-271.

[17] Lambais, M. R., Goldman, M. H. S., Camargo, L. E. A., Glodman, G. H. A genomic approach to the understanding of Xylella fastidiosa pathogenicity. *Curr. Opin. Microbiol.* 2000, 3, 459-462.

[18] Silva, F. R., Vettore, A. L., Kemper, E. L., Leite, A. et al. Fastidian gum: the Xylella fastidiosa exopolysaccharide possibly involved in bacterial pathogenicity. *FEMS Microbiol. Lett.* 2001, 203, 165-171.

[19] Leite, B., Ishida, M. L., Alves, E., Carrer, H. et al. Genomics and X-ray microanalysis indicate that Ca2+ and thiols mediate the aggregation and adhesion of Xylella fastidiosa. *Braz. J. Med. Biol. Res.* 2002, 35, 645-650.

[20] Meidanis, J., Braga, M. D. V., Verjovski-Almeida, S. Whole-genome analysis of transporters in the plant pathogen Xylella fastidiosa. *Microbiol. Mol. Biol. Rev.* 2002, 66, 272-299.

Tabela 1. Características gerais do genoma da linhagem 9a5c da *X. fastidiosa* (modificado de Simpson et al.[14]).

Cromossomo principal	
Tamanho (pares de base)	2.679.305
Razão G + C	52,7 %
Número de genes	2838
genes com função conhecida	1283
genes com similaridade a proteínas sem função conhecida	310
genes sem similaridade significante	1083
Plasmídio pXF51	
Tamanho (pares de base)	51.158
Razão G + C	49,6 %
Número de genes	65
genes com função conhecida	31
genes com similaridade a proteínas sem função conhecida	8
genes sem similaridade significante	24
Plasmídio pXF1.3	
Tamanho (pares de base)	1.285
Razão G + C	55,6 %
Número de genes	2
genes com função conhecida	1
genes com similaridade a proteínas sem função conhecida	0
genes sem similaridade significante	1

O banco de dados do genoma (acessado pelo endereço eletrônico http://aeg.lbi.ic.unicamp.br/xf) representa uma fonte rica de informações que agora serve de base para o avanço mais rápido dos estudos da *X. fastidiosa* e sua patogenicidade.

2. Proteômica

O conhecimento que vem se acumulando com os sequenciamentos de genomas demonstra que é preciso caminhar mais adiante. Na era pós-genômica, tem ficado evidente que apenas com sequências genômicas completas não é suficiente para elucidar muitos dos processos biológicos [21]. Adicionalmente às informações obtidas

[21] Pandey, A., Mann, M. Proteomics to study genes and genomes. *Nature* 2000, 405, 837-846.

pela análise do genoma, é preciso ter conhecimento, por exemplo, sobre quais proteínas estão realmente sendo expressas, quando e em quais níveis esta expressão ocorre e sobre as eventuais modificações pós-transcricionais. O fato das células de um organismo terem o mesmo genoma mas apresentarem as mais variadas funções e morfologias, fruto de diferentes composições de proteínas, ilustra a importância de se estudar a regulação da expressão gênica para compreender a função biológica.

O termo proteômica foi inicialmente introduzido em 1995 e foi definido como sendo a caracterização em larga escala do conjunto de proteínas expresso numa célula ou tecido, isto é, o proteoma[22,23,24,25]. Diferentemente do genoma, o proteoma tem natureza dinâmica e se altera frente a diferentes condições e estímulos. O estudo global destas alterações representa uma forma de investigar e integrar processos em sistemas vivos para melhor entender o funcionamento de uma célula ou tecido ao nível molecular.

A proteômica visa estudar a estrutura, função e o controle dos sistemas biológicos pela análise das várias propriedades das proteínas. Isto inclui a sequência (identidade), abundância, atividade e estrutura das proteínas expressas por uma célula, assim como as modificações, interações e translocações sofridas por cada proteína. Apenas a análise direta das proteínas (as principais moléculas efetoras de atividade biológica) possibilita o estudo de muitas destas propriedades estruturais e funcionais.

Analogamente à proteômica, a transcriptômica se baseia na análise global de RNAs mensageiros (mRNA) presentes numa célula ou tecido. Neste caso, informações precisas sobre regulação transcricional são geradas, o que representa outro tipo de informação, pois mesmo que uma proteína não possa ser sintetizada sem seu mRNA estar presente, é possível que uma proteína esteja na célula quando seu mRNA não está mais presente e, ainda, é possível haver muito mRNA sem que haja tradução da mensagem em proteína[26]. Devido a mecanismos pós traducionais de regulação de expressão gênica, a quantidade de proteína não é necessariamente proporcional a quantidade de seu mRNA correspondente.

---

[22] Wilkins, M. R., Pasquali C., Appel R. D., Ou K. et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology* 1996, 14, 61-65.

[23] Celis, J. E., Gromov, P., Ostergaard, M., Madsen, P., et al. Human 2-D PAGE databases for proteome analysis in health and disease: http://biobase.dk/cgi-bin/celis. *FEBS Lett.* 1996, 398, 129-134.

[24] Anderson, N. G., Anderson, N. L. Twenty years of two-dimensional electrophoresis: past, present and future. *Electrophoresis* 1996, 17, 443-453.

[25] Williams, K. L., Hochstrasser, D. F. Introduction to the proteome. Em: *Proteome Research: New Frontiers in Functional Genomics* (Wilkins, M. R., Williams, K. L., Apple, R. D., Hochstrasser, D. F., eds.), Springer, 1997, pp. 1-12.

[26] Anderson, L., Seilhamer, J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 1997,18, 533-537.

A identificação e quantificação sistemática das proteínas é geralmente o objetivo inicial central numa análise proteômica. A tecnologia mais aplicada atualmente é a combinação da eletroforese de duas dimensões em gel de poliacrilamida (2DE) para separação, detecção e quantificação das proteínas, com espectrometria de massas para identificação das proteínas separadas. Enquanto a técnica de 2DE foi inicialmente introduzida na década de 70[27,28], a maioria dos métodos que possibilitou identificar proteínas separadas em 2DE de forma rápida, conclusiva e com sensibilidade foi apenas desenvolvida nos últimos 10 anos[29]. Os principais fatores que possibilitaram o surgimento destes métodos foram avanços na técnica de espectrometria de massas, disponibilidade de sequências genômicas completas e desenvolvimento de programas computacionais para correlação dos dados obtidos por espectrometria de massas com os bancos de dados de sequências.

Atualmente, identificação e quantificação de proteínas em larga escala e comparações de proteomas por análise diferencial de perfis 2DE representam importantes ferramentas para detectar e identificar proteínas efetoras ou controladoras de processos biológicos.

2.1. Eletroforese de Duas Dimensões

A 2DE é um dos métodos disponíveis mais poderosos para separar misturas complexas de proteínas. Ela combina duas técnicas de separação baseadas em gel de poliacrilamida (Figura 4). A primeira separação (primeira dimensão) é a focalização isoelétrica (IEF), na qual as proteínas são separadas em um gradiente de pH até alcançarem a posição estacionária onde a carga total é zero (ponto isoelétrico). Na segunda dimensão, as proteínas separadas pela IEF são então separadas por eletroforese em gel de poliacrilamida contendo dodecil sulfato de sódio (SDS-PAGE), separação esta baseada na massa molecular (MM) das proteínas. As proteínas são então visualizadas diretamente no gel, resultando num perfil bidimensional de pontos (chamados "spots"), cada um contendo múltiplas cópias de uma proteína (Figura 4). A quantidade de proteína migrando em cada *spot* é quantificada baseando-se na intensidade e área do respectivo *spot*. A imagem do gel é então digitalizada e analizada

[27] Klose, J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* 1975, 26, 231-243.
[28] O'Farrel, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 1975, 250, 4007-4021.
[29] Aebersold, R., Goodlett, D. Mass spectrometry in Proteomics. *Chem. Rev.* 2001, 101, 269-295.

densitometricamente por programas de computador. Diferentes géis representando diferentes proteomas podem ser comparados com o objetivo de detectar proteínas diferencialmente expressas.

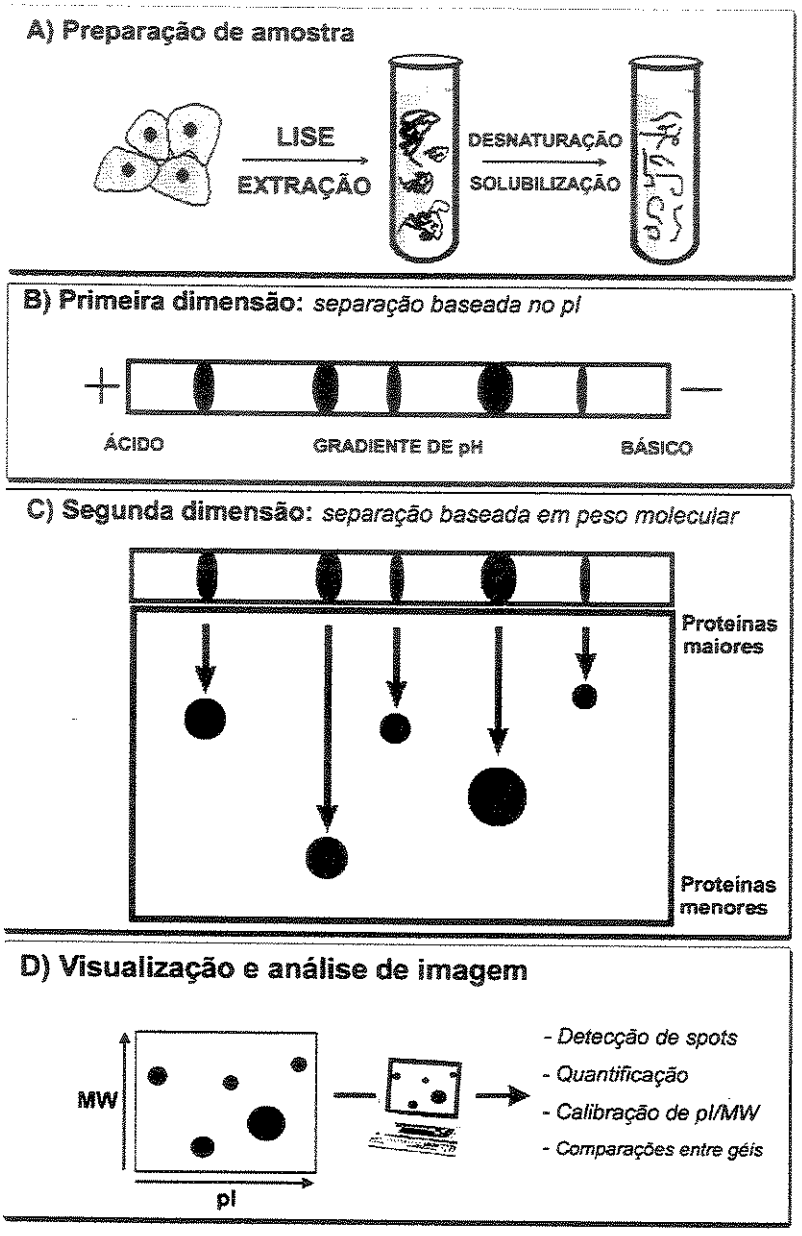


FIGURA 4: Passos do procedimento da 2DE. (A) Preparação de amostra: proteínas são primeiramente extraídas das células (ou tecidos) e solubilizadas. (B) Separação na primeira dimensão é feita por IEF, onde as proteínas são separadas pelo seu ponto isoelétrico. (C) Separação na segunda dimensão é feita por SDS-PAGE, que separa proteínas pela massa molecular aparente. (D) As proteínas separadas são então coradas diretamente no gel, resultando num mapa de pontos, *spots*. A quantidade de proteína migrando em cada *spot* é quantificada baseando-se na intensidade de coloração do respectivo *spot*. Geralmente, a imagem do gel é digitalizada e analizada densitometricamente por programas de computador.

Apesar da 2DE ter o potencial de separar centenas a milhares de proteínas e ser a técnica mais utilizada para a separação de misturas complexas de proteínas, ela também apresenta significantes limitações. Classes específicas de proteínas como as muito ácidas (pI < 3.5), muito básicas (pI > 9.0), pequenas (< 6000Da), muito hidrofóbicas e muito grandes (> 120000Da), bem como as proteínas de baixa abundância, geralmente não são detectáveis[30,31]. Porém, para análise das proteínas que são visíveis, a 2DE é um método de separação poderoso porque todas as proteínas numa amostra são simultaneamente separadas. Além disso, informação sobre pI, massa molecular e abundância relativa pode ser obtida para todas as proteínas detectadas, estando elas ainda passíveis para análises pós separação, como por exemplo espectrometria de massas. A presença de modificações pós-traducionais é geralmente aparente devido ao fato das modificações provocarem mobilidade eletroforética alterada.

2.2. Métodos de Detecção e Quantificação

Vários métodos diferentes de detecção de proteínas separadas por eletroforese em gel de poliacrilamida podem ser usados. A maioria deles se baseia em coloração de proteínas com corantes ou reagentes fluorescentes, ou detecção de proteínas marcadas radioativamente por fluorografia ou autoradiografia[32,33]. Os métodos mais comuns são os de coloração com coomassie blue (CBB) ou nitrato de prata. Ambos são procedimentos simples e de baixo custo, podendo ser compatíveis com análises subsequentes por espectrometria de massas. Coloração por prata é um dos métodos mais sensíveis (1-10 ng/*spot*), mas é difícil de ser usado para quantificações precisas pois a quantidade de prata ligada é específica para cada proteína e porque a intensidade de coloração é dependente do tempo de revelação, tornando difícil comparações quantitativas entre géis[34,35,36]. Coloração com CBB apresenta linearidade

[30] Urquhart, B. L., Cordwell, S. J., Humphery-Smith, I. Comparison of predicted and observed properties of proteins encoded in the genome of Mycobacterium tuberculosis H37Rv. *Biochem. Biophys. Res. Commun.* 1998, 253, 70-79.

[31] Gygi S. P., Corthals G. L., Zhang Y., Rochon Y. et al. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. USA* 2000, 15, 9390-9395.

[32] Patton, W. F. A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis*, 2000, 21, 1123-1144.

[33] Patton, W. F. Detecting proteins in polyacrylamide gels and electroblot membranes. Em: *Proteomics: from protein sequence to function* (Pennington, S. R. and Dunn, M. J., eds.), Bios, 2001, pp. 31-59.

[34] Rabilloud, T. Mechanisms of protein silver staining in polyacrylamide gels: a 10-year synthesis. *Electrophoresis* 1990, 11, 785-794.

[35] Rabilloud, T. Silver staining of 2-D electrophoresis gels. *Methods Mol. Biol.* 1999, 112, 297-305.

[36] Rabilloud, T. A comparison between low background silver diammine and silver nitrate protein stains. *Electrophoresis* 1992, 13, 429-439.

quantitativa na faixa de concentração de proteína de 100-1000 ng/*spot*, porém, a sensibilidade é 10 a 50 vezes mais baixa que coloração por prata [32,37].

Métodos baseados em fluorescência e marcação radioativa são bastante sensíveis e quantitativos por uma maior faixa de abundância (faixa dinâmica). Porém, tendem a ser mais caros e, no caso de marcação radioativa, limitados a amostras que podem ser marcados radioativamente (geralmente por marcação metabólica).

Para análise de imagem, os géis são digitalizados e os arquivos gerados são analizados em programas especializados que detectam e quantificam densitometricamente os *spots* e calculam seus respectivos pI e massa molecular observados. Análise comparativa de diferentes géis é usado para detectar proteínas diferencialmente expressas. A determinação de pI e massa molecular é importante porque estes parâmetros são úteis para a confirmação da identidade da proteína determinada por MS ou outros métodos. O sucesso da análise de imagem depende fundamentalmente da qualidade e reprodutibilidade dos perfis 2DE.

Atualmente, 2DE é o método mais utilizado para quantificação comparativa dos níveis de abundância de uma proteína em diferentes amostras. A quantificação por análise densitométrica implicitamente assume que a intensidade de coloração de um *spot* indica precisamente a quantidade de proteína presente no *spot*. Porém, a implementação prática deste simples conceito tem se mostrado difícil, principalmente devido a limitações de sensibilidade, linearidade quantitativa e reprodutibilidade dos métodos disponíveis de coloração e separação eletroforética.

Novos métodos têm procurado eliminar, ou ao menos aliviar, muitos dos desafios da quantificação por análise de imagem. Métodos sensíveis de coloração por fluorescência apresentam uma resposta linear por uma maior faixa dinâmica[38,39]. Mais recentemente, a introdução do sistema de marcação com fluorescência de duas cores possibilitou a comigração eletroforética no mesmo gel 2DE de duas amostras marcadas diferentemente e assim a determinação da razão de abundância de proteínas em cada amostra por análise em diferentes espectros de onda do mesmo gel[40,41]. Neste caso, o problema de reprodutibilidade entre géis é completamente eliminado.

---

[37] Dunn, M. J. Two-dimensional polyacrylamide gel elctrophoresis. Em: *Advances in Electrophoresis* (Chrambach, A., Dunn, M. J., Radola, B. J., eds.), VCH, 1987, pp. 1-109.

[38] Lopez, M. F., Berggren, K., Chernokalskaya, E., Lazarev, A., et al. A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* 2000, 21, 3673-3683.

[39] Steinberg, T. H., Jones, L. J., Haugland, R. P., Singer, V. L. SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Anal. Biochem.* 1996, 239, 223-237.

[40] Unlu, M., Morgan, M. E., Minden, J. S. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 1997, 18, 2071-2077.

2.3. Identificação de Proteínas por Espectrometria de Massas (MS)

A identificação de proteínas presentes nos perfis de 2DE é um passo importante na análise de proteomas. De uma técnica puramente descritiva, a 2DE transforma-se numa poderosa ferramenta integrada para análise estrutural e funcional de proteínas. Quando combinada à técnica de MS, a identificação é geralmente feita por correlação de dados experimentais obtidos de uma proteína com os conteúdos de bancos de dados de seqüências (geralmente de proteínas, cDNA, DNA genômico ou "expressed sequence tags" (EST). Vários tipos de dados podem ser usados para identificar proteínas por busca em bancos de dados, e atualmente existe uma grande variedade de técnicas que podem ser usadas para gerá-los[29]. Na maioria dos casos, a proteína é fragmentada, geralmente por digestão tríptica, e seus fragmentos (não a proteína intacta) são submetidos para análise por MS.

O método de identificação de proteínas utilizando a massa precisa de um número de peptídios derivados da mesma proteína é chamado de "peptide-mass finger printing" (PMF) (Figuras 5 e 6). O aparelho utilizado para PMF é geralmente o MALDI-TOF MS, e programas para PMF já foram desenvolvidos independentemente por vários grupos[42,43,44,45,46].

Na prática, o processo começa por cortar cada *spot* individualmente do gel 2DE. As proteínas nos *spots* são então submetidas à digestão dentro do gel ("in-gel"). Para aplicações de pequena escala, isto é geralmente realizado manualmente, e para projetos de larga escala sistemas robotizados foram desenvolvidos. Após digestão, os peptídios são extraídos do gel e contaminantes residuais como detergentes, sais, corantes e outros, são removidos para evitar interferência com a análise por MS.

No espectrômetro, os peptídios são ionizados, separados segundo suas razões massa/carga (m/z) e então detectados. Em diferentes tipos de espectrômetros de massa os processos de ionização e separação ocorrem por diferentes princípios físicos.

---

[41] Tonge, R., Shaw, J., Middleton, B., Rowlinson, R. et al. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* 2001, 1, 377-396.

[42] Mann, M., Hojrup, P., Roepstorff, P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* 1993, 22, 338-345.

[43] Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C. et al. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* 1993, 90, 5011-5015.

[44] Yates, J. R. 3rd, Speicher, S., Griffin, P. R., Hunkapiller T. Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* 1993, 214, 397-408.

[45] James, P., Quadroni, M., Carafoli, E., Gonnet G. Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun.* 1993, 195, 58-64.

Todos os instrumentos geram espectros de massa que são tipicamente representados como um gráfico de duas dimensões contendo valores de m/z no eixo x e intensidades relativas no eixo y (Figura 6A). Nos espectros gerados em instrumentos de alta resolução, cada peptídio aparece como um conjunto de moléculas protonadas isotopoméricas (Figura 6B), conseqüência do fato de que 1.1 % dos carbonos naturalmente existente é o isotopicamente pesado ^{13}C. Como os peptídios contêm grandes quantidades de átomos de carbono, fração significativa dos peptídios irá conter um ^{13}C (e um valor de massa aumentado em 1 unidade de massa comparado ao íon contendo apenas ^{12}C), ou dois ^{13}C (e um valor de massa aumentado em 2 unidades de massa comparado ao íon contendo apenas ^{12}C), e assim por diante. A escala do eixo x do espectro é definido pelo usuário e tipicamente ajustado entre 500 e 3000 Da, a região de massa mais utilizada para análise de peptídios por MS.

Descrição mais detalhada dos tipos de espectrômetros mais usados, capacidades e limitações do PMF e análises MS/MS, bem como dos algoritmos usados para procura em bancos de dados, pode ser encontrada no Apêndice da presente tese e em outras revisões [29,47,48]

[46] Pappin, D. J. Peptide mass fingerprinting using MALDI-TOF mass spectrometry. *Methods Mol. Biol.* 1997, 64, 165-173.

[47] Patterson, S. D., Aebersold, R. Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* 1995, 16, 1791-1814.

[48] Corthals, G. L., Gygi, S. P., Aebersold, R., Patterson, S. D. Identification of proteins by mass spectrometry. Em: *Proteome Research: 2D Gel Electrophoresis and Detection Methods* (Rabilloud, T., ed.), Springer, 1999, pp. 197-231.

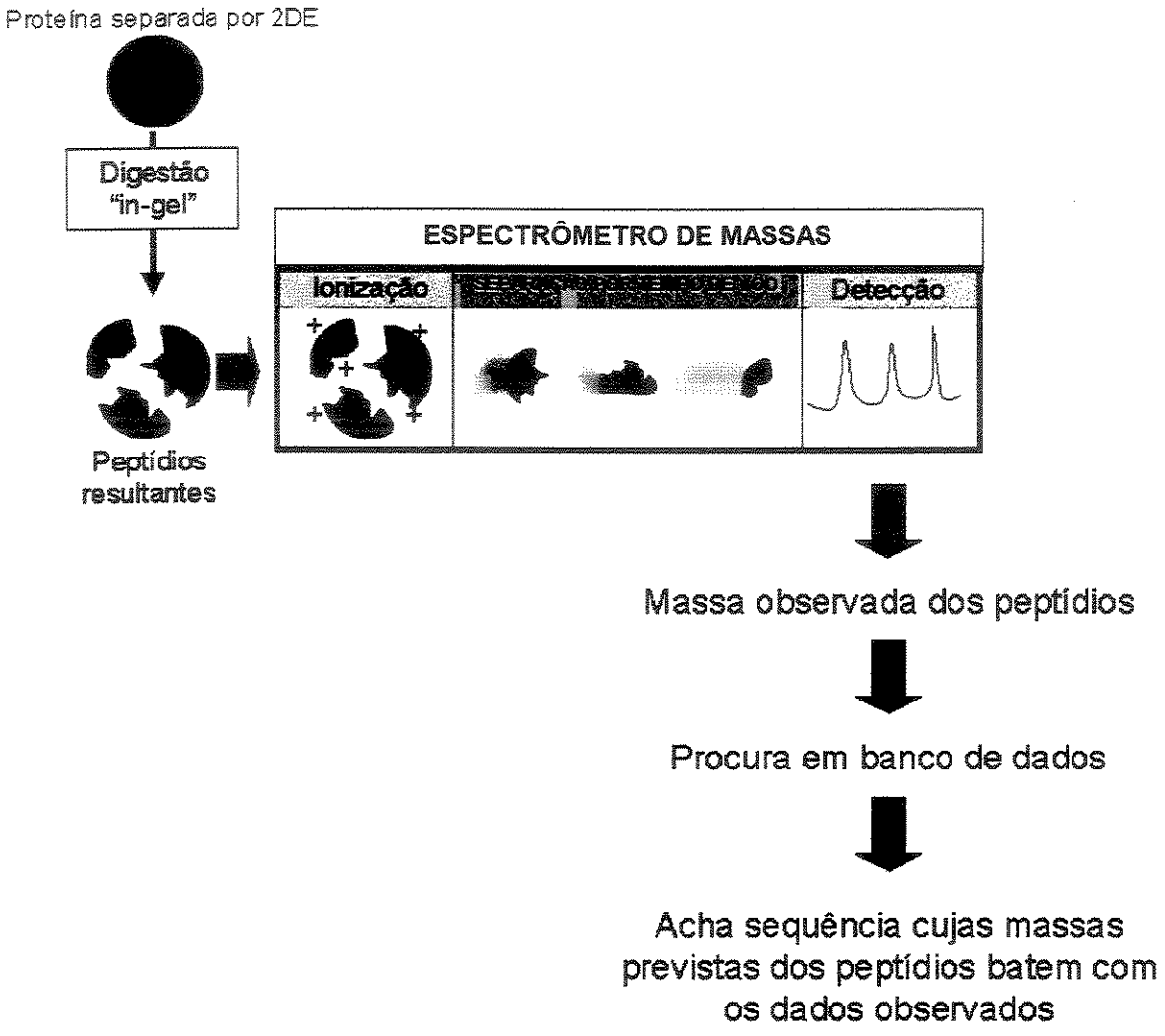


FIGURA 5: Identificação de proteína por PMF. A proteína separada por 2DE é digerida *in gel* (geralmente com tripsina), as massas dos peptídios resultantes são determinadas em espectrômetro MALDI-TOF e usadas para identificar a proteína por procura em bancos de dados de seqüências.

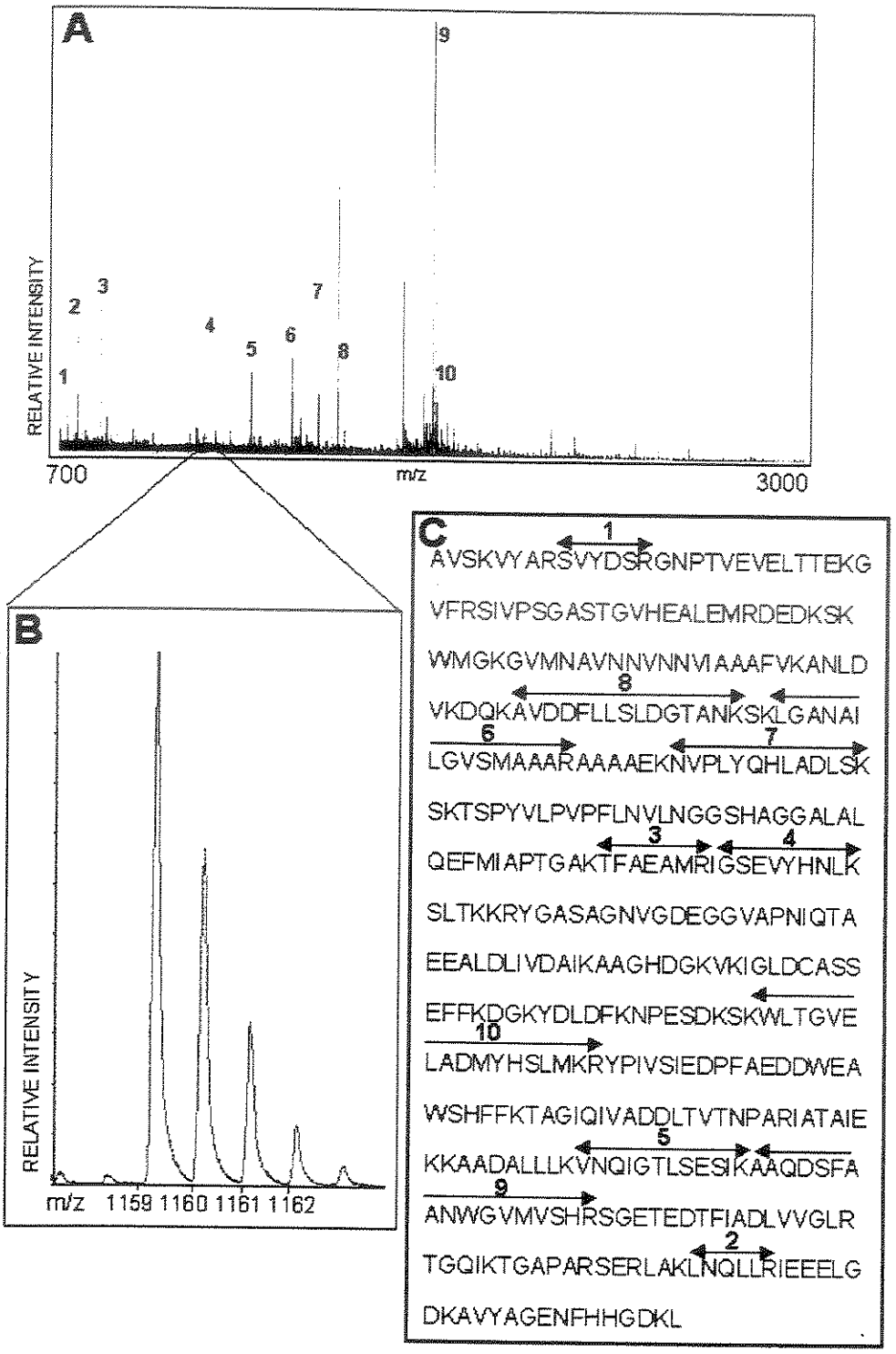


FIGURA 6: Mapeamento de peptídios da proteína enolase de *S. cerevisae*. O *spot* separado em 2DE foi digerido com tripsina e os peptídios resultantes analisados em espectrômetro de massas MALDI-TOF. Procura em bancos de dados indicou a proteína como sendo enolase 2. (A) Espectro de massa dos peptídios trípticos. Os números indicam os peptídios correspondentes na Figura 5C. (B) Ampliação da região entre 1158 e 1164Da mostrando o conjunto de isótopos do peptídio com massa de 1159.66Da. (C) Seqüência de amino ácidos da enolase 2. Note que a tripsina cliva na parte C-terminal da lisina (L) e arginina (R).

2.4. Reagentes ICAT ("Isotope Coded Afinity Tags")

A combinação do procedimento de rotulação de proteínas com reagentes ICAT e análise por espectrometria de massas dos peptídeos resultantes por proteólise é um novo método de análise quantitativa de proteomas[49]. O método permite identificação sistemática das proteínas numa mistura complexa e quantificação precisa das diferenças em abundância de cada proteína presente em duas ou mais amostras protéicas.

O reagente ICAT consiste de três componentes funcionais (Figura 7). O primeiro é um grupamento reativo à tióis, que é seletivo para os grupamentos sulfidrilas das cadeias laterais de cisteínas reduzidas. O segundo componente, é um grupo de etileno glicol que pode conter átomos de deutério (isotopicamente "pesado") ou não (isotopicamente "leve") e que possibilita realizar quantificação por MS. O terceiro componente, um grupamento de biotina, é a "etiqueta de afinidade" utilizado como base para o isolamento seletivo dos peptídios contendo cisteína (marcados com ICAT) de uma mistura complexa de peptídios, através de cromatografia de afinidade em coluna de avidina.

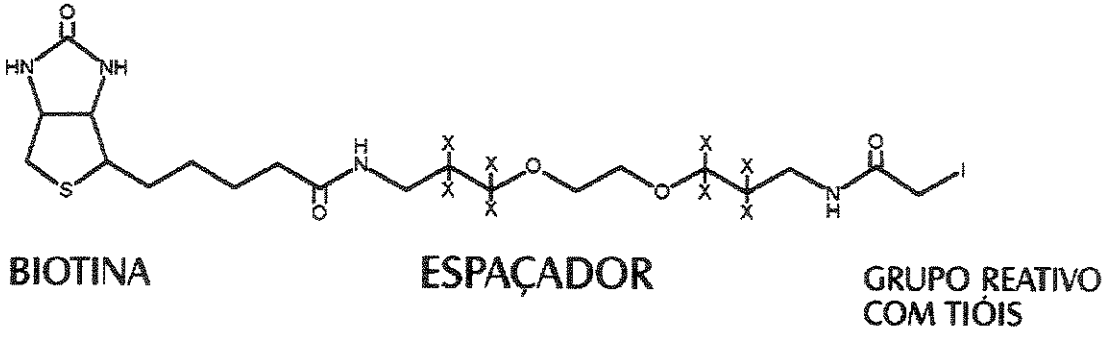


FIGURA 7: Estrutura do reagente ICAT, composto de três elementos: um grupamento reativo à tióis, que se liga à resíduos de cisteínas reduzidas; um espaçador que pode conter deutérios (reagente pesado, com os deutérios nas posições indicadas por X) ou não (reagente leve, com hidrogênios nas posições indicadas por X); e um grupamento de biotina utilizado para o isolamento seletivo dos peptídios contendo cisteína (marcados com ICAT).

[49] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999, 17, 994-999.

Num típico experimento (Figura 8), os resíduos de cisteína reduzida das proteínas nas duas amostras a serem comparadas são rotulados separadamente. Numa amostra é usada a forma isotopicamente "leve" e em outra amostra a forma "pesada" do reagente ICAT. As duas amostras são então misturadas, digeridas com tripsina, e os peptídios contendo cisteína são isolados por cromatografia de afinidade. Através da análise por MS/MS, é determinada a identidade da proteína que deu origem ao peptídio e a abundância relativa de cada proteína nas amostras sendo comparadas. A abundância relativa de um peptídio, e logo, a abundância relativa da proteína de onde o peptídio originou é determinada pela razão das intensidades do sinal espectrométrico das formas isotopicamente leve e pesada de um peptídio marcado com ICAT.

Para análise de misturas complexas, a mistura de peptídios contendo cisteína é fracionada, geralmente por cromatografias líquidas multidimensionais, permitindo a análise de um grande número de peptídios[50,51,52,53,54].

Este método é facilmente automatizável e tende a permitir a análise de classes de proteínas não detectáveis pela 2DE, como as hidrofóbicas, básicas e pouco abundantes. Porém, é importante salientar que as capacidades e limitações da 2DE já são bem conhecidas, enquanto que as tecnologias baseadas em ICAT ainda permanecem, de certa forma, enigmáticas neste sentido uma vez que estão apenas começando a ser aplicadas para estudos proteômicos. Já se sabe que o ICAT não permite a análise de proteínas sem cisteína, e ainda, como a análise é feita em peptídios gerados pela digestão de proteínas, importantes informações geralmente analisáveis em proteínas intactas (como presença de isoformas e modificações pós-traducionais) são perdidas.

Como os objetivos do estudo proteômico global não podem ser alcançados por nenhuma destas plataformas tecnológicas isoladamente, a combinação destes e de outros métodos de análise de proteínas é provavelmente uma necessidade para se analisar proteomas.

---

[50] Griffin, T.J., Gygi, S.P., Rist, B., Aebersold, R. et al. Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer. *Anal. Chem.* 2001, 73, 978-986.

[51] Han, D. K., Eng, J., Zhou, H., Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* 2001, 19, 946-951.

[52] Griffin, T. J., Aebersold, R. Advances in proteome analysis by mass spectrometry. *J. Biol. Chem.* 2001, 276, 45497-45500.

[53] Griffin, T. J., Han, D. K., Gygi, S. P., Rist, B. et al. Toward a high-throughput approach to quantitative proteomic analysis: expression-dependent protein identification by mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2001, 12, 1238-1246.

[54] Griffin, T. J., Goodlett, D. R., Aebersold, R. Advances in proteome analysis by mass spectrometry. *Curr. Opin. Biotechnol.* 2001, 12, 607-612

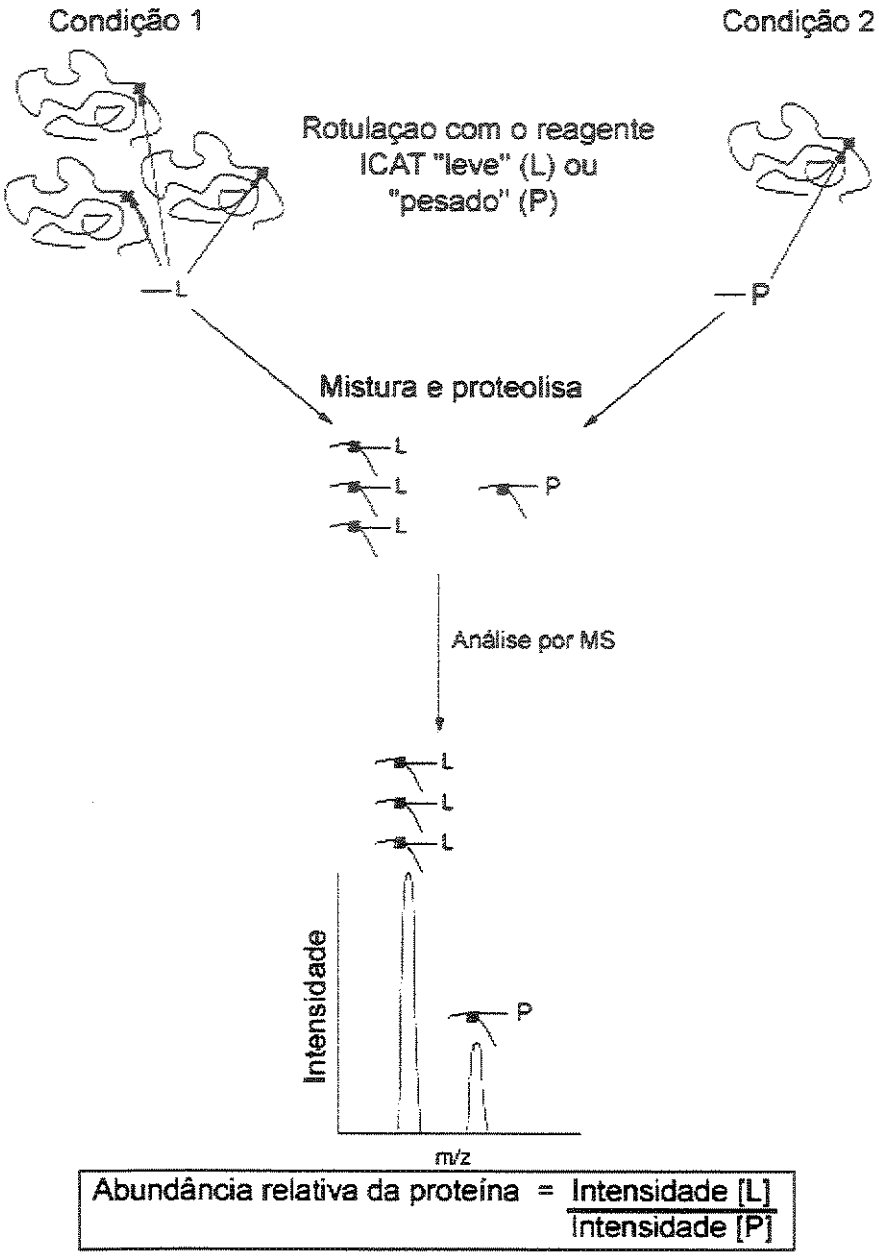


FIGURA 8: Esquema representativo de um experimento de análise quantitativa da abundância relativa de proteína utilizando o reagente ICAT (Modificado de Griffin et al, 2001 [55]).

Objetivos

Parte I: Análise Proteômica da *Xylella fastidiosa*.

Os objetivos foram: (i) caracterização das proteínas expressas pela estirpe 9a5c da *X. fastidiosa* causadora da clorose variegada do citros por eletroforese de duas dimensões (2DE) de alta resolução e espectrometria de massas (MS); (ii) mapeamento das proteínas potencialmente relacionadas aos mecanismos de colonização e patogenicidade; e (iii) construção de um banco de dados para acesso via internet aos dados gerados.

Parte II: Desenvolvimento de Novo Método para Análise Quantitativa de Proteomas por Espectrometria de Massas

O objetivo do presente trabalho foi desenvolver um método que permita a quantificação precisa das diferenças em abundância de cada proteína em diferentes amostras, utilizando para isto a combinação da 2DE, marcação de proteínas com reagentes do tipo ICAT ("Isotope Coded Afinity Tags") e espectrometria de massas.

Justificativa

Parte I: Análise Proteômica da *Xylella fastidiosa*.

Em contraste ao grande volume de informações genômicas disponíveis, os estudos de caracterização bioquímica das proteínas expressas pela *X. fastidiosa* são quase inexistentes. A análise proteômica é uma forma global de identificar e caracterizar proteínas envolvidas em processos biológicos. Informações sobre níveis de expressão, organização estrutural, localização e modificações pós-traducionais permitirão uma maior caracterização dos alvos funcionais envolvidos nos mecanismos de colonização e patogenicidade. O estabelecimento das técnicas de análise proteômica e a construção de um banco de dados deverão contribuir para a exequibilidade de estudos de expressão de proteínas em modelos experimentais com a *X. fastidiosa*.

Parte II: Desenvolvimento de Novo Método para Análise Quantitativa de Proteomas por Espectrometria de Massas

Atualmente, 2DE é o método mais utilizado para quantificação comparativa dos níveis de abundância de uma proteína em diferentes amostras. A quantificação por análise densitométrica implicitamente assume que a intensidade de coloração de um *spot* indica precisamente a quantidade de proteína presente no *spot*. Porém, a implementação prática deste simples conceito tem se mostrado difícil, principalmente devido a limitações de sensibilidade, linearidade quantitativa e reprodutibilidade dos métodos disponíveis de coloração e separação eletroforética. São necessárias novas tecnologias que quantifiquem de forma precisa, reprodutível e robusta as proteínas separadas por 2DE.

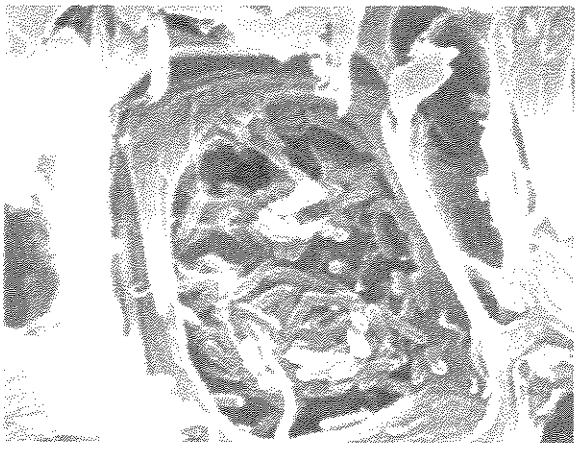
O primeiro capítulo da presente tese traz o artigo aceito para publicação referente à análise de proteínas expressas pela *Xylella fastidiosa*. Resultados não incluídos no artigo são mostrados no item Resultados Suplementares.

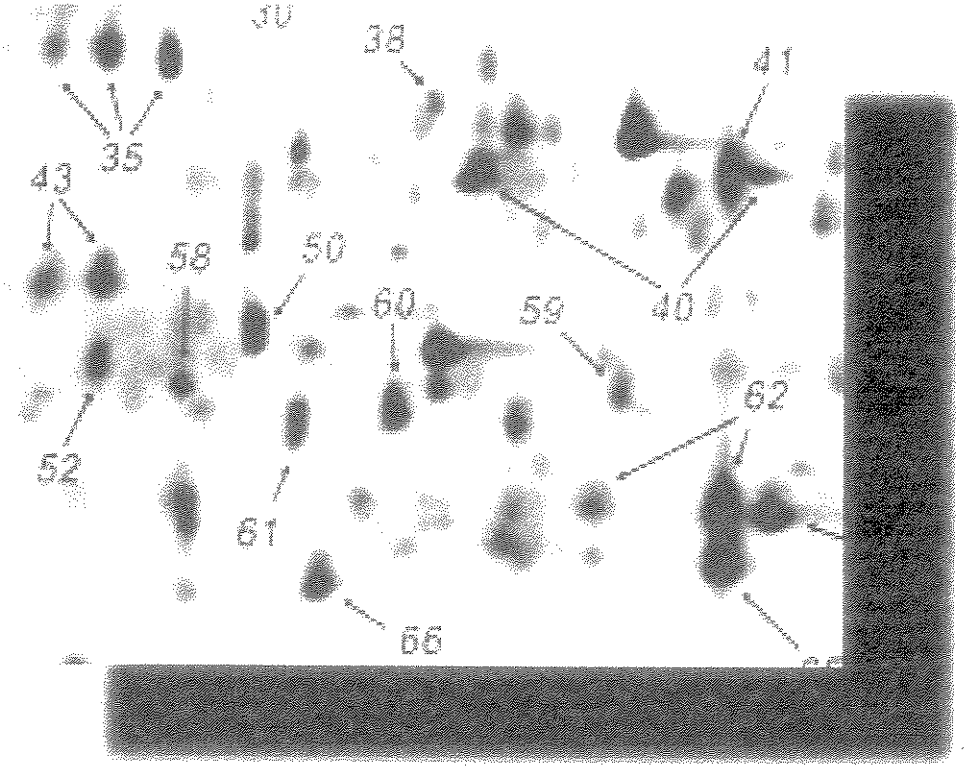
No segundo capítulo, dois artigos publicados descrevem o desenvolvimento de novo método para análise quantitativa de proteomas por espectrometria de massas e a otimização do procedimento de rotulação de proteínas com o reagente ICAT.

Um capítulo de livro a ser publicado foi incluído como apêndice e aborda de forma detalhada os procedimentos para realização da eletroforese de duas dimensões e identificação de proteínas por espectrometria de massas.

Capítulo 1:

Análise Proteômica da Xylella fastidiosa





1. **Artigo aceito:** Análise Proteômica do Fitopatógeno *Xylella fastidiosa* Revela as Principais Proteínas Celulares e Extracelulares e uma Distribuição Peculiar do Codon Bias

Artigo aceito na revista Proteomics

Marcus Bustamante Smolka, Daniel Martins, Flavia Vischi Winck, Carlos Eduardo Santoro, Rafael Ramos Castellari, Fernanda Ferrari, Itaraju Junior Brum, Eduardo Galembeck, Helvécio Della Coletta Filho, Marcos Antonio Machado, Sergio Marangoni, Jose Camillo Novello. **Proteome Analysis of the Plant Pathogen *Xylella fastidiosa* Reveals Major Cellular and Extracellular Proteins and a Peculiar Codon Bias Distribution.** *Proteomics (Accepted for publication)*

Proteome Analysis of the Plant Pathogen *Xylella fastidiosa* Reveals Major Cellular and Extracellular Proteins and a Peculiar Codon Bias Distribution

Marcus Bustamante Smolka[1], Daniel Martins[1], Flavia Vischi Winck[1], Carlos Eduardo Santoro[1], Rafael Ramos Castellari[1], Fernanda Ferrari[1], Itaraju Junior Brum[1], Eduardo Galembeck[1], Helvécio Della Coletta Filho[2], Marcos Antonio Machado[2], Sergio Marangoni[1], Jose Camillo Novello[1]

[1] Departamento de Bioquimica, UNICAMP, Campinas, SP, Brazil

[2] Centro de Citricultura Sylvio Moreira / IAC, Cordeiropolis, SP, Brazil

Running title: Proteome analysis of *Xylella fastidiosa*

Corresponding author:

Prof. Dr. Jose Camillo Novello

Laboratório de Química de Proteinas, LAQUIP

Departamento de Bioquimica, Instituto de Biologia, UNICAMP

Campinas, SP, Brazil - 13083-970

PHONE / FAX: (55) 19-37886133

email: jcn@unicamp.br

Abbreviations: BCYE: buffered charcoal yeast extract; BPB: bromophenol blue; CB: codon bias.

Keywords: *Xylella fastidiosa*, Citrus Variegated Chlorosis, Extracellular proteins, Adhesins, Codon bias

Summary

The bacteria *Xylella fastidiosa* is the causative agent of a number of economically important crop diseases, including Citrus Variegated Chlorosis. Although its complete genome is already sequenced, *X. fastidiosa* is very poorly characterized by biochemical approaches at the protein level. In an initial effort to characterize protein expression in *X. fastidiosa* we used one and two-dimensional gel electrophoresis and mass spectrometry to identify the products of 142 genes present in a whole cell extract and in an extracellular fraction of the citrus isolated strain 9a5c. Of particular interest for the study of pathogenesis are adhesion and secreted proteins. Homologs to proteins from three different adhesion systems (type IV fimbriae, mrk pili and hsf surface fibrils) were found to be co-expressed, the last two being detected only as multimeric complexes in the high molecular weight region of one dimensional electrophoresis gels. Using a procedure to extract secreted proteins as well as proteins weakly attached to the cell surface we identified 30 different proteins including toxins, adhesion related proteins, antioxidant enzymes, different types of proteases and 16 hypothetical proteins. These data suggest that the intercellular space of *X. fastidiosa* colonies is a multifunctional microenvironment containing proteins related to *in vivo* bacterial survival and pathogenesis. A codon usage analysis of the most expressed proteins from the whole cell extract revealed a low biased distribution, which we propose is related to the slow growing nature of *X. fastidiosa*. A database of the *X. fastidiosa* proteome was developed and can be accessed via the internet (URL: www.proteome.ibi.unicamp.br).

Introduction

Xylella fastidiosa is a slow growing Gram-negative bacterium involved in many economically important plant diseases, such as citrus variegated chlorosis (CVC) in sweet orange [1,2], Pierce's disease (PD) in grapevine [3], leaf scorch in almond and other species such as coffee, plum, oleander and mulberry [4]. In all cases, *X. fastidiosa* infects the plant xylem and compromises fruit production. The bacteria multiply in the foregut of sharpshooter leafhoppers, which feed on the sap of the plant xylem. The insect vector delivers the bacteria directly into the xylem system of host plants, where they multiply and cause symptoms of chlorosis. Fruits from infected plants are small, hardened and cannot be used for commercial and industrial purposes.

CVC is a major threat to the Brazilian citrus industry, being present in all the main citrus growing areas where it is responsible for significant losses of orange production [5]. The complete genome sequence of the CVC strain 9a5c of *X. fastidiosa* was published in 2000 [6] and represents the first complete genome sequence of a plant pathogen. The determination of the genome sequence is an important step towards a better understanding of *X. fastidiosa* pathogenicity and improved disease control. A total of 2849 genes were annotated in the circular bacterial chromosome and the two plasmids.

Although the exact mechanisms of pathogenicity are unclear, the adhesion and secreted proteins are expected to be key elements. An important issue for *X. fastidiosa* is the necessity to adhere to both plant and insect hosts. To colonize the insect's foregut, *X. fastidiosa* needs to adhere to the insect tissue so that it can resist the high flow of xylem sap passing through. In the plant, adhesion to the xylem walls should enable appropriate conditions for bacterial growth and biofilm formation. Important aspect to bacterial pathogenesis is cell aggregation (bacterium-bacterium interaction),

which has been proposed to lead to vascular occlusion of the xylem, causing water and nutrient stress in the plant [4,7]. Moreover, secreted proteins such as proteases, cellulases and lipases may be involved in the infection process by disrupting the plant tissue and allowing spreading of bacteria through out the vascular system [7]. Also, protein toxins can be secreted and may be involved in observed disease symptoms.

In contrast to the huge amount of available genomic information, *X. fastidiosa* is very poorly characterized by biochemical approaches at the protein level. The present paper represents the first step to study protein expression in *X. fastidiosa.* We used one and two-dimensional gel electrophoresis and mass spectrometry to identify abundant proteins from a citrus isolated strain 9a5c of *X. fastidiosa* grown in vitro on solid culture medium. Particular attention has been paid to the adhesion and secreted proteins due to their possible role in bacterial pathogenesis and their usefulness as molecular targets for understanding and controlling the disease. Additionally, an unusual codon bias distribution of the most abundant proteins was observed and its probable significance is discussed.

Materials and methods

Bacterial strain and culture condition

The pathogenic strain, 9a5c, of *X. fastidiosa*, originally isolated from Brazilian sweet orange,was used in the current study. The complete genome sequence of this isolate is available at the *Xylella* Genome Project Web site (URL: http://aeg.lbi.ic.unicamp.br/xf/). Bacteria were cultured on solid buffered charcoal yeast extract (BCYE) medium [8] at 28ºC for 21 days. *X. fastidiosa* colonies were collected and immediately stored under liquid nitrogen for later processing.

Sample Preparation

For the preparation of whole cell protein extracts, approximately 100 mg (wet weight) of bacteria were washed three times in 1 mL of washing buffer containing 10 mM Tris (pH 8.8), 3 mM KCl, 50 mM NaCl, 5 mM EDTA and 1mM PMSF and centrifuged for 2 minutes at 1600g. The pelleted cells were then lysed with 200 µL of the following solution: 10 mM Tris (pH 8.8), 0.5 % (w/v) SDS, 5 mM EDTA, 1mM PMSF. After adding 100mM DTT, the sample was boiled for 3 minutes and stored at -70ºC.

An extracellular protein fraction (containing secreted proteins as well as proteins weakly attached to the cell surface) was obtained by washing the cells (100 mg) twice with 800 µL of a washing buffer containing 10 mM Tris (pH 8.8), 3 mM KCl, 50 mM NaCl, 5 mM EDTA and 1mM PMSF. After pelleting the cells, the proteins present in the supernatant were precipitated with a cold ethanol: acetone: acetic acid (50 :50 :0.1 v/v/v) solution in ice for 30 minutes. Proteins were then solubilized in 50 µL of 10 mM Tris (pH 8.8), 0.5 % (w/v) SDS, 100mM DTT, 5 mM EDTA and 1mM PMSF, boiled for 3 minutes and stored at -70ºC.

An aliquot of each of the resulting protein extracts was used to determine the protein concentration using a commercial protein assay kit (Bio-Rad).

SDS-PAGE

For the analysis of high molecular weight proteins, 8 µL (approximately 120 µg of protein) of whole cell protein extract or 6 µL (approximately 40 µg of protein) of extracellular protein extract were added to 8 µL of sample buffer 1D [6 % (w/v) SDS, 100 mM Tris (pH 6.8), 30 % glycerol, 100 mM DTT and 0.001 % (w/v) bromophenol blue (BPB)], boiled for 5 minutes and separated on 9 %T polyacrylamide gels (14 X 16 X 0.15 cm) containing 10 % glycerol. Proteins were visualized using Coomassie blue staining.

Two-dimensional gel electrophoresis

Prior to 2-DE, 5 µL of whole cell protein extract (approximately 75 µg of protein) or 9 µL of extracellular protein extract (approximately 60 µg of protein) were diluted to a final volume of 350 µl with a solution containing 8 M urea, 4 % (w/v) CHAPS, 2 % (v/v) carrier ampholytes pH 3-10, 70 mM DTT and 0.001 % (w/v) bromophenol blue (BPB). Except for CHAPS which was obtained from Sigma, all reagents were obtained from Amersham Biosciences (Uppsala, Sweden). After centrifugation at 8000 g for 2 min, the supernatant was used for first dimension run. Samples were applied to IPG gel strips with a non-linear separation range of pH 3-10 (Amersham Biosciences). After a 10 hour rehydration, isoelectric focusing was carried out, at 20°C, for 1 hour at 500 V, for 1 an additional hour at 1000 V and then for 10 hours at 8000 V in an IPGphor apparatus (Amersham Biosciences) maintaining a limiting current of 50 µA per strip. First dimension strips were subjected to the standard reduction and alkylation steps prior to

second dimension electrophoresis. Strips were soaked for 10 minutes in a buffer containing 50 mM Tris-HCl (pH 6.8), 6 M urea, 30 % (v/v) glycerol, 2 % (w/v) SDS and 2 % (w/v) DTT, and for an additional 10 minutes in the same buffer containing 2.5 % (w/v) iodoacetamide in place of the DTT. Second dimension electrophoresis (SDS-PAGE) was performed on 12.5 %T polyacrylamide gels run on a SE-600 system connected to a Multitemp II refrigerating system (Amersham Biosciences). The IPG gel strips were sealed to the surface of the second dimension gel using 0.5 % (w/v) agarose. Electrophoresis was carried out for 1 hour at 90V at which time a constant current of 30 mA per gel was applied until the dye front reached the lower end of the gel. Proteins were detected by a silver nitrate staining protocol adapted from [9].

Image analysis

ImageMaster 2D software, version 3.01, (Amersham Biosciences) was used for spot detection and pI/M_r calibration. The observed isolectric points and molecular weights of the bacterial proteins were determined by comparions with known, 2D and 1D marker proteins (Bio-rad) run independently and comigrated with *X. fastidiosa* samples. The M_r and pI values of the markers were plotted as "y" and "x" coordinates respectively of the 2-DE gels and equations were fitted by First-order LaGrange regression. Using these equations, the "x" and "y" coordinates of each spot were converted to observed pI and M_r values, respectively.

Peptide mass fingerprinting

Peptides were generated and extracted from the gel-separated proteins following established *in gel* trypsin digestion protocols [10] using sequencing grade modified porcine trypsin from Promega (Madison, USA). The peptides were analyzed on a Voyager DE-PRO MALDI-TOF mass spectrometer (Applied Biosystems). Prior to their

application to the sample plate the samples were purified using C18 ZipTips (Millipore), and eluted directly with a matrix solution of 2 % (w/v) α-ciano-4-hydroxycinnamic acid (Sigma), 60% acetonitrile and 0.1 % (v/v) TFA. Internal mass calibration was performed using trypsin autodigestion products (842.509, 1045.563 and 2211.104 Da). The masses of monoisotopic peaks with relative intensity over 2 % of the most intense peak in the spectrum were used for database searching.

For the identification of the bacterial proteins, the measured masses of the tryptic peptides were searched against the *X. fastidiosa* genome database using the MS-Fit program [11]. All searches were performed using a protein mass window between 1 and 100 kDa, unless the protein spot or band was present on a region over 80 kDa where no mass restriction was applied. The search parameters allowed one missed cleavage, oxidation of methionine and carbamidomethylation of cysteine. For a positive protein identification, at least four peptides had to match with a mass accuracy better than 40 ppm. Relative peak intensity of the matched peaks, linearity of mass error over the spectrum and comparison of predicted and observed M_r/pI values were checked for confirmation of the identification. Excessions were allowed for obvious positive identifications that showed differences in observed and theoretical M_r/pI values, what occured mainly due to the presence of multimers or fragments and N-terminal processed proteins.

For identification of multiple proteins in a SDS-PAGE band, peak masses were submitted for database searching and after the first positive identification, the corresponding matched masses were subtracted from the list and a new search was performed. This process was repeated until no more positive identification was possible.

Calculation of codon bias

The codon bias value of the gene sequences was based on the Codon Bias Index, as defined by Bennetzen and Hall [12]. It was calculated using CodonW program, accessible at URL: www.molbiol.ox.ac.uk/cu. Based on all the gene sequences present in the *X. fastidiosa* genome, the program computes the relative frequency of each codon and attributes a value for each gene that varies between -1 and +1. The more of the most used codons are present in a given gene sequence, the greater is the codon bias value.

Results and Discussion

Whole cell extract proteins

The bacteria grown for 21 days in solid BCYE media were used for the construction of the reference 2-DE map and identification of the most abundant expressed proteins. Washed cells were lysed and the proteins were solubilized as described in Material and Methods. 2-DE separation was performed after diluting the sample in the conventional sample buffer containing CHAPS and urea (described under materials and methods). Figure 1A shows the 2-DE profile of proteins over the pI range 4 to 8.5. For the detection and identification of high molecular weight proteins, the same sample was also analysed by SDS-PAGE using a 9 %T acrylamide gel containing 10 % glycerol (Figure 1B). Seventeen high molecular weight proteins were identified by one dimensional electrophoresis that were not detected by 2-DE, including two proteins (XF1981 and XF0395) present as multimers (bands 99 and 100 respectively in Figure 1B). Table I lists the 111 proteins resolved by 2-DE and 1-D gels that were identified by mass spectrometry.

Major fimbriae and adhesion related proteins

Proteins related to the type IV fimbriae were present in high abundance. This system is thought to play an important role in bacterial adhesion, protease secretion and twitching motility [13, 14]. The detected proteins related to type IV fimbriae were: the fimbriae subunit protein (XF2542, spot 90, Figure 1A), two PilY1 proteins (XF0032, spot1, Figure 1A and XF1224, band 108, Figure 1B) involved in fimbrial biogenesis and a PilT protein (XF1633, spot 34, Figure 1A). PilT protein is supposedly responsible for retracting and extending the fimbriae, in a mechanism known as twitching motility. It is possible that this motility mechanism is important in host colonization, allowing the bacteria to move

systemically within the plant vascular system. In the case of Pierce disease, movement of bacteria through 17 internodes of grapevine plant was observed, whereas it was also seen that weakly virulent strains moved more slowly and avirulent strains failed to move at all [7].

In the region over 300 kDa of 9 %T acrylamide SDS-PAGE gels a protein (XF1981, band 99, Figure 1B) was detected with a high similarity to Hsf adhesin of *Haemophilus influenza*. In this organism, this protein was shown to form surface fibrils and to represent an additional adhesion mechanism to the type IV fimbriae [15, 16]. The theoretical molecular weight of the XF1981 gene product is 118,447 kDa, so it is most likely to be forming a multimer complex resistant to SDS and DTT denaturation. Moreover, the fact that this protein complex was not extracted in the extracellular fraction (see results below), suggests that it is strongly bound to the cellular membrane. The results show that *X. fastidiosa* may have the ability to co-express different adhesion systems.

Additionally, prediction programs that recognize specific transmembrane hydrophobic segments in protein sequences indicated some hypothetical proteins to be membrane located and they may also be involved in adhesion. For example, the protein XF1036 (spot 2, Figure 1A) was predicted to be membrane located and, in fact, showed a partial similarity in the N-terminal region to a *Myxococcus xanthus* OmpA-related protein precursor reported to be involved in cellular adhesiveness during development [17].

An OprF porin is the major outer membrane protein

In addition to housekeeping, ribosomal and metabolic proteins, an abundant outer membrane protein (spot 28, XF0343) was also resolved by 2-DE (Figure 1A). In a 1-D and a 2-DE profile of a membrane-enriched extract, this protein represented the most

abundant protein (data not shown), indicating that it is probably the major outer membrane protein. XF0343 has a high similarity to OprF porin proteins that are related to the OmpA family and in *Pseudomonas* species it is reported to be involved in the ability of bacteria to grow and survive in low-osmolarity niches [18]. Based on the fact that the plant xylem is an extremely hyposmotic environment, the high abundance of XF0343 could be critical for the ability of *X. fastidiosa* to colonize and survive in the plant.

Iron uptake/storage

Proteins involved in iron uptake and storage were also present in high abundance, including a number of hypothetical outer membrane proteins with low similarities to ferric enterobactin receptors. The detected proteins were: the bacterioferritin [XF0395, spot 84 in Figure 1A and as a multimer in band 100, Figure 1B (in most bacteria, homologous bacterioferritins are known to form an oligomer of 24 similar subunits)], a two-component system regulatory protein (XF0389, spot 60, Figure 1A) reported to be important for the high affinity iron acquisition in *Rhizobium leguminosarum* and two hypothetical membrane proteins (XF0339, spot 6 and XF2237, spot 4, Figure 1A) with homology to ferric enterobactin receptors. The aconitase protein (XF0290, spot 8) was also resolved by 2-DE (Figure 1A). Aconitase is reported to be involved in iron homeostasis and in the regulation of several extracellular pathogenicity factors [19].

Extracellular proteins

In order to identify the major *X. fastidiosa* proteins present in the intercellular space of the colonies growing in solid media, cells were washed with a buffer containing salt (50 mM NaCl) and protease inhibitors. This washing step is intended to extract secreted proteins as well as proteins weakly attached to the cell surface. After pelleting the cells,

the supernatant containing the extracellular proteins was collected, the proteins were precipitated and analyzed by 2-DE (Figure 2A). High molecular weight proteins were analyzed by a low percentage acrylamide SDS-PAGE gel (Figure 2B). As the sample was not too complex, several proteins in the molecular weight region below 40 kDa were also identified from 12.5 %T polyacrylamide gels using peptide mass fingerprint (Figure 2C).

In the 2-DE protein profile shown in Figure 2A, it was possible to observe the presence of three of the most abundant cellular proteins XF2340 (spot C1), XF0615 (spot C2) and XF0343 (spot C3). Because of their high abundance in the cell, it is hard to determine if they are contaminating proteins from residual cell lysis or, in fact, a fraction of these proteins that is surface exposed or secreted. Several other high abundant ribosomal proteins (for example XF2634, XF2438 and XF2559) and elongation factors (for example, XF2628, XF2579 and XF2473) were not detected. Seventeen proteins were found to be enriched or only present in the extracellular fraction (Table 2). Other eighteen high molecular weight proteins were only identified following one dimensional gel electrophoresis (Table 2, Figure 2B). Three of the 18 proteins identified were low molecular weight proteins present as multimers (XF0395, XF0077 and XF0078). Twelve proteins with molecular weight below 40 kDa were only detected by one dimensional gel electrophoresis on 12.5 %T polyacrylamide gels (Figure 2C). The majority had predicted pIs >9.0 and none of them were identified by 2-DE using IPG strips 3-10. In this case, very alkaline immobilized pH gradients (up to pH 12) may allow the detection of several of these basic proteins using 2-DE [20, 21], although this analysis was not performed here. Table 2 lists all the proteins identified in the extracellular fraction and Figure 3 shows their functional class distribution. Proteins related to adhesion, proteolysis, antioxidant defense, toxicity and iron uptake/storage represented the major functional classes among the identified proteins in the

extracellular fraction. Several proteins putatively localized in the outer membrane were detected in the extracellular fraction, and the fact that they were extracted with a low salt containing buffer indicate that they were either weakly bound to the cell membrane or, in fact, secreted.

Secreted proteases and toxins

The extracellular fraction was rich in proteases and proteins involved in toxicity, whose homologous proteins in other bacteria are known to be secreted. The secreted proteases detected were a zinc (XF0816, spot 1 in Figure 2A and band 23 in Figure 2B) and two serine proteases (XF1026, spot 2 and XF1851, spot 3, both in Figure 2A; band 26, Figure 2B). Secretion of these proteases may be important for nutritional purposes and also for the degradation of plant pit membranes, allowing the bacteria to move to uncolonized xylem vessels.

Two secreted haemolysin-like proteins (XF2407, band 20 and XF0668/XF1011, band 25, Figure 2B) that belong to the RTX toxin family were detected, and are believed to represent important virulence factors [22, 23]. A NonF related protein (XF1137, spot 11, Figure 2A) involved in the production of polyketides was abundantly present both in the whole cell extract and in the extracellular fraction. Polyketide toxins are frequently virulence factors in bacterial pathogens [24].

Adhesion related proteins in the extracellular fraction

The proteins XF0077 and XF0078 were both detected in the upper region (> 300 kDa) of the 9 %T acrylamide SDS-PAGE gel. These proteins had a high similarity to each other as well as to the mrkD gene, whose product is the adhesin of a thin and flexible pili from *Klebsiella pneumoniae* [25, 26]. The predicted molecular weights of XF0077 and XF0078 are 36,854 Da and 37,601 Da, respectively, so like the Hsf adhesin (XF1981)

detected in the whole cell extract, they probably formed high molecular weight protein multimers that were resistant to strong denaturation by SDS and DTT. The fact that this adhesin was only detected in the extracellular fraction suggests that it was weakly bound to the cells and could be extracted by washing with the salt solution. Interestingly, the *mrkA* gene, that encodes the pili subunit, as well as the other *mrk* genes necessary for the assembly of the fimbrial filament were not annotated in the genome sequence of *X. fastidiosa* [25].

The type IV fimbriae subunit protein (XF2542) was present in both the whole cell and extracellular fraction, with a characteristic unfocused pattern (spot 90, Figure 1A and spot 17, Figure 2A). Its presence in both fractions is in agreement with the model that this fimbriae structure begins in the intracellular region, crosses the inner membrane, periplasm, and outer membrane and expands to the extracellular region [13]. The proteins fimD (XF0081) and PilY1 (XF0032) were also detected in the extracellular fraction and consist of outer membrane proteins involved in fimbrial biogenesis.

Extracellular antioxidant enzymes

Four antioxidant enzymes were detected in the extracellular fraction: superoxide dismutase [MN] (XF2614, spot 13 in Figure 2A), alkyl hydroperoxide reductase (XF1530, spot 12 in Figure 2A), catalase (XF2232, band 27 in Figure 2B) and an organic hydroperoxide resistance protein (XF1827, band 35 in Figure 2C). The first two proteins were also detected in the whole cell extract, while the two last proteins were only detected on the extracellular fraction using SDS-PAGE and may have not been detected in the whole cell extract because the sample was too complex for identifying them in SDS-PAGE gels at their specific molecular weight regions. Considering that these enzymes are active outside the cell, they may contribute to an external antioxidant

protection to free radicals produced by plants, which is known to be an important defense mechanism against microorganisms. This mechanism of extracellular antioxidant protection may be especially important when *X. fastidiosa* establishes biofilm formation in the xylem and antioxidant enzymes can be concentrated in the intercellular space.

Other proteins in extracellular fraction

Several proteins with diverse functions such as lipase (XF1253), malate dehydrogenase (XF1211), polynucleotide phosphorylase (XF0239) and carbonic anhydrase (XF2095) were also detected and probably have a functional role in the extracellular space. Secretion of carbonic anhydrase may be related to concentration of carbon dioxide for carbon assimilation.

Extracellular hypothetical proteins

Approximately one third of the proteins from the extracellular fraction did not show significant homologies and were classified as hypothetical proteins. Using prediction programs for protein localization sites, of the secreted hypothetical proteins, 49 % had a high score for outer membrane location (data not shown), which is a high proportion considering that only 13 % of the non-hypothetical proteins from the extracellular fraction had such a high score.

A surprising finding was that a plasmid protein (XFa0032) is one of the most abundant protein both in the whole cell extract as well as in the extracellular protein fraction. This protein shows a low similarity (e-10^7) to a cyanohydrin lyase from *Arabdopsis thaliana* and may be involved in a biochemical pathway for detoxification to cyanide, a substance known to be secreted by plants as a defense mechanism against microorganism [27].

***Xylella fastidiosa* proteins have a peculiar codon bias distribution**

The higher the codon bias (CB) value for a gene, the more of the organism's preferred codon for a specific amino acid is present in its sequence [28]. In several unicellular species it has been observed that codon usage bias is related to expression patterns [29, 30]. The extent of codon usage bias was found to correlate with levels of gene expression, where highly expressed proteins have CB values >0.2 [29-31]. Figure 4A shows the codon bias distribution for all genome orfs and for the most abundant proteins in *E. coli*. In this example, it is clear that codon usage in highly expressed genes is biased towards the most used codons. This reflects a coadaptation between codon usage and tRNAs abundance to optimize the efficiency of proteins synthesis [32-34]. It has been shown experimentally that mRNA consisting of preferred codons is translated faster than mRNA artificially modified to contain rare codons [35, 36].

When the CB value was calculated for the 50 most expressed genes present in the 2-DE map of *X. fastidiosa* whole cell extract the distribution differed from that observed for *E. coli* (Figure 4B). In fact, the CB distribution of these highly expressed genes followed the same distribution pattern for all the genes in the genome. It was intriguing that several highly expressed chaperones and ribossomal proteins showed very low CB values. Moreover, the distribution for the whole *X. fastidiosa* genome differed from that observed for most organisms in showing an unusually wide distribution extending to the lower value region.

Codon usage bias may be explained partly by translation efficiency-related selective pressures and it was previously demonstrated that natural selection shapes codon usage in *E. coli* and *S. cerevisiae* [37, 38]. It was possible that the results observed for *X. fastidiosa* were a consequence of a lack of selective pressure and, perhaps, *X. fastidiosa*'s recent phylogenetic origin. Furthermore, it may indicate

abundant recent events of lateral gene transfer. At the time of transfer, the introduced genes have codon usage that is typical for the donor genome. With time, the codon usage evolves to match the codon usage of the host genome, in a process called "amelioration" [39, 40]. According to this model, *X. fastidiosa* would represent an "unameliorated" genome, where most of the proteins have low CB values.

Another interesting aspect is that *X. fastidiosa* is a fastidious organism with special nutritional requirements and slow growth with doubling times over 10 hours. The lack of an optimized system for the use of tRNA probably results in a slower velocity in protein production, which could well be the cause for the slow bacterial growth. Figure 5 shows the CB distribution of all genome orfs from 24 different bacteria with whole sequenced genomes. *X. fastidiosa* is the only bacteria that have the majority of the genes in the region of CB value below zero. Its profile was in the extremely "unameliorated" side, together with *Borrelia burgdorferi*. The fact that *B. burgdorferi* is also a slow growing bacteria, with a generation time of about 9 hours reinforces the idea that "unamelioration" is related to slow growth rate. In fact, a tight correlation has already been observed between growth rates, translational elongation rates and the efficiency with which aminoacyl-tRNA species are matched with the cognate codon on the ribosome of different *E. coli* isolates [28, 41].

But is this slow growing nature a simple consequence of the low CB in higly expressed genes? It is also possible to imagine that a slow growing nature is beneficial for the pathogen to grow in a nutrient limited environment. In this case, natural selection would be pressuring for a low codon bias value in highly expressed proteins, representing a mechanism to regulate the rate of bacterial growth.

Concluding remarks

Since *X. fastidiosa* was first associated to CVC, in 1990, there has been significant progress in the study of this plant pathogen, mainly concerning its morphological characterization, epidemiology and diagnostic procedures. But the information related to its molecular characterization was very limited until the complete sequencing of the genome in 2000 [6]. The huge amount of genetic information made available contrasted with a total lack of knowledge on the protein expression of this organism. In the present work, the products of 142 different genes from a whole cell extract and an extracellular fraction of *X. fastidiosa* grown on BCYE culture media were identified for the first time. The use of SDS-PAGE was useful for the detection of proteins not resolved by 2-DE, specifically high molecular weight and basic proteins. The identification of the adhesion proteins XF1981, XF0077 and XF0078, was only possible since SDS-PAGE on 9 %T acrylamide gels could resolve the extremely high molecular weight multimeric complexes formed by these proteins. This approach clearly illustrated the necessity of using combined techniques for protein separation.

Proteins putatively associated with three different adhesion systems were found to be constitutively co-expressed. This suggests the presence of structures posessing different adhesive properties that may be important for *X. fastidiosa* to colonize hosts (insects and plants) with different tissue structural organizations. Moreover, co-expressing different adhesion structures might be important for the bacteria to grow adhered to the xylem walls but still to be able to adhere to the foregut of insects that feed on the plant. The fact that the adhesins were detected as multimeric complexes may be related to their function in forming large adhesive structures. Further analyses should be performed to obtain a detailed structural and functional characterization of these adhesins.

Concerning the study of extracellular proteins, homology studies and available computational programs are usually insufficient for identifying which proteins are in fact secreted or exported to the outer membrane or periplasm, so practical experimentation is required. The use of solid medium for growing the bacteria was found to be useful for extracting extracellular proteins, mainly because they are concentrated in the intercellular space of bacterial colonies, allowing considerable amounts to be easily recovered. Analysis of the extracellular fraction revealed the presence of 30 proteins with varied functions, suggesting that the intercellular space of *X. fastidiosa* colonies is a multifunctional microenvironment containing proteins related to *in vivo* bacterial survival and pathogenesis. If actively secreted or not, they are in fact very likely to be extracellularly located and a number of them may be important for the bacteria to colonize and successfully parasitize the hosts. Further functional analysis on the 16 hypothetical proteins identified in the extracellular fraction should be performed.

In the present work, we also detected unusually low codon bias values among highly expressed genes and we proposed that it was responsible for the slow growing nature of *X. fastidiosa*. Knowing that *X. fastidiosa* lacks an optimized mechanism of preferential codon usage for protein synthesis may contribute to understanding several biological aspects of this plant pathogen. This metabolic "pitfall" in the process of protein synthesis may be ultimately used advantageously in a strategy for disease control.

References

[1] Rossetti, V., Garnier, M., Bové, J. M., Beretta, M. J., Teixeira, A. R. R., Quaggio, J. A., DeNegri, J. D., *C. R. Acad. Sci.* 1990, 310, 345-349.

[2] Chang, C. J., Garnier, M., Zreiki, L., Rossetti, V., Bové, J. M., *Curr. Microbiol.* 1993, 27, 137-142.

[3] Davis, M. J., Purcell, A. H., Thomson, S. V., *Science* 1978, 199, 75-77.

[4] Purcell, A. H., Hopkins, D. L., *Annu. Rev. Phytopathol.* 1996, 34, 131-151.

[5] Amaro, A. A., Maia, M. L., Gonzales, M. A., in: Donadio, L. C., Moreira, C. S. (Eds), *Citrus Variegated Chlorosis*, Est. Exp. Citricultura, Bebedouro, SP, Brazil 1998, pp. 123-139.

[6] Simpson, A. J. G., Reinach, F. C., Arruda, P., Abreu, F. A., *et al. Nature* 2000, 406, 151-157.

[7] Hopkins, D. L., *Ann. Rev. Phytopathol.* 1989, 27, 271-290.

[8] Wells, J. M., Raju, B. C., Nyland, G., Lowe S. K., *Appl. Environ. Microbiol.* 1981, 42, 357-363.

[9] Blum, H., Beier, H., Gross, H.J. *Electrophoresis* 1987, 8, 93-99.

[10] Schevechenko, A., Wilm, M., Vorm, O., Mann, M. *Anal. Chem.* 1996, 68, 850-858.

[11] Clauser K. R., Baker P. R., Burlingame A. L. *Anal. Chem.* 1999, 71, 2871-2882.

[12] Bennetzen J. L., Hall B. D. *J. Biol. Chem.* 1982, 257, 3026-3031.

[13] Strom M. S., Stephen L. *Annu. Rev. Microbiol.* 1993, 47, 565-596.

[14] Soto G. E., Hultgreen S. J. *J. Bacteriol.* 1999, 181, 1059-1071.

[15] Geme J. W., Cutter D. *Mol. Microbiol.* 1995, 15, 77-85.

[16] Geme J. W., Cutter D., Barenkamp S. *J. Bacteriol.* 1996, 178, 6281-6287.

[17] Martinez-Canamero M., Munoz-Dorado J., Farez-Vidal E., Inouye M. *et al. J. Bacteriol.* 1993, 175, 4756-4763.

[18] Rawling E. G., Brinkman F. S., Hancock R. E. *J. Bacteriol.* 1998, 180: 3556-3562.

[19] Wilson T. J., Bertrand N., Tang J. L., Feng J. X. *et al. Mol Microbiol.* 1998, 28: 961-967.

[20] Gorg A., Obermaier C., Boguth G., Csordas A. *et al. Electrophoresis* 1997, 18: 328-337.

[21] Fountoulakis M., Takacs B., Langen H. *Electrophoresis* 1998, 19: 761-766.

[22] Lally E. T., Hill R. B., Kieba I. R., Korostoff J. *Trends Microbiol.* 1999, 7, 356-361.

[23] Oresnik I. J., Twelker S., Hynes M. F. *Appl. Environ. Microbiol.* 1999, 65, 2833-2840.

[24] Bender C. L., Rangaswamy V., Loper J. *Annu. Rev. Phytopathol.* 1999, 37, 175-196.

[25] Gerlach G., Clegg S., Allen B. L. *J. Bacteriol.* 1989, 171, 1262-1270.

[26] Allen B. L., Gerlach G., Clegg S. *J. Bacteriol.* 1991, 173, 916-920.

[27] Nahrstedt A. *Pl. Syst. Evol.* 1985, 150, 35-47.

[28] Kurland C. G. *FEBS Lett.* 1991, 285, 165-169.

[29] Gouy M., Gautier C. *Nucleic. Acids Res.* 1982, 10, 7055-7074.

[30] Sharp P. M., Tuohy M. F., Mosurski K. R. *Nucleic. Acids Res.* 1986, 14, 5125-5143.

[31] Gygi S. P., Corthals G. L., Zhang Y., Rochon Y. *et al. Proc Natl Acad Sci U S A.* 2000, 15, 9390-9395.

[32] Moriyama E. N. , Powell J. R. *J. Mol. Evol.* 1997, 45, 514-523.

[33] Duret L. *Trends Genet.* 2000, 16, 287-289.

[34] Kanaya S., Yamada Y., Kudo Y., Ikemura T. *Gene* 1999, 238, 143-155.

[35] Robinson M., Lilley R., Little S., Emtage J. S. *et al. Nucleic. Acids Res.* 1984, 12, 6663-6671.

[36] Sorensen M. A., Kurland C. G., Pedersen S. *J. Mol. Biol.* 1989, 207, 365-377.

[37] Grantham R., Gautier R., Gouy M., Jacobzone M. *et al. Nucleic. Acids Res.* 1981 9, 43-79.

[38] Ikemura T. *Mol. Biol. Evol.* 1985, 2, 13-34.

[39] Ermolaeva M. D. *Curr. Issues Mol. Biol.* 2001, 3, 91-97.

[40] Lawrence J. F., Ochman H. *J. Mol. Evol.* 1997, 44, 383-397.

[41] Mikkola R., Kurland C. G. *Mol. Biol. Evol.* 1992, 9, 394-402.

Acknowledgments

This work was supported by grants from Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), as part of the *Xylella fastidiosa* Functional Genome Program, and from Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq). We thank Prof. Phillip Cash for the valuable suggestions and careful correction of the manuscript.

Tables

Table 1. List of the identified proteins from whole cell protein extract. Corresponding spots or bands are indicated in Figures 1A and 1B.

SPOT	PRODUCT	GENOME ACCESSION CODE	pI obs	pI prev	MW obs	MW prev	N° of peptides matched	Sequence Coverage (%)
		2-DE						
1	PilY1 gene product	XF0032	6.56	7.25	124689	133494	34	34
2	hypothetical protein	XF1036	5.68	6.11	107601	111097	20	31
3	aconitate hydratase 2	XF0292	5.47	5.49	104811	96019	22	36
4	conserved hypothetical protein	XF2237	5.62	6.3	102858	103030	18	29
5	zinc protease	XF0816	6.15	6.34	100347	108457	21	35
6	conserved hypothetical protein	XF0339	6.05	6.7	93150	97859	30	54
7	hypothetical protein	XFa0017	5.44	5.53	90780	85099	33	68
8	aconitase	XF0290	5.5	6.24	76861	98410	12	15
9	DnaK protein	XF2340	5.18	5.19	73307	68446	8	16
10	hypothetical protein	XF1887	6.9	7.22	69753	79376	16	46
11	dihydrolipoamide dehydrogenase	XF0868	5.62	5.83	66496	63899	9	31
12	30S ribosomal protein S1	XF2436	5	5.03	65140	62879	11	39
13	lipase/esterase	XF0781	5.98	6.23	64988	64268	7	14
14	MOPA OR GROEL	XF0615	5.41	5.45	62414	57757	15	59
15	alkaline phosphatase	XF0657	6.38	6.73	61505	61993	10	35
16	periplasmic protease	XF2241	8.12	8.91	59082	54140	9	24
17	inosine-5'-monophosphate dehydrogenase	XF2430	6.46	6.58	57265	52118	4	10
18	heat shock protein	XF1485	5.36	5.33	56962	51633	9	31
19	hypothetical protein	XF0820	5.92	6.12	56064	57664	12	26
20	3-isopropylmalate dehydratase large subunit	XF2375	5.82	5.9	52571	51183	9	27
21	dihydrolipoamide dehydrogenase	XF1548	6.72	7.68	52571	52119	11	37
22	aminopeptidase A/I	XF0138	6.01	6.08	52117	51723	6	22
23	ATP synthase, beta chain	XF1143	5.05	5	51814	50721	9	31
24	peptidyl-prolyl cis-trans isomerase	XF1186	5.36	5.34	51057	48595	7	36
25	peptidyl-prolyl cis-trans isomerase	XF0838	5.56	5.94	49088	50748	7	23
26	dihydrolipoamide S-succinyltransferase	XF1549	5.82	6.11	44870	42275	6	15
27	elongation factor Tu [a)]	XF2628/XF2640	5.89	5.48	44438	42876	11	59
28	outer membrane protein	XF0343	6.71	8.45	44265	42174	12	39
29	3-oxoacyl-[ACP] synthase II	XF0673	5.36	5.26	43660	43282	13	65
30	succinyl-CoA synthetase, beta subunit	XF2547	5.16	5.08	42796	40988	8	38
31	alcohol dehydrogenase	XF2389	6.01	6.17	42753	37437	7	23
32	aspartate-B-semialdehyde dehydrogenase	XF1371	5.69	5.77	41802	37449	6	35
33	hypothetical protein	XF0007	5.67	6.99	41197	43778	11	32
34	twitching motility protein PilT	XF1633	6.45	6.42	39371	38497	12	60
35	fructose-bisphosphate aldolase	XF0826	5.06	4.98	39166	36131	7	28
36	malate dehydrogenase	XF1211	6.39	7.03	38864	36226	14	43
37	glutaredoxin-like protein	XF2394	4.72	4.51	38259	29544	7	27
38	conserved hypothetical protein	XF0241	5.33	5.27	38000	31013	8	31
39	lipase	XF1253	5.78	5.91	37135	35045	6	28
40	peptidyl-prolyl cis-trans isomerase	XF1605	5.52	5.66	36098	31610	8	42
41	elongation factor Ts	XF2579	5.51	5.79	36098	32056	8	46
42	2-dehydro-3-deoxyphosphooctonate aldolase	XF1289	5.85	6.13	34197	31796	10	31
43	thioredoxin	XF2174	4.97	5.13	33290	37368	5	21
44	hypothetical protein	XF0813	4.35	4.36	33246	22585	5	26
45	dihydroxydipicolinate synthase	XF0963	5.83	6.37	33160	33362	8	40
46	succinyl-CoA synthetase, alpha subunit	XF2548	7.69	8.35	32728	29557	7	35
47	conserved hypothetical protein	XF1361	6.61	6.63	32339	31044	9	32
48	ABC transporter ATP-binding protein	XF1475	6.09	6.13	32253	30774	11	63
49	methionine aminopeptidase	XF0111	5.83	5.92	32209	28507	8	35
50	septum site-determining protein	XF1321	5.17	5.1	31097	28823	5	17
51	3-methyl-2-oxobutanoate hydroxymethyltransferase	XF0229	5.86	5.53	31388	25729	5	27
52	2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehy	XF1726	4.97	4.93	31000	25996	11	39
53	conserved hypothetical protein	XF1840	5.92	6.37	30967	25104	10	47
54	phosphoglyceromutase	XF1893	6.2	6.31	30967	28474	9	57
55	chaperone protein precursor	XF0082	7.61	8.85	30902	29286	11	46
56	toluene tolerance protein	XF0421	7.04	7.07	30869	28024	7	31

Table 1. (Cont.)

57	conserved hypothetical protein	XFa0032	5.7	5.85	30837	27097	5	38
58	elongation factor P	XF2473	5.08	5.05	30544	21380	5	29
59	Conserved hypothetical protein	XFa0050	5.45	5.45	30446	26356	9	53
60	two-component system, regulatory protein	XF0389	5.3	5.27	30056	25323	9	63
61	two-component system, regulatory protein	XF2534	5.21	5.18	29763	25022	8	45
62	subunit C of alkyl hydroperoxide reductase	XF1530	5.5	6.08	28136	22891	5	30
63	NonF-related protein	XF1137	5.53	5.55	27974	24664	8	49
64	2-keto-3-deoxy-6-phosphogluconate aldolase	XF1061	5.43	5.52	27063	23110	7	42
65	superoxide dismutase [MN]	XF2614	5.5	5.6	27063	22643	7	44
66	ATP-dependent Clp protease proteolytic subunit	XF1187	5.23	5.07	26542	22730	7	39
67	inorganic pyrophosphatase	XF2171	4.93	4.97	24626	19792	6	33
68	ribosome recycling factor	XF1051	6.53	6.64	24590	20620	4	26
69	adenylate kinase	XF0275	5.74	5.61	23712	21288	5	44
70	transcription antitermination factor	XF2638	5.74	5.99	23712	21208	9	60
71	biotin carboxyl carrier prot. of acetyl-CoA carb.	XF0048	4.41	4.6	23647	17214	4	33
72	outer membrane protein P6 precursor	XF1896	5.76	8.6	22573	20431	6	31
73	conserved hypothetical protein	XF1412	6.14	6.19	22508	19633	5	30
74	conserved hypothetical protein	XF1382	5.13	5.7	22215	22508	7	31
75	heat shock protein GrpE	XF2341	4.56	4.77	22085	22493	3	18
76	ABC transporter vitamin B12 uptake permease	XF1604	7.37	8.51	21126	21409	5	27
77	single-stranded DNA binding protein	XF1392	5.73	5.9	21040	18211	7	49
78	MOPA OR GROEL, fragment	XF0615	4.73	5.45	20493	57757	7	20
79	tryptophan repressor binding protein	XF1133	6.22	6.29	20321	20466	6	68
80	hypothetical protein [a)]	XF0687/XF2521	5.22	4.51	19890	8089	5	69
81	transcriptional elongation factor	XF1108	5.28	5.09	19574	16988	11	74
82	low molecular weight heat shock protein	XF2234	5.5	5.55	19516	17858	6	63
83	conserved hypothetical protein	XF0196	6	6.91	18826	19896	12	53
84	bacterioferritin	XF0395	5.06	4.79	18021	17851	4	36
85	conserved hypothetical protein	XF0363	7.05	9.69	17561	25806	8	39
86	single-strand binding protein	XFa0061	5.28	5.22	17475	15384	6	36
87	hypothetical protein	XF2155	5.41	5.58	17044	15065	5	35
88	peptidyl-prolyl cis-trans isomerase	XF1212	6.54	6.58	16843	17788	4	24
89	bacterioferritin comigratory protein	XF0961	6.21	6.28	16613	17813	5	41
90	fimbrial protein	XF2542	5.57	8.46	16153	15034	4	42
91	50S ribosomal protein L9	XF2559	6.47	6.31	15894	15726	7	64
92	RNA polymerase omega subunit	XF1502	4.54	4.64	14859	11235	6	76
93	50S ribosomal protein L7/L12	XF2634	4.17	4.96	13513	12713	6	57
94	translation initiation inhibitor	XF0353	5.09	4.99	13190	13573	5	69
95	conserved hypothetical protein	XF1810	5.68	5.93	9764	12953	10	79
96	10kDa chaperonin	XF0616	5.56	5.8	9361	10092	4	67
97	hypothetical protein	XF0284	6	6.09	8998	12556	5	60
98	temperature acclimation protein B	XF2622	6.44	6.01	7870	9436	4	58
	SDS-PAGE							
99	surface protein	XF1981	-	5.12	>300000	118447	35	33
100	bacterioferritin	XF0395	-	5.06	>300000	17851	8	51
101	chemotaxis-related protein kinase	XF1952	-	4.67	207158	193985	28	16
102	conserved hypothetical protein	XF2091	-	6.12	170651	185737	43	34
	conserved hypothetical protein	XF1252	-	9.05		178921	40	29
103	glutamate synthase, alpha subunit	XF2710	-	6.34	165641	163598	37	31
104	RNA polymerase beta subunit	XF2633	-	5.63	157767	154782	29	32
105	RNA polymerase beta subunit	XF2632	-	8.12	152040	158719	24	22
106	phosphoribosylformyl. synthetase	XF1423	-	5.89	139155	143721	26	25
107	PilY1 gene product	XF0032	-	7.25	132713	133494	19	18
	hypothetical protein	XF1232	-	8.92		127417	18	15
108	PilY1 gene product	XF1224	-	7.9	128418	132379	21	18
	transcription-repair coupling factor	XF0045	-	5.93		132405	23	21
109	carbamoyl-phosp. synth. large chain	XF1107	-	5.11	122692	117384	25	28
	hypothetical protein	XF2734	-	6.43		129883	28	31
110	initiation factor IF-2	XF0235	-	7.12	115939	97010	18	17
111	conserved hypothetical protein	XF0550	-	6.52	110350	113870	31	39

a) Cases where mass spectrometry analysis could not define a single corresponding gene because of the existence of two or more identical or highly similar genes in the genome.

Table 2. List of the proteins identified in the extracellular fraction. Corresponding spots or bands are indicated in in Figures 2A, 2B and 2C.

SPOT	PRODUCT	GENOME ACCESSION CODE	pI obs	pI prev	MW obs	MW prev	N° of peptides matched	Sequence Coverage (%)
2-DE								
1	zinc protease	XF0816	6.15	6.34	100347	108457	19	31
2	serine protease	XF1026	4.85	5.6	84455	95212	17	24
3	serine protease	XF1851	5.04	6.15	73259	105429	7	9
4	aconitase	XF0290	5.54	6.24	76565	98410	7	10
5	hypothetical protein	XF1887	6.9	7.22	69753	79376	16	46
6	malate dehydrogenase	XF1211	6.39	7.03	38864	36226	15	52
7	lipase	XF1253	5.78	5.91	37135	35045	6	24
8	peptidyl-prolyl cis-trans isomerase	XF1605	5.53	5.66	36098	31610	8	42
9	conserved hypothetical protein	XFa0032	5.59	5.85	30707	27097	5	38
10	conserved hypothetical protein	XF2170	5.17	5.27	28169	22754	9	52
11	NonF-related protein	XF1137	5.53	5.55	27974	24664	8	49
12	subunit C of alkyl hydroperoxide reductase	XF1530	5.5	6.08	28136	22891	5	30
13	superoxide dismutase [MN]	XF2614	5.5	5.6	27063	22643	7	44
14	ATP-depend. Clp protease proteolytic subunit	XF1187	5.23	5.07	26542	22730	5	35
15	tryptophan repressor binding protein	XF1133	6.22	6.29	20321	20466	8	41
16	conserved hypothetical protein	XF0196	6	6.91	18826	19896	6	34
17	fimbrial protein	XF2542	5.57	8.46	16153	15034	4	27
SDS-PAGE								
18	fimbrial adhesin precursor	XF0078	-	8.56	>300000	36854	6	30
	fimbrial adhesin precursor	XF0077	-	6.31		37601	6	33
19	bacterioferritin	XF0395	-	5.06	>300000	17851	5	40
20	bacteriocin	XF2407	-	4.31	147070	218777	22	34
21	hypothetical protein	XF2445	-	6.45	112186	118472	12	12
22	PilY1 gene product	XF0032	-	7.25	109619	133494	22	20
	conserved hypothetical protein	XF0550	-	6.52		113870	24	31
23	zinc protease	XF0816	-	6.15	103630	108457	31	55
24	conserved hypothetical protein	XF2713	-	6.16	97427	96309	23	30
	aconitase	XF0290	-	6.24		98410	10	14
25	peptidase	XF1479	-	6.06	89674	86186	24	38
	hemolysin-type calcium binding protein [a)]	XF0668/XF1011	-	4.41		128404	13	12
	conserved hypothetical protein	XF2551	-	6.48		98694	13	17
26	serine protease	XF1026	-	5.6	81683	95212	15	22
	serine protease	XF1851	-	6.15		105429	12	14
27	polynucleotide phosphorylase	XF0239	-	5.56	78187	78146	14	20
	catalase/peroxidase	XF2232	-	6.22		86891	14	19
	outer membrane usher protein precursor	XF0081	-	9.48		98163	19	25
	peptidyl-dipeptidase	XF1944	-	6.04		80508	22	37
28	hypothetical protein	XF1887	-	7.22	74191	79375	15	24
	hypothetical protein	XF1451	-	8.95		71749	13	20
	dihydrolipoamide dehydrogenase	XF0868	-	5.83		63899	6	15
29	soluble lytic murein transglycosyl. precur.	XF1363	-	8.96	71028	79951	13	19
30	hypothetical protein	XF2283	-	5.91	36067	34299	13	49
31	peptidyl-prolyl cis-trans isomerase	XF0644	-	9.43	30932	24854	5	30
	conserved hypothetical protein	XF0363	-	9.69		25806	4	25
	conserved hypothetical protein	XFa0050	-	5.45		26356	4	27
	hypothetical protein	XFa0064	-	10.65		28479	5	31
32	superoxide dismutase [MN]	XF2614	-	5.6	26068	22643	4	23
	hypothetical protein	XF1803	-	9.54		21196	7	49
33	hypothetical protein	XF1010	-	6.2	21202	19881	9	53
34	fimbrial protein	XF2542	-	8.46	17549	15034	4	42
	carbonic anhydrase	XF2095	-	6.23		22765	4	35
35	organic hydroperoxide resistance protein	XF1827	-	5.82	14995	14899	7	64
	peptidoglycan-assoc. outer memb. lipoprot. prec.	XF1547	-	10.03		15653	6	51
	hypothetical protein	XF0898	-	9.3		12973	5	52
	hypothetical protein	XF1911	-	9.06		17131	4	40

a) Cases where mass spectrometry analysis could not define a single corresponding gene because of the existence of two or more identical or highly similar genes in the genome.

Legend for Figures

Figure 1. Gel electrophoresis analysis of whole cell protein extract from *X. fastidiosa* grown for 21 days on solid BCYE media. (A) 2-DE separation was performed using an IPG gel with a pH range of 3-10 (first dimension) and 12.5 %C acrylamide gels (second dimension). Proteins were visualized using silver staining and the indicated spots were identified by peptide mass fingerprinting. Table 1 details the identification of these spots. (B) For the analysis of high molecular weight proteins, the same protein extract was separated by SDS-PAGE on a 9 %C acrylamide gel and proteins were stained with coomassie blue. Bands present in the molecular weight region (>110 kDa) were subjected to protein identification by peptide mass fingerprinting. Identified proteins from the indicated bands are shown in Table 1.

Figure 2. Gel electrophoresis analysis of extracellular protein fraction. The sample preparation procedure was developed to extract secreted proteins as well as proteins weakly attached to the cell surface: bacteria grown for 21 days on solid BCYE media were washed with a 50 mM NaCl containing buffer and proteins extracted in the supernatant were then precipitated. (A) 2-DE separation was performed using an IPG gel with a pH range of 3-10 (first dimension) and 12.5 %C acrylamide gels (second dimension). Proteins were visualized with silver staining and the indicated spots were identified by peptide mass fingerprinting. Table 2 describes the resulted identifications. The numbers in parenthesis indicate the corresponding protein spots in Figure 1A. Spots indicated with the letter "C" correspond to the most abundant intracellular proteins (C1: XF2340, C2: XF0615 and C3: XF0343), some of them also present as fragments. Because of their high abundance in the cell, it is not clear whether they represent contamination from residual cell lysis or are in fact a fraction of that proteins that are

being surface exposed. (B) For the analysis of high molecular weight proteins, the same extracellular protein fraction was separated by SDS-PAGE on a 9 %C acrylamide gel and proteins were stained with coomassie blue. Bands present with molecular weights over 70 kDa were subjected to protein identification by peptide mass fingerprinting. Identified proteins from the indicated bands are shown in Table 1. (C) Since the extracellular protein fraction was not too complex, protein identification was also performed for bands present in the 10 to 40 kDa molecular weight region of 12.5 %C acrylamide gels stained with coomassie blue.

Figure 3. Functional class distribution of proteins identified in the extracellular fraction shown in Table 2.

Figure 4. Codon bias distribution for total genome orfs (—) and for the 50 most abundant proteins (- - -) in (A) *E. coli* and (B) *X. fastidiosa*. Determination of the 50 most abundant proteins was based on the *E. coli* Swiss 2D-PAGE database (http://www.expasy.ch/cgi-bin/map2/def?ECOLI) in the case of *E. coli* and on the 2-DE profile shown in Figure 1A for *X. fastidiosa*. For *E. coli*, it is possible to observe that abundant proteins show higher CB value in relation to most of the genome orfs, which was not observed for *X. fastidiosa*. The measure of directional codon bias used was based on the Codon Bias Index [12].

Figure 5. Codon bias distribution of total genome orfs for 24 bacteria with complete sequenced genome. On the upper side, are the organisms presenting highest CB values for most of the genome orfs ("ameliorated" genomes). *X. fastidiosa* was observed to be the only bacteria to have the largest fraction of its genes in the region of CB value below zero. The measure of directional codon bias use was based on the Codon Bias Index [12].

Figure 1

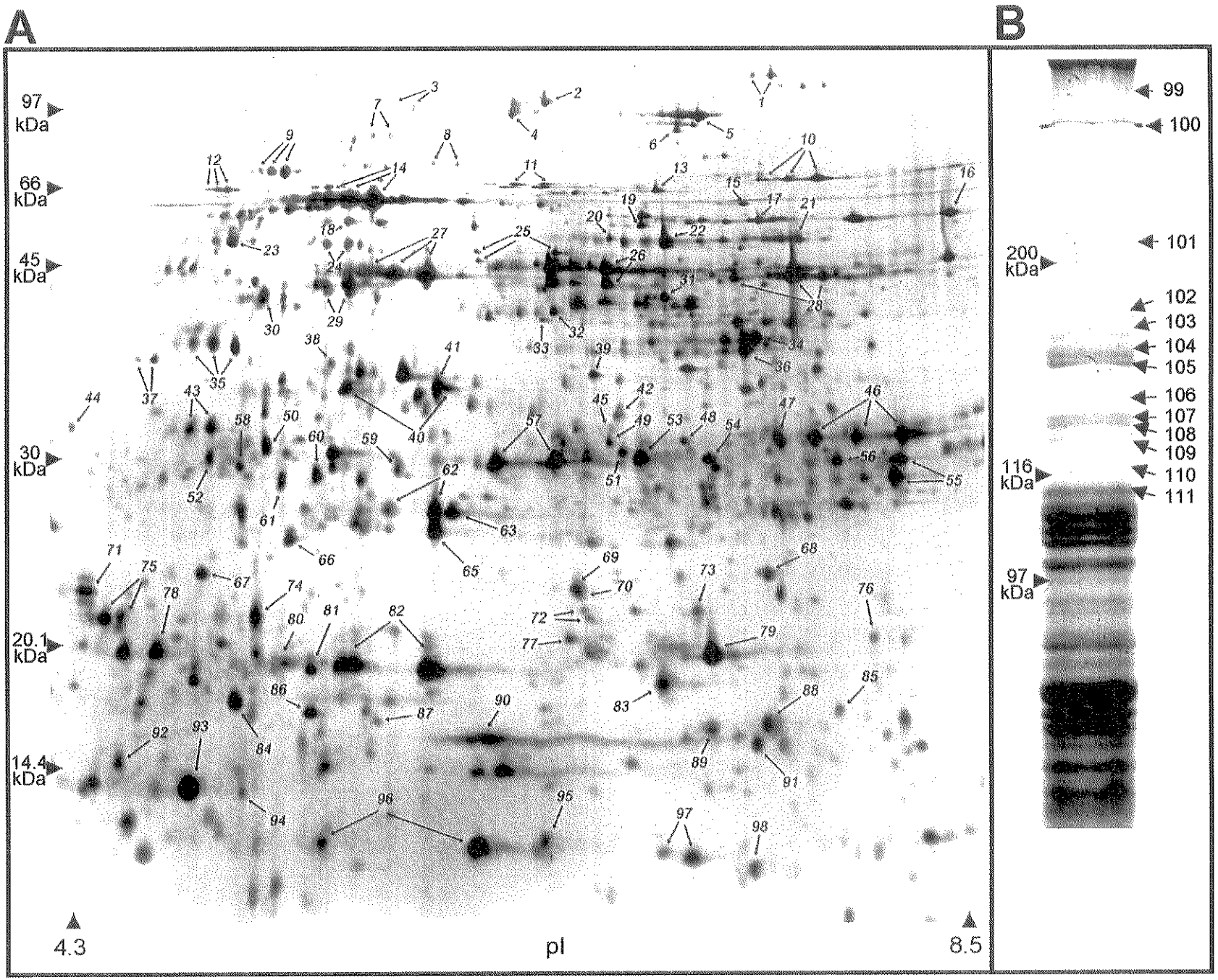

A
B
97 kDa
66 kDa
45 kDa
30 kDa
20.1 kDa
14.4 kDa
4.3
pI
8.5
200 kDa
116 kDa
97 kDa
99
100
101
102
103
104
105
106
107
108
109
110
111

Figure 2

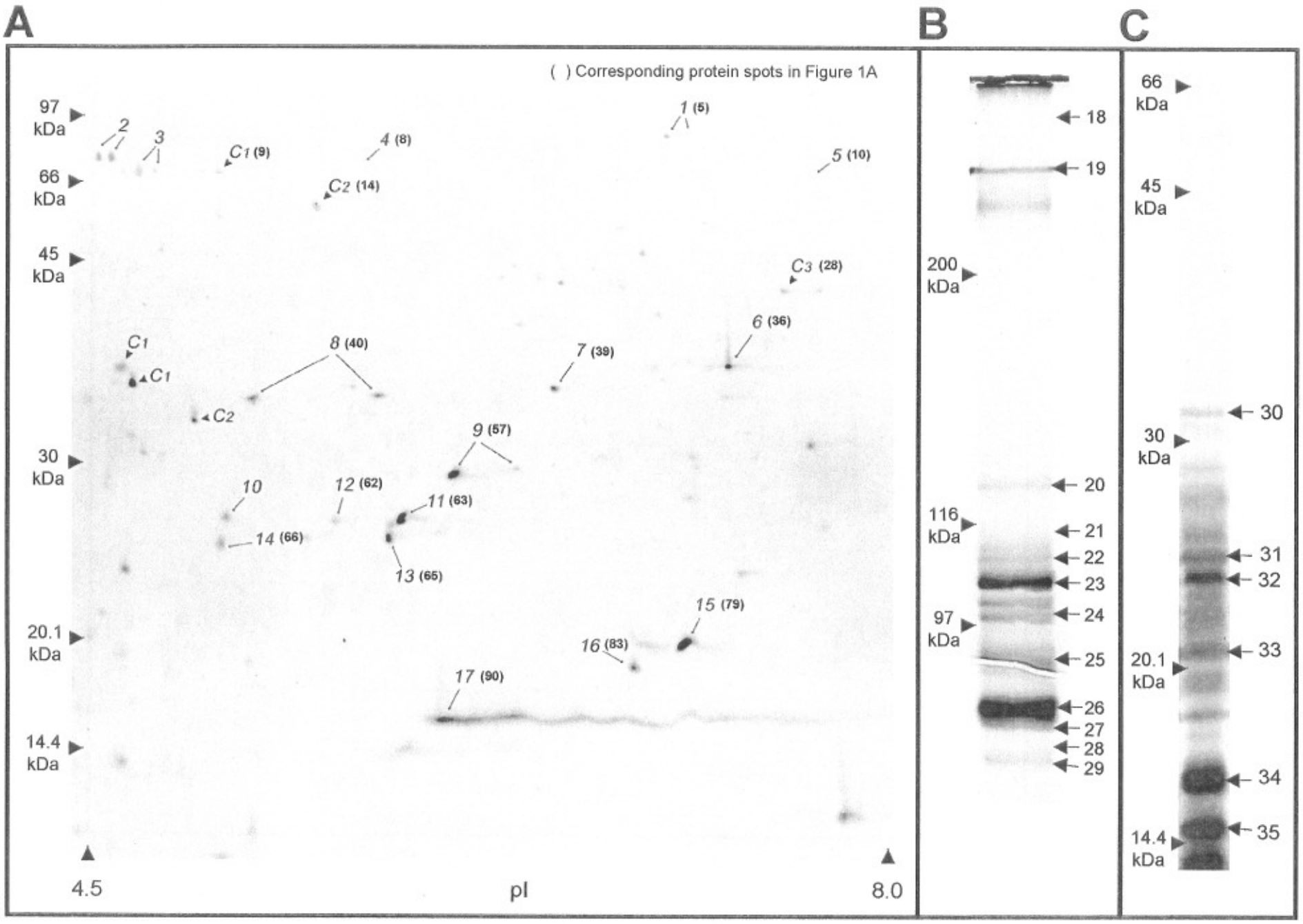

A
B
C
( ) Corresponding protein spots in Figure 1A
97 kDa
66 kDa
45 kDa
30 kDa
20.1 kDa
14.4 kDa
1 (5)
2
3
C1 (9)
4 (8)
5 (10)
C2 (14)
C3 (28)
6 (36)
C1
C1
8 (40)
7 (39)
C2
9 (57)
10
12 (62)
11 (63)
14 (66)
13 (65)
15 (79)
16 (83)
17 (90)
4.5
pI
8.0
200 kDa
116 kDa
97 kDa
18
19
20
21
22
23
24
25
26
27
28
29
66 kDa
45 kDa
30 kDa
20.1 kDa
14.4 kDa
30
31
32
33
34
35

Figure 3

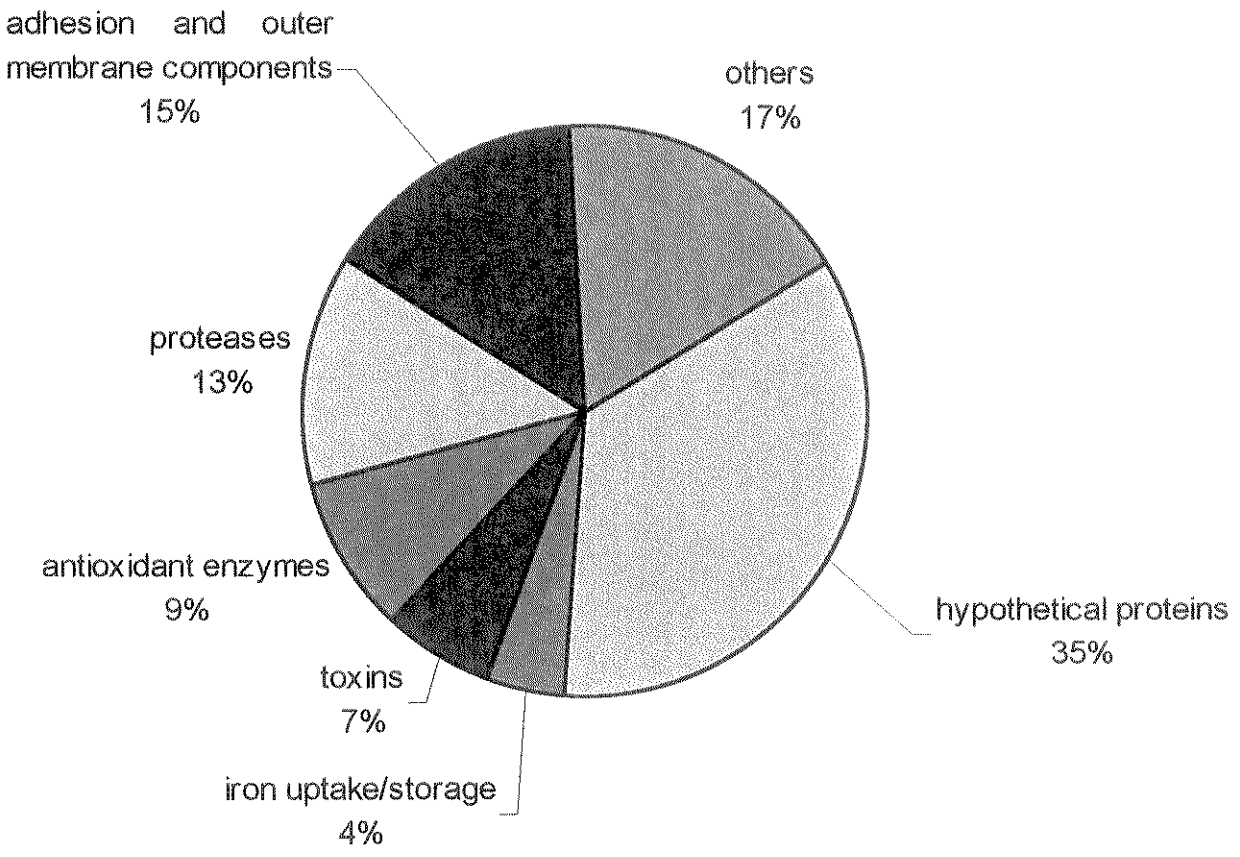

adhesion and outer membrane components 15%
others 17%
proteases 13%
antioxidant enzymes 9%
hypothetical proteins 35%
toxins 7%
iron uptake/storage 4%

Figure 4

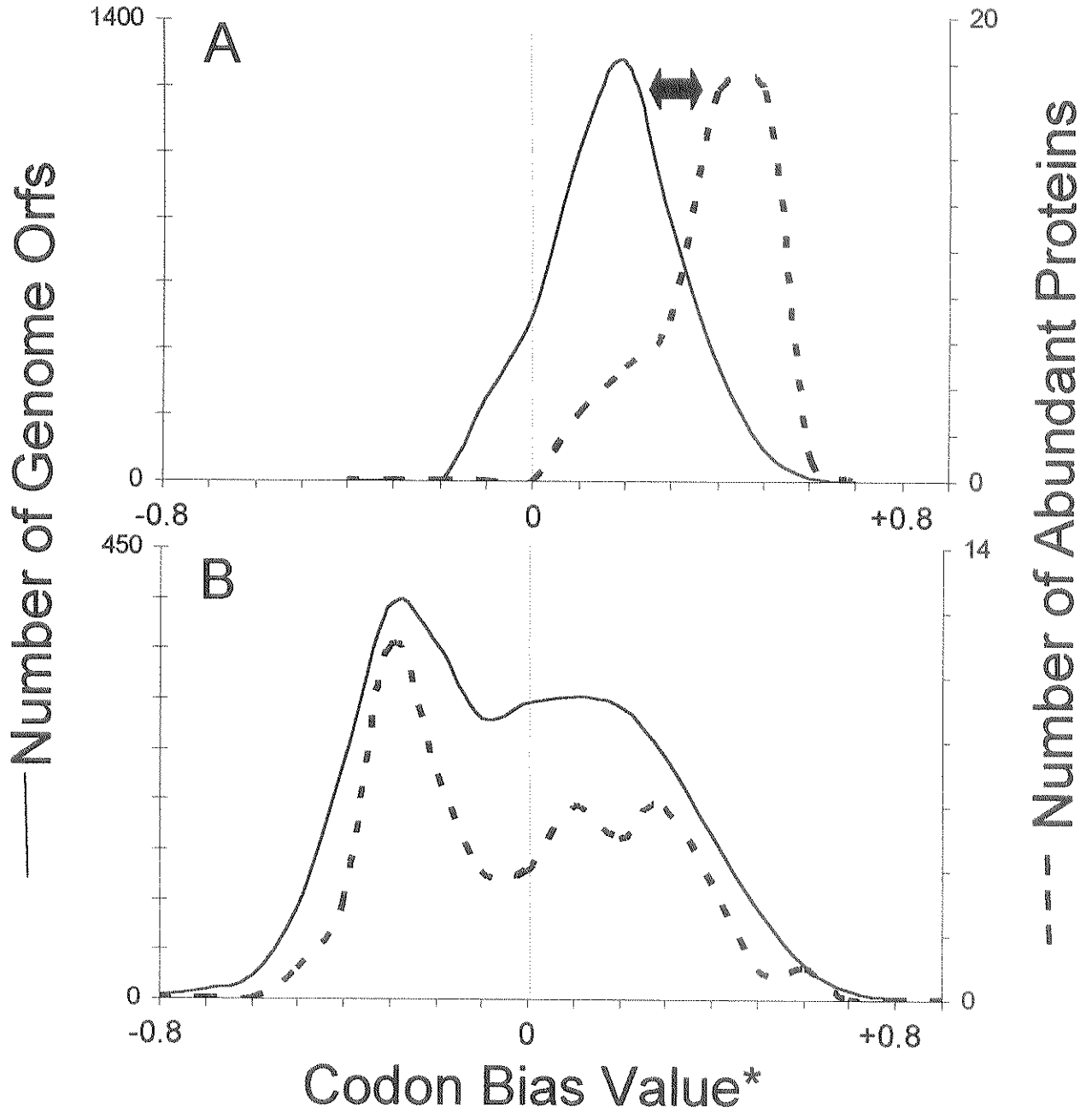
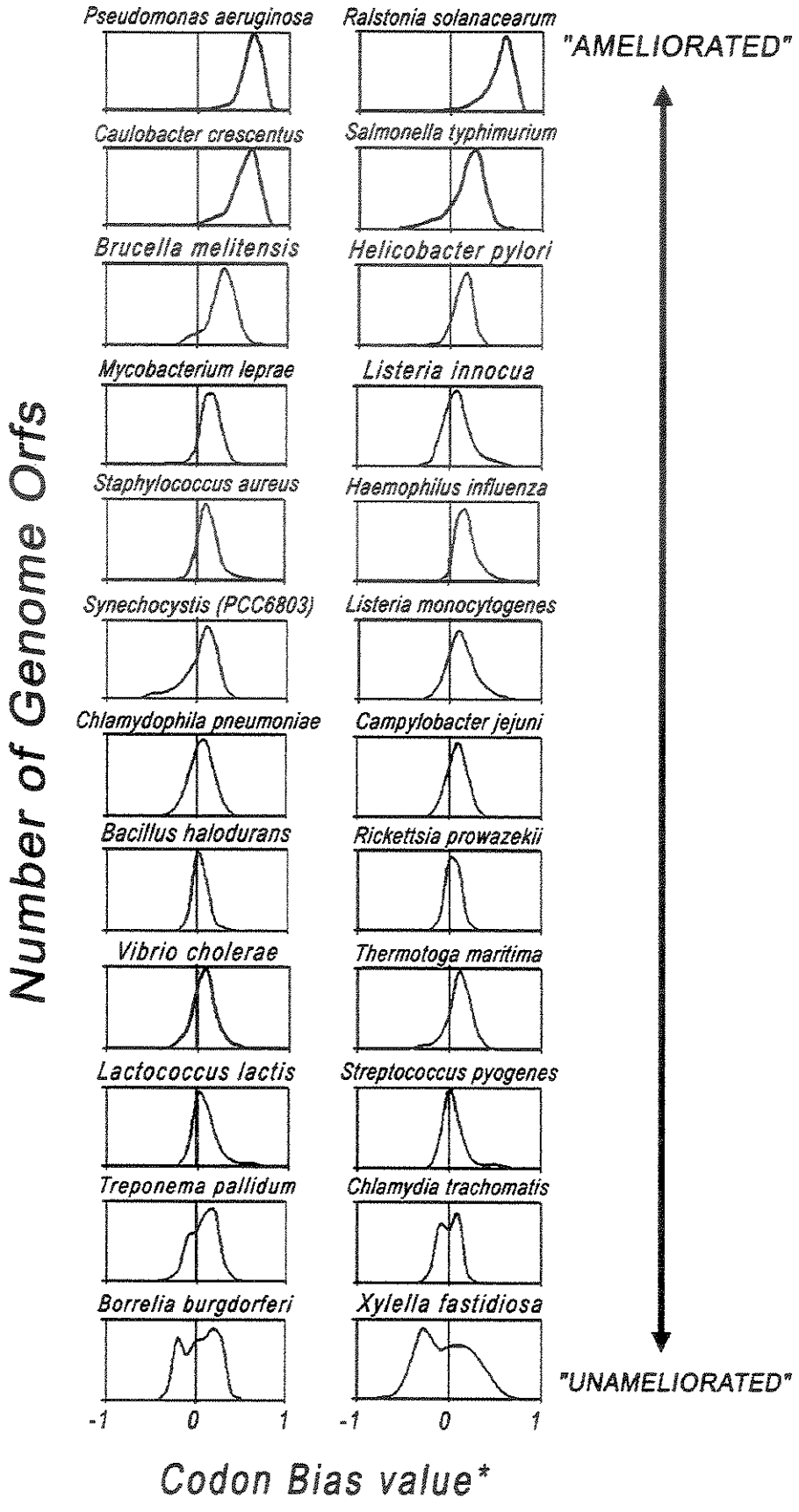

1400
A
20
0
0
-0.8
0
+0.8
450
B
14
0
0
-0.8
0
+0.8
Codon Bias Value*
Number of Genome Orfs
Number of Abundant Proteins

Figure 5


Pseudomonas aeruginosa
Ralstonia solanacearum
"AMELIORATED"
Caulobacter crescentus
Salmonella typhimurium
Brucella melitensis
Helicobacter pylori
Mycobacterium leprae
Listeria innocua
Staphylococcus aureus
Haemophilus influenza
Synechocystis (PCC6803)
Listeria monocytogenes
Chlamydophila pneumoniae
Campylobacter jejuni
Bacillus halodurans
Rickettsia prowazekii
Vibrio cholerae
Thermotoga maritima
Lactococcus lactis
Streptococcus pyogenes
Treponema pallidum
Chlamydia trachomatis
Borrelia burgdorferi
Xylella fastidiosa
"UNAMELIORATED"
-1
0
1
-1
0
1
Number of Genome Orfs
Codon Bias value*

2. Resultados suplementares

2.1. Proteínas extracelulares estão presentes majoritariamente em sua forma mais ácida

Proteínas detectadas no extrato celular total como isoformas com diferentes pIs e que também foram detectadas na fração extracelular, estavam majoritariamente em sua forma mais negativa quando translocadas para o exterior de célula (Figura 9). Isto foi observado para as proteínas XFa0032, XF1530, XF0290 e XF0816.

Provavelmente, isto é resultado de fosforilação. Fosforilações são abundantes em bactérias e resultam na adição de carga(s) negativa(s), tornando uma proteína mais ácida. Apenas uma pequena fração das proteínas do extrato celular estaria fosforilada, enquanto uma fração maior destas proteínas estaria fosforilada na fração extracelular. Apesar de ser necessário utilizar outros experimentos mais específicos para se comprovar a existência de fosforilação nas quatro proteínas citadas, já foi demonstrado que as fosforilações provocam migrações eletroforéticas alteradas (para a região mais ácida) em proteínas separadas por 2DE, resultando em perfis similares aos mostrados na Figura 9 [55].

É possível que exista um mecanismo pós-traducional de regulação de secreção de proteínas, e que envolva fosforilação. Uma outra possibilidade é que quinases ativas estejam presentes no meio extracelular. Não foi detectada nenhuma quinase na fração extracelular, mas é conhecido que quinases estão presentes em quantidades extremamente inferiores às proteínas mais abundantes, e não são geralmente detectáveis por 2DE.

[55] Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. et al. A sampling of the yeast proteome. *Mol. Cell Biol.* 1999, 19, 7357-7368.

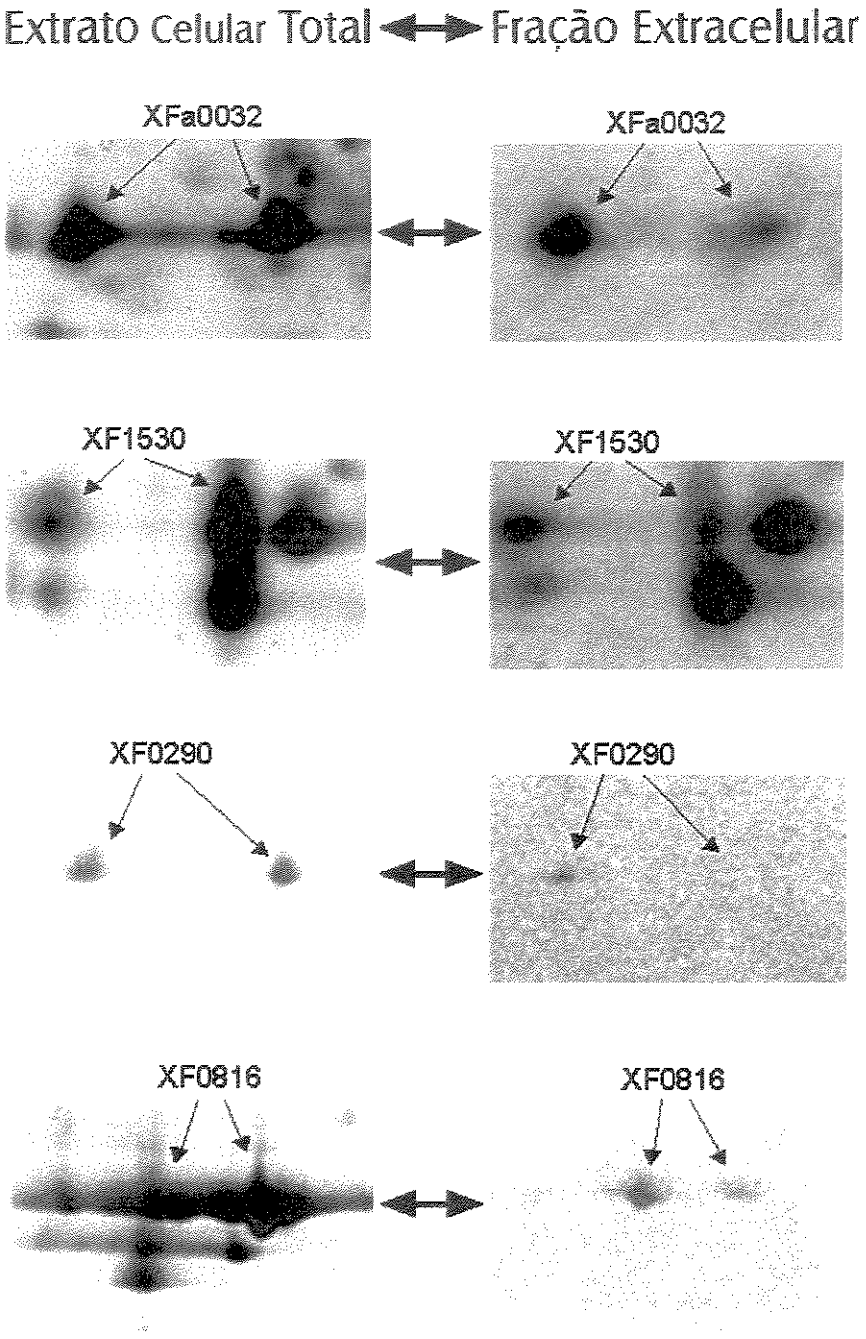


FIGURA 9: Comparação de regiões específicas do mapa de proteínas do extrato toal e do mapa de proteínas extracelulares. É possível notar a predominância da forma mais negativa das isoformas no mapa de proteínas extracelulares.

2.2. Identificação de proteínas da *Xylella fastidiosa* específicas do crescimento em biofilme

Biofilmes podem ser definidos como comunidades estruturadas de micróbios agregados, envoltos numa matriz polimérica própria e aderidos numa superfície [56]. A formação de biofilme é um fenômeno típico de ambientes aquosos e ocorre pela adesão de bactérias na interface sólido-líquido[57, 58].

Agregados de *X. fastidiosa* com características típicas de biofilmes foram detectados tanto no xilema da planta como no sibário do inseto[59,60]. A formação de biofilme pela *X. fastidiosa* foi recentemente caracterizada *in vitro*[61] e provavelmente representa um fator importante para sua virulência.

Dada sua natureza de formar biofilme, a *X. fastidiosa* cultivada em meio líquido se agrega e adere nas paredes do frasco de cultura (Figura 10B), numa formação típica de biofilme. Com o objetivo de identificar proteínas específicas deste crescimento em biofilme, foi feita uma comparação do perfil protéico da bactéria crescida como colônias isoladas em meio sólido de cullltura com o perfil protéico dela crescida em biofilme no meio líquido.

O procedimento utilizado para crescimento das bactérias em meio sólido foi similar ao descrito no trabalho publicado. O meio usado para crescimento em meio líquido também foi o BCYE, porém as células foram crescidas por quatro dias sob agitação constante em meio sem agar e carvão ativado.

Na Figura 11A são mostrados os mapas 2DE representativos de proteínas do extrato total da *X. fastidiosa* crescida em meio sólido e líquido. Os círculos indicam regiões dos mapas contendo proteínas com grande alteração em abundância. Oito repetições da comparação mostrada na Figura 11A foram feitas e só foram consideradas as alterações acima de 200 % em abundância (medida por área vezes X intensidade densitométrica do *spot*) reprodutíveis em todas as repetições.

---

[56] Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. et al. Microbial biofilms. *Annu. Rev. Microbiol.* 1995, 49, 711-45.

[57] Van Loosdrecht, M.C., Lyklema, J., Norde, W., Zehnder, A.J. Influence of interfaces on microbial activity. *Microbiol. Rev.* 1990, 54, 75-87.

[58] Watnick, P., Kolter, R. Biofilm, city of microbes. *J. Bacteriol.* 2000, 182, 2675-2679.

[59] Hopkins, D. L. Xylella fastidiosa: xylem-limited bacterial pathogen of plants. *Annu. Rev. Phytopathol.* 1989, 27, 271-290.

[60] Purcell, A. H., Hopkins, D. L. Fastidious xylem-limited bacterial plant pathogens. *Annu. Rev. Phytopathol.* 1996, 34, 131-151.

[61] Marques, L. R., Ceri, H., Manfio, G. P., Reid, D. M. Characterization of biofilm formation by xylella fastidiosa in vitro. *Plant Dis.* 2002, 86, 633-638.

Pela análise dos mapas 2DE obtidos, ficou clara a maior expressão das proteínas XF2539, XF0156 e XF2151 no crescimento em biofilme (Figuras 11B, 11C e 11D, respectivamente) e da XF2542 no crescimento em meio sólido (Figura 11E). A Tabela 2 traz uma descrição resumida das proteínas baseada em análises de bancos de dados e dados sobre a identificação por espectrometria de massas.

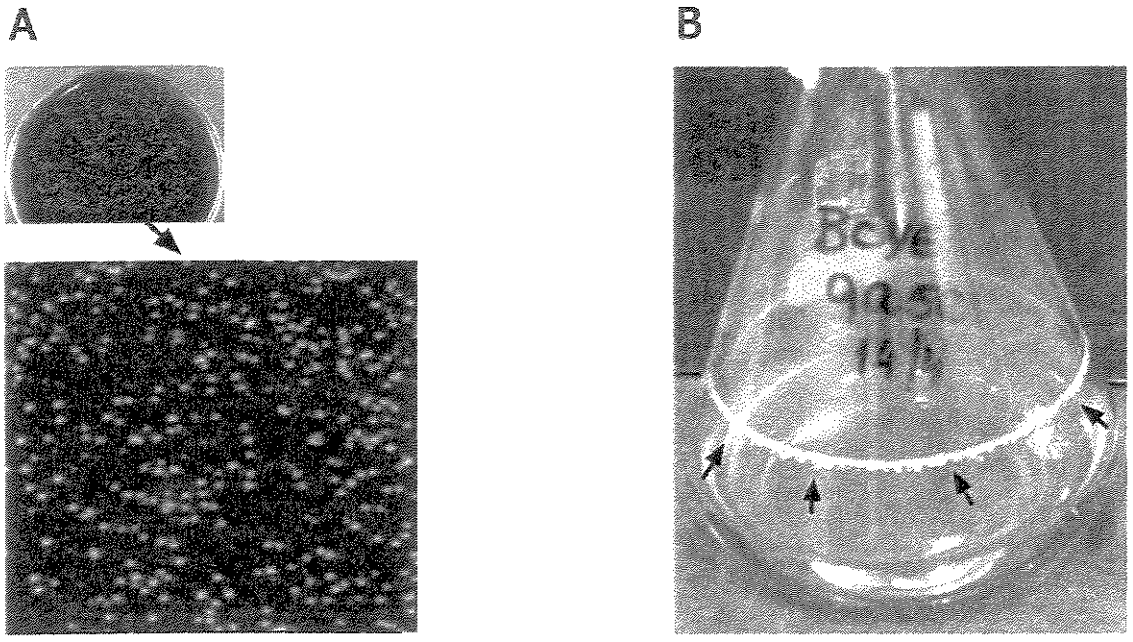


FIGURA 10: Fotos mostrando o crescimento da *X. fastidiosa* em meio (A) sólido e (B) líquido de cultura. Em meio sólido (A), a bactéria cresce como colônias isoladas e em meio líquido (B) ela se agrega e adere nas paredes do frasco de cultura (indicado pelas setas), numa formação típica de biofilme.

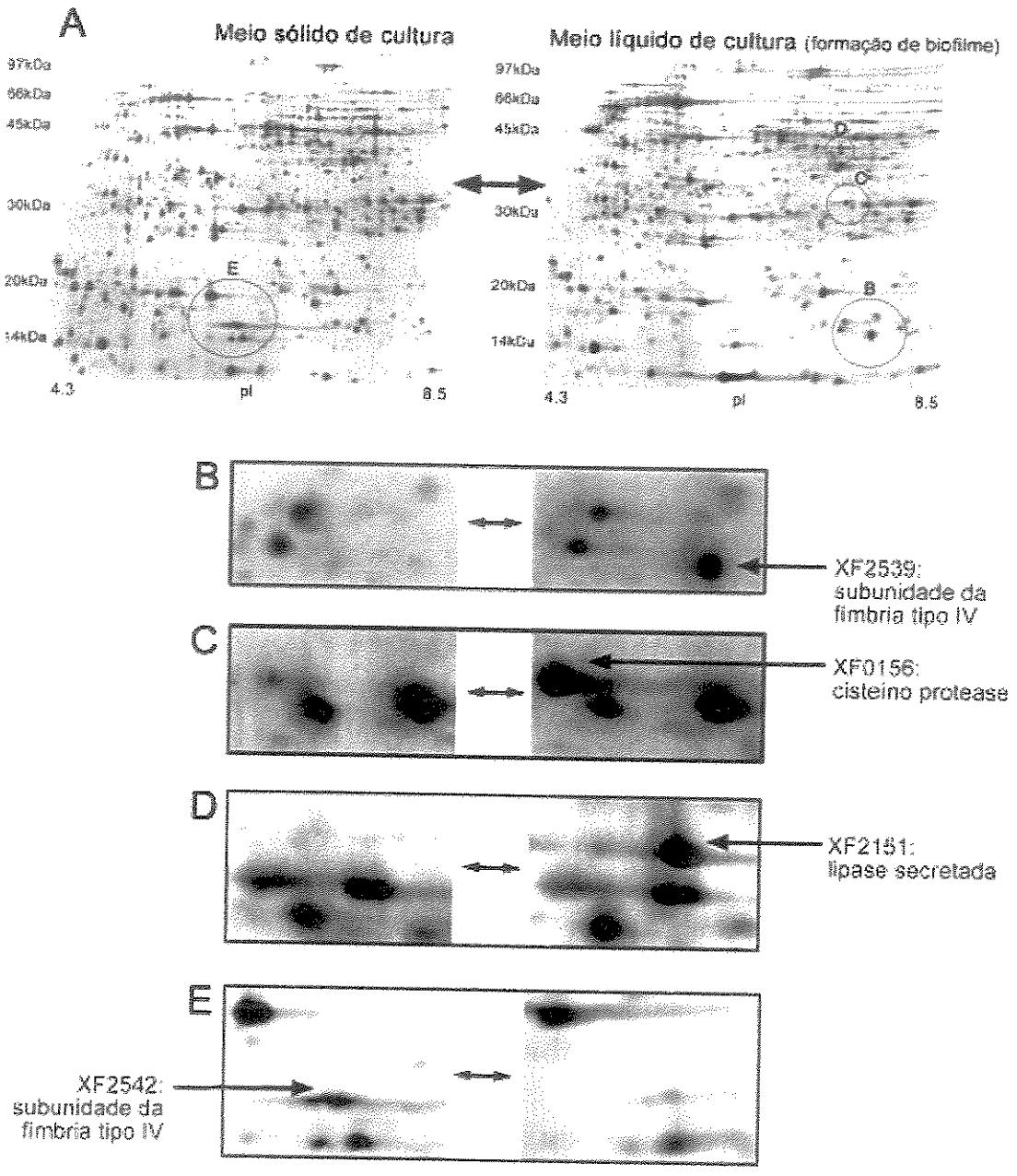


FIGURA 11: Análise diferencial por eletroforese de duas dimensões do perfil protéico de proteínas do extrato total da *X. fastidiosa* crescida em meio BCYE sólido e líquido (A). As regiões demarcadas com os círculos contêm proteínas que apresentam significativas alterações em abundância e são mostradas em destaque em (B), (C), (D) e (E).

Tabela 2. Descrição das proteínas com expressão significativamente alterada na comparação do crescimento em meio BCYE sólido e líquido, como mostrado na Figura 3.

Proteínas mais expressas em meio líquido			
Número de acesso	**Descrição**	**Identificação por espectrometria de massas**	
		N° de peptídios identificados	**Cobertura da sequência (%)**
XF2539	Subunidade da fímbria do tipo IV.	6	60
XF0156	Cisteíno protease. Não apresenta similaridade significativa com proteína conhecida, mas tem um domínio de peptidase C1. Proteínas com este domínio são chamadas de "papaína-like" e representam cisteíno proteases secretadas.	8	39
XF2151	Lipase LIP. Anotada como proteína hipotética no banco de dados genômico da *X. fastidiosa*, também não apresenta similaridade significativa com proteínas de função conhecida, mas tem um domínio de lipase secretada do tipo LIP.	16	46
Proteínas mais expressas em meio sólido			
Número de acesso	**Descrição**	**Identificação por espectrometria de massas**	
		N° de peptídios identificados	**Cobertura da sequência (%)**
XF2452	Subunidade da fímbria do tipo IV. Apresenta alta similaridade com a XF2539 detectada no crescimento em meio líquido.	4	42

Subunidade da fimbria do tipo IV (XF2539 e XF2542)

Os genes XF2539 e 2542 apresentam homologia com genes que codificam subunidades formadores da fímbria do tipo IV. Estas fímbrias são formadas por subunidades protéicas idênticas organizadas numa configuração de hélice contendo cinco subunidades por volta. São filamentos polares flexíveis de aproximadamente 5.4 nanômetros em diâmetro e 2.5 micrômetros em comprimento, que apresentam importantes funções na adesão bacteriana, secreção de proteases e motilidade[62].

Os resultados mostram que o gene da subunidade da fímbria do tipo IV expresso na *X. fastidiosa* é diferente no crescimento em meio sólido e no meio líquido. Quando crescida em meio sólido, a bactéria apresenta expressão do gene XF2542. Na condição de formação de biofilme em meio líquido, a expressão do gene XF2542 parece ser repremida e a bactéria passa a expressar em alta abundância o gene XF2539. Estes dois genes apresentam alta similaridade entre si (Figura 12) e estão localizados muito próximos no cromossomo, apenas 1297 pares de base um do outro (Figura 13).

```
 Score =  177 bits (445), Expect = 2e-46
 Identities = 92/145 (63%), Positives = 112/145 (76%), Gaps = 2/145 (1%)

XF2542: 1   MKKQQGFTLIELMIVIAIIAILAAIALPMYQNYVARSQIAAALAEITPGKVQAEIRIADG 60
           MKKQQGF LIELMIVIAIIA+LAAIALPMYQNYVARSQ+ AALA+ITPGKVQAE  IADG
XF2539: 1   MKKQQGFNLIELMIVIAIIAVLAAIALPMYQNYVARSQLTAALADITPGKVQAESLIADG 60

XF2542: 61  QAATTPNAIGLRAPTPRCGTIVVDIAPSAASAITCTMIGNAQVNNQTITLTRIADNNAGQ 120
           ++ +  + IGLR  T RCG I V +  +  + ITC + GN+QVN++TI   R +DN+AG
XF2539: 61  KSTSNASDIGLRTDTTRCG-ITVKVDAAGTANITCKVKGNSQVNDKTIAWDRTSDNSAGT 119

XF2542: 121 GGVNTGGNWTCTTTAPA-ALTPAGC 144
            GVN GG WTC++T  + AL P+GC
XF2539: 120 NGVNNGGVWTCSSTVTSDALRPSGC 144
```

FIGURA 12: Alinhamento das sequências de aminoácidos referente aos genes XF2542 e XF2539, homólogos da subunidade da fímbria do tipo IV.

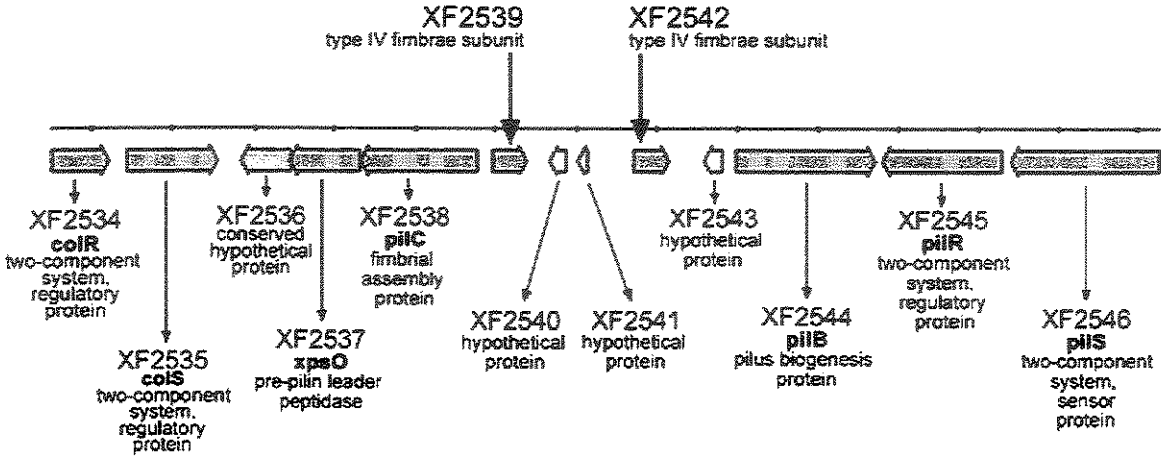


FIGURA 13: Mapa da região cromossômica contendo os genes XF2539 e XF2542.

Possivelmente, esta alteração de expressão resulta numa fímbria com diferentes características estruturais e funcionais, sendo que a fímbria formada pela subunidade XF2539 estaria relacionada à capacidade de adesão da *X. fastidiosa* às paredes do frasco de cultura e formação de biofilme.

As maiores similaridades dos genes XF2539 e XF2542 ocorrem com subunidades de fímbria tipo IV (também chamadas de fimbrilinas pilA ou FimA) de espécies de *Xanthomonas* e *Pseudomonas.* Diferentemente de *Xanthomonas spp* e *Xylella fastidiosa*, em *Pseudomonas* existe apenas um gene codificador da subunidade da fímbria tipo IV. A existência de dois genes para a subunidade da fímbria tipo IV e de um mecanismo específico de regulação da expressão destes genes em diferentes condições de crescimento poderia ser um fator importante para a sobrevivência da *X. fastidiosa.*

É interessante como a expressão de dois genes localizados tão próximos é antagonicamente regulada. Na Figura 13 é possível observar a presença de genes do sistema regulatório de dois componentes proximamente localizados abaixo ("downstream") e acima ("upstream") dos genes da subunidade da fímbria tipo IV. O sistema de dois componentes funciona como um mecanismo básico de integração estímulo-resposta, possibilitando que organismos "sintam" e respondam a alterações

[62] Parge, H. E., Forest, K. T., Hickey, M. J., Christensen, D. A. et al. Structure of the fibre-forming protein pilin at 2.6 A resolution. *Nature* 1995, 378, 32-38.

em diversas condições ambientais[63]. É bem descrito que os genes XF2545 (pilR) e XF2546 (pilS), localizados "downstream", regulam a expressão da fímbria do tipo IV[64], porém não se sabe se o genes XF2534 e XF2535 estão relacionados à esta regulação também. Estes genes apresentam alta similaridade com genes do sistema de dois componentes da subclasse OmpR-PhoB. A maior similaridade (e-73) foi obtida com os genes colR e colS, respectivamente, de *Pseudomonas fluorescens* que ainda são pouco estudados. Já foi mostrado que mutações nestes genes afetam a capacidade da bactéria colonizar raízes de plantas[65], porém, não se sabe ainda quais genes são regulados por eles. A proximidade dos genes colR e ColS aos genes da subunidade da fímbria tipo IV em *X. fastidiosa* sugere que eles também possam regular a expressão da fímbria, da mesma forma que os genes pilS e pilR.

É possível concluir que a *X. fastidiosa* possui dois genes homólogos à subunidade da fímbria tipo IV cuja expressão pode ser regulada em função da condição de crescimento da bactéria. Maiores estudos são necessários para melhor entender o mecanismo de regulação de expressão destes genes e a importância funcional deste sistema regulatório, principalmente no processo de formação de biofilme.

Cisteíno protease (XF0156)

A *Orf* XF0156 não apresenta similaridade significativa com proteína já bem caracterizada funcionalmente, mas possui um domínio de peptidase C1 entre os amino ácidos 37 e 267. Proteínas com este domínio são chamadas de "papaína-like" e representam cisteíno proteases secretadas que necessitam de clivagem do pro-peptídio N-terminal para ativação[66].

A maior similaridade com uma proteína de função conhecida é de e-9 com uma cisteíno protease de Giardia. Apesar de baixa, esta similaridade é homogênea por toda sua sequência. Na Giardia, esta protease é localizada em vesículas que são liberadas para que o parasita possa sair do cisto e infectar o hospedeiro[67].

Na bactéria *Porphyromonas gingivalis*, uma cisteíno protease secretada (a gengipaína) é considerada um dos principais fatores de patogenicidade desta bactéria

[63] Stock, A. M., Robinson, V. L., Goudreau, P. N. Two-component signal transduction. *Annu. Rev. Biochem.* 2000, 69, 183-215.

[64] Hobbs, M., Collie, E. S., Free, P. D., Livingston, S. P. et al. PilS and PilR, a two-component transcriptional regulatory system controlling expression of type 4 fimbriae in Pseudomonas aeruginosa *Mol. Microbiol.* 1993, 7, 669-682.

[65] Dekkers, L.C., Bloemendaal, C.J., de Weger, L.A., Wijffelman, C.A. et al. A two-component system plays an important role in the root-colonizing ability of Pseudomonas fluorescens strain WCS365. *Mol. Plant Microbe Interact.* 1998, 11, 45-56.

[66] Rawlings, N. D., Barrett, A. J. Families of cysteine peptidases. *Meth. Enzymol.* 1994, 244, 461-486.

formadora de biofilme responsável por doenças periodontais. Apesar de não haver homologia de sequência com a XF0156, o modelo de ação da gengipaína pode oferecer pistas iniciais sobre a função de uma cisteíno protease na formação de biofilme. Já foi demonstrado que a gingipaína modula o processo de adesão da bactéria com o epitélio de gengiva e a formação de biofilme[68,69,70,71]. Ela atua aumentando a capacidade de ligação da fímbria bacteriana com o epitélio[72], o que parece ser feito pela degradação de proteínas estruturais do epitélio.

Ainda, a gengipaína está envolvida no processo de "fimbriação", atuando intracelularmente na clivagem da porção N-terminal de proteínas precursoras das subunidades de fímbrias[73]. A cis-protease XF0156 da *X. fastidiosa* poderia estar atuando neste processamento, permitindo a transição da fímbria formada pelo gene XF2542 para a fímbria formada pelo gene XF2539.

Lipase (XF2151)

Anotada como proteína hipotética no banco de dados genômico da *X. fastidiosa*, esta proteína não apresenta similaridade significativa com proteínas de função conhecida, mas tem um domínio de lipase secretada do tipo LIP entre os aminoácidos 127 e 399.

É possível que esta proteína altamente expressa no crescimento em biofilme também esteja envolvida na modulação da adesão e formação de biofilme. Lipases deste tipo hidrolisam ligações éster na interface de triacilglicerídeos insolúveis e a fase aquosa onde a enzima está dissolvida[74]. No caso de *C. albicans*, lipases da família LIP são expressas e secretadas durante o ciclo de infecção e acredita-se que contribuem para a sobrevivência e virulência do fungo em tecidos humanos[75]. Já foi proposto que

---

[67] Ward, W., Alvarado, L., Rawlings, N. D., Engel, J. C. et al. A primitive enzyme for a primitive cell: the protease required for excystation of Giardia. *Cell* 1997, 89, 437-444.

[68] Ellen, R. P., Lepine, G., Nghiem, P. M. In vitro models that support adhesion specificity in biofilms of oral bacteria. *Adv. Dent. Res.* 1997, 11, 33-42.

[69] Kuramitsu, H., Tokuda, M., Yoneda, M., Duncan, M. et al. Multiple colonization defects in a cysteine protease mutant of Porphyromonas gingivalis. *J. Periodontal Res.* 1997, 32, 140-142.

[70] Baba, A., Abe, N., Kadowaki, T., Nakanishi, H. Arg-gingipain is responsible for the degradation of cell adhesion molecules of human gingival fibroblasts and their death induced by Porphyromonas gingivalis. *Biol. Chem.* 2001, 382, 817-824.

[71] Yoshioka, M., Wang, P. L., Ohura, K. Effect of proteases from Porphyromonas gingivalis on adhesion molecules of human periodontal ligament fibroblast cells. *Nippon Yakurigaku Zasshi* 1997, 110, 347-355.

[72] Kontani, M., Kimura, S., Nakagawa, I., Hamada, S. Adherence of Porphyromonas gingivalis to matrix proteins via a fimbrial cryptic receptor exposed by its own arginine-specific protease. *Mol. Microbiol.* 1997, 24, 1179-1187.

[73] Nakayama, K., Yoshimura, F., Kadowaki, T., Yamamoto, K. Involvement of arginine-specific cysteine proteinase (Arg-gingipain) in fimbriation of Porphyromonas gingivalis. *J, Bacteriol.* 1996, 178, 2818-2824.

[74] Anthonsen, H. W., Baptista, A., Drablos, F., Martel, P. et al. Lipases and esterases: a review of their sequences, structure and evolution. *Biotechnol. Annu. Rev.* 1995, 1, 315-371.

[75] Hube, B., Stehr, F., Bossenz, M. Mazur, A. et al. Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members. *Arch. Microbiol.* 2000, 174, 362-374.

lipases secretadas são importantes fatores de virulência em várias bactéiras, com funções na adesão e nutrição[76].

[76] Gribbon, E. M., Cunliffe, W. J., Holland, K. T. Interaction of Propionibacterium acnes with skin lipids in vitro. *J. Gen. Microbiol.* 1993, 139, 1745-1751.

2.3. O Banco de Dados do Proteoma da *Xylella fastidiosa*

Para rápida divulgação e organização dos dados gerados, um banco de dados 2DE foi desenvolvido e pode ser acessado pela internet http://proteome.ibi.unicamp.br/ (Figura 14).

É possível selecionar um dos vários mapas 2DE depositados no banco de dados (Figura 14B). O mapa selecionado é então mostrado com os *spots* já identificados marcados por uma cruz vermelha. Informação resumida de cada um destes *spots* aparece quando o cursor é direcionado por sobre a área do *spot* (Figura 15A). Clicando no *spot* remete a informações mais detalhadas como sequência completa, anotação funcional, método de identificação, massa molecular aparente e prevista, o pI estimado e previsto, dentre outras (Figura 15B). Pelo "link" "search protein" é possível acessar uma lista contendo todas as proteínas identificadas. Os protocolos experimentais usados para preparação de amostra e 2DE também foram disponibilizados no "site" do banco de dados.

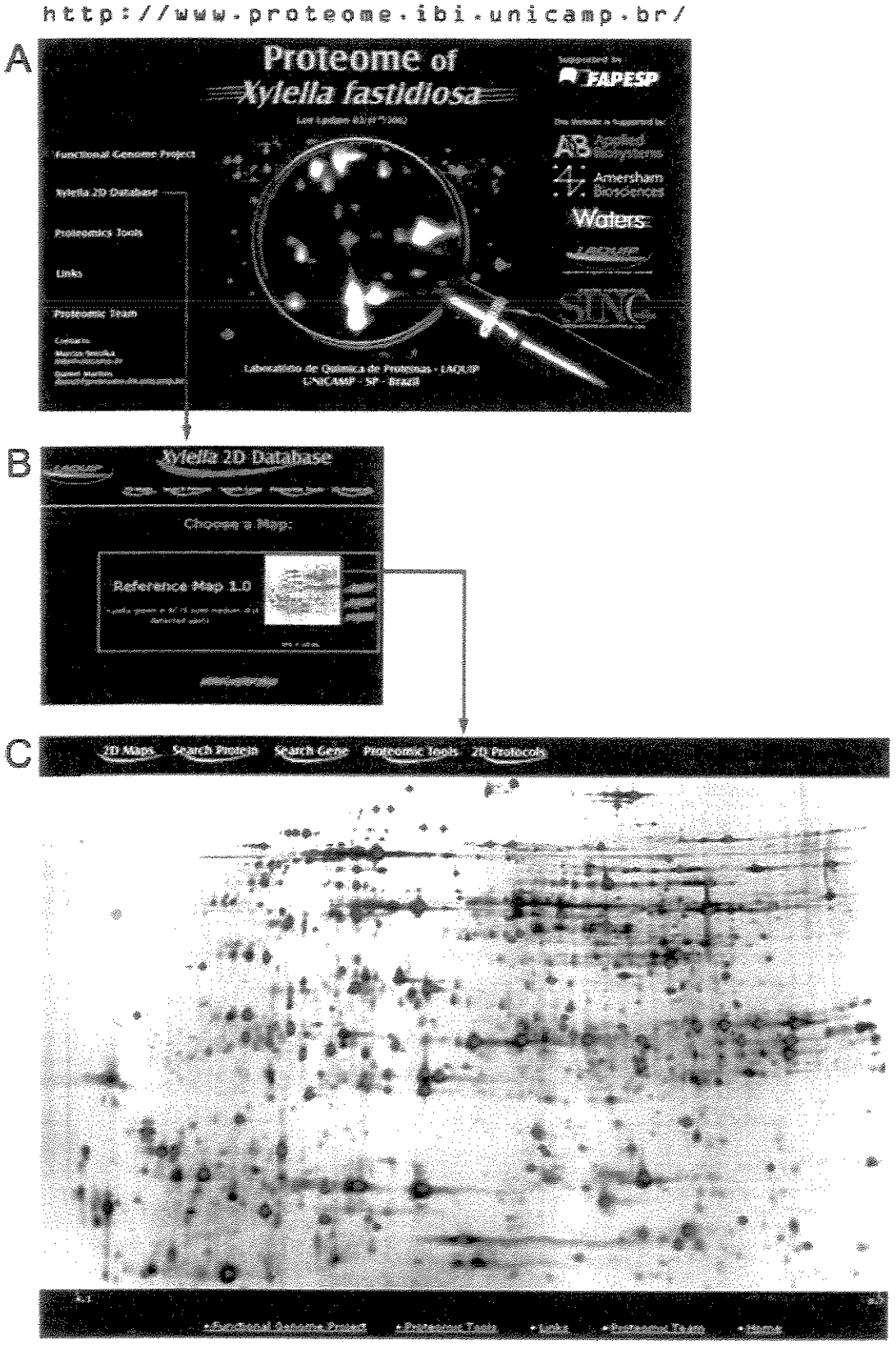


FIGURA 14: Acesso e funcionamento do banco de dados do proteoma da *X. fastidiosa*. (A) "Homepage" do projeto. (B) Página de seleção do mapa 2DE desejado. (C) Apresentação do mapa 2DE com os *spot* identificados marcados em vermelho.

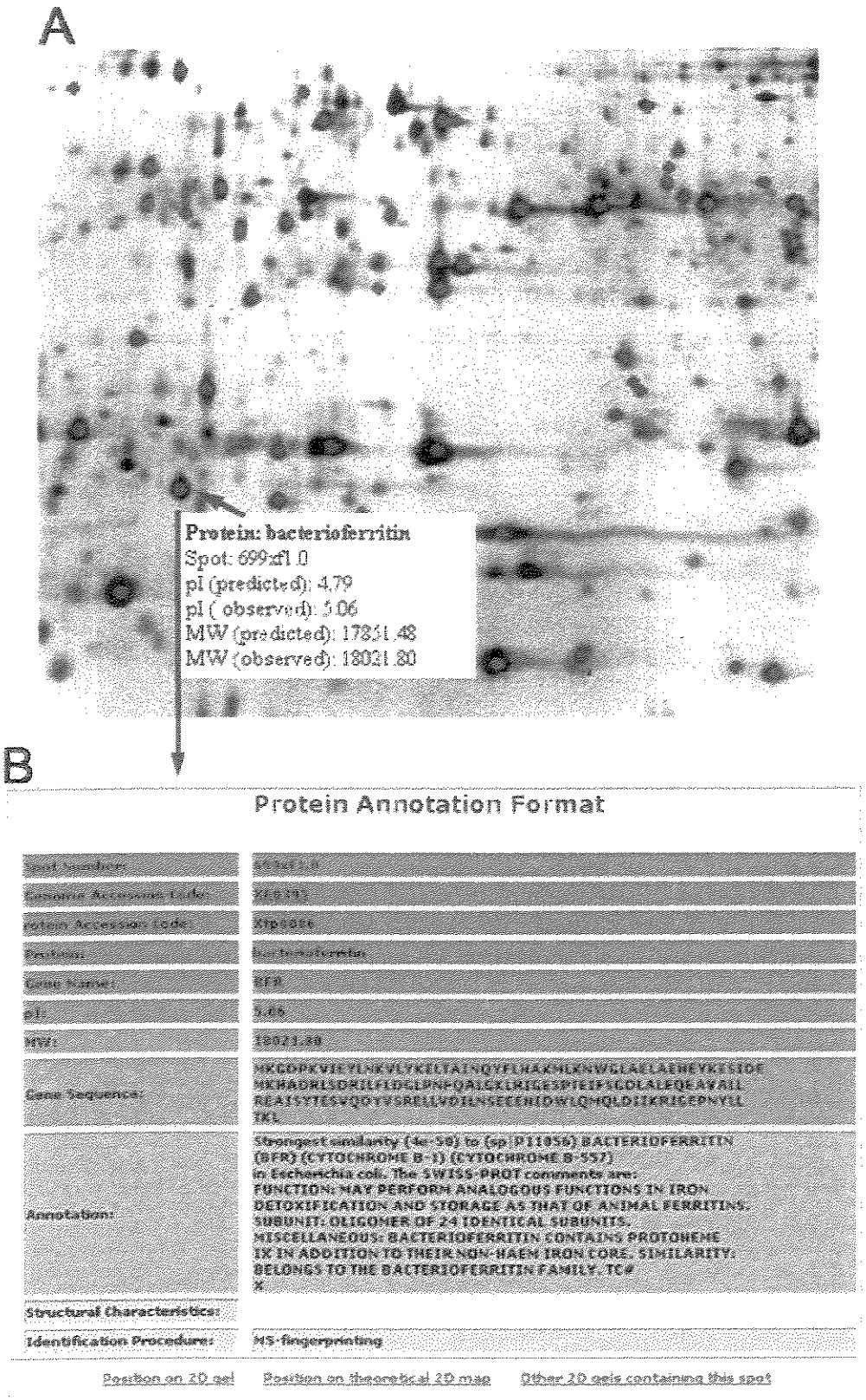


FIGURA 15: Forma de apresentação das informações de cada *spot* identificado. (A) Informação resumida de cada *spot* aparece quando o cursor é direcionado por sobre a área do *spot*. (B) Clicando no *spot* remete a informações mais detalhadas.

A página "Proteomic tools" disponibiliza várias ferramentas computacionais de análise de sequências. A ferramenta "pI and MW Prediction Tool" permite o cálculo dos valores previstos de pI e massa molecular de várias proteínas simultaneamente. Estes dados foram fundamentais para a etapa de confirmação das identificações por MS, onde os valores previstos de pI e massa molecular de uma proteína eram comprados com os respectivos valores experimentalmente determinados pela posição de migração no mapa 2DE. A ferramenta permite ainda que seja gerado uma 2DE "virtual" baseando-se nos valores previstos de pI e massa molecular de todos os genes do genoma da *X. fastidiosa* (Figura 16). Toda parte de programação computacional da ferramenta "pI and MW Prediction Tool" bem como do banco de dados foi desenvolvida pelo aluno de iniciação científica Itaraju Brum.

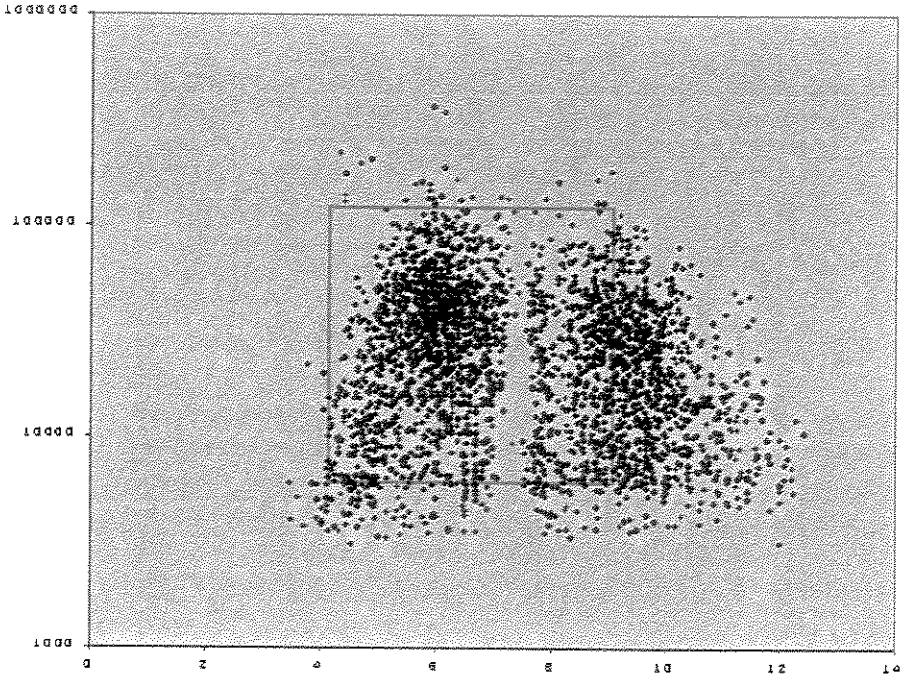
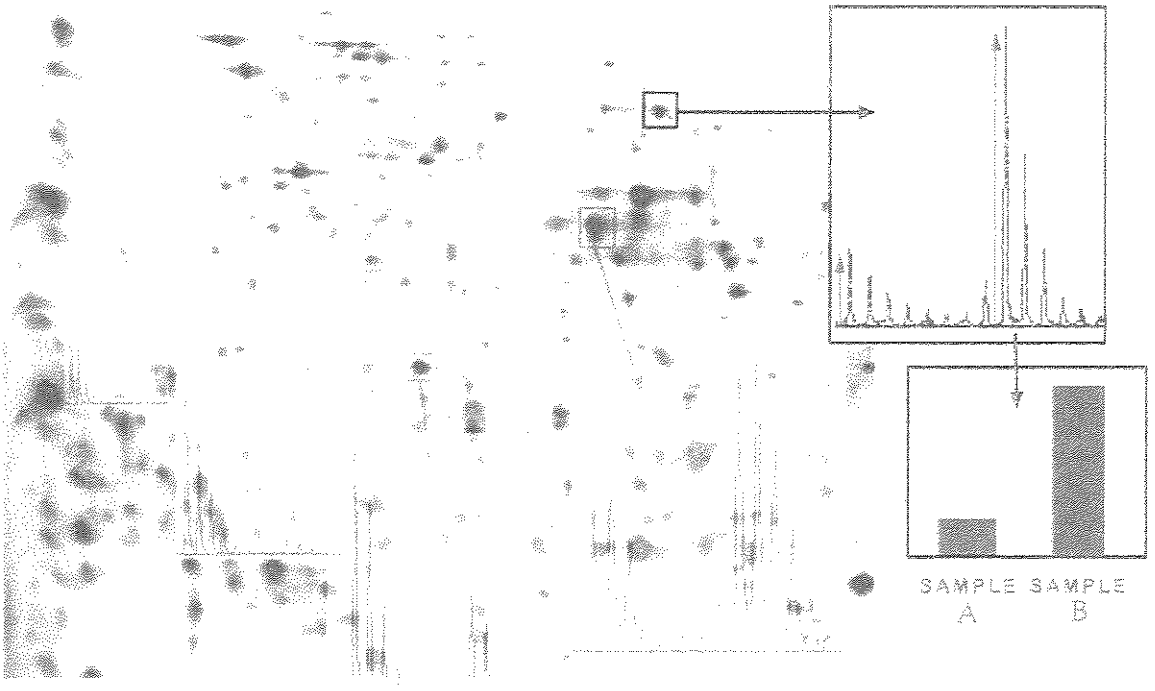


FIGURA 16: Mapa 2DE teórico da *Xylella fastidiosa*. Os valores de pI e MM previstos a partir da seqüência de todos os genes do genoma forma plotados num gráfico cartesiano (pI no eixo *x* e MM no eixo *y*), simulando um mapa 2DE real. O retângulo cinza indica a região na qual a 2DE real consegue resolver e detectar proteínas. É possivel notar a limitação da 2DE em representar as proteínas de baixa e alta MM, e principalmente as de alto pI.

Capítulo 2:

Desenvolvimento de Novo Método para Análise Quantitativa de Proteomas por Espectrometria de Massas



1. **Artigo Publicado:** Análise Quantitativa de Proteínas Utilizando Eletroforese de Duas Dimensões, Rotulação com Etiquetas Isotopicamente Marcadas e Espectrometria de Massas

Artigo publicado na revista Molecular and Cellular Proteomics

Smolka M, Zhou H, Aebersold R. **Quantitative Protein Profiling Using Two-dimensional Gel Electrophoresis, Isotope-coded Affinity Tag Labeling, and Mass Spectrometry.** *Mol Cell Proteomics* 2002 Jan;1(1):19-29.

Quantitative Protein Profiling Using Two-dimensional Gel Electrophoresis, Isotope-coded Affinity Tag Labeling, and Mass Spectrometry*

Marcus Smolka‡, Huilin Zhou§, and Ruedi Aebersold§¶

Quantitative protein profiling is an essential part of proteomics and requires new technologies that accurately, reproducibly, and comprehensively identify and quantify the proteins contained in biological samples. We describe a new strategy for quantitative protein profiling that is based on the separation of proteins labeled with isotope-coded affinity tag reagents by two-dimensional gel electrophoresis and their identification and quantification by mass spectrometry. The method is based on the observation that proteins labeled with isotopically different isotope-coded affinity tag reagents precisely co-migrate during two-dimensional gel electrophoresis and that therefore two or more isotopically encoded samples can be separated concurrently in the same gel. By analyzing changes in the proteome of yeast (*Saccharomyces cerevisiae*) induced by a metabolic shift we show that this simple method accurately quantifies changes in protein abundance even in cases in which multiple proteins migrate to the same gel coordinates. The method is particularly useful for the quantitative analysis and structural characterization of differentially processed or post-translationally modified forms of a protein and is therefore expected to find wide application in proteomics research. ***Molecular & Cellular Proteomics 1:19–29, 2002.***

Proteomics attempts to study the structure, function, and control of biological systems and processes by the systematic and quantitative analysis of the many properties of proteins. These include the sequence (identity), abundance, activity, and structure of the proteins expressed in a cell, as well as the modifications, interactions, and translocations each protein might experience. There is currently no single experiment or platform that permits the analysis of all these properties on a proteome-wide scale. Therefore, technologies and platforms have been developed that specifically study a subset of these properties.

The systematic identification of the proteins contained in a complex sample and the determination of the relative abundance of each protein (quantitative protein profiling) if two or more samples are being compared are central objectives of proteomics. For this purpose two methods have been developed and are routinely being used. The first is the combination of the well established techniques of two-dimensional gel electrophoresis (2DE)[1] and mass spectrometry (MS) (reviewed in Ref. 1). The second is a recently developed procedure based on isotope-coded affinity tag (ICAT™) reagent protein labeling, liquid chromatography, and tandem mass spectrometry (2, 3).

At present the 2DE/MS method is most commonly used. Proteins separated by 2DE and detected by staining are identified, one-by-one, by mass spectrometric analysis of peptide fragments derived from each protein. The method is supported by robust and automated instruments that perform specific steps in the process, such as gel imaging, spot picking, protein digestion, peptide mass spectrometry, and sequence data base searching (1, 4–6). In this method protein quantification is achieved by image analysis of the spot patterns generated by staining the proteins separated in the gel, implicitly assuming that the staining intensity of a protein spot accurately indicates the amount of protein contained in the spot. Unfortunately, the practical implementation of this simple concept has been difficult, mainly because of limitations of the available staining and gel electrophoresis methods. These include limited dynamic range, sensitivity, and reproducibility, respectively, of the popular Coomassie Blue and silver staining procedures (7, 8), limited pattern reproducibility of 2DE gels, and a recurrence of several proteins to migrate to the same gel coordinates. The development of sensitive fluorescence staining methods with a linear signal response over a wide dynamic range (9, 10) and the introduction of a two-color fluorescent

From the ‡Departamento de Bioquímica, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Sao Paulo 13083-970, Brazil and the §Institute for Systems Biology, Seattle, Washington 98105

Received, September 5, 2001, and in revised form, October 9, 2001
Published, MCP Papers in Press, October 11, 2001, DOI 10.1074/mcp.M100013-MCP200

[1] The abbreviations used are: 2DE, two-dimensional (isoelectric focusing/SDS-PAGE) gel electrophoresis; BSA, bovine serum albumin; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; DTT, dithiothreitol; ICAT™, isotope-coded affinity tag reagents; IEF, isoelectric focusing; LACB, bovine β-lactoglobulin; LCA, bovine α-lactalbumin; MALDI, matrix-assisted laser desorption-ionization; MS, mass spectrometry; MW, molecular weight; OVAL, chicken ovalbumin; TOF, time-of-flight.

labeling system allowing the concurrent electrophoresis of two differentially labeled protein samples in the same 2DE gel and thus the determination of the ratio of abundance by spectral analysis of each protein spot (11, 12) have eliminated or at least alleviated some of the challenges related to 2DE gel imaging.

As 2DE gel-based proteomic studies depend on the analysis of the separated proteins by MS, it has been suggested to use stable-isotope dilution to achieve accurate protein quantification in the mass spectrometer, obviating the need for gel imaging (13, 20). In these experiments proteins were metabolically labeled with stable isotopes (^{15}N or ^{13}C), and two samples with distinguishable isotope signatures were combined and concurrently separated by 2DE. Proteins in specific spots were then digested, and the resulting peptides were extracted and analyzed by MS. The ratio of signal intensities of the isotopically normal and heavy forms of a particular peptide was then used to calculate the ratio of abundance of the respective protein. Although quantification was shown to be very accurate, this method is limited to cells that can be cultured in isotopically defined culture conditions.

In this paper we show that post-isolation isotopic protein labeling using the ICAT™ reagents is also compatible with 2DE and accurate protein quantification by MS. We show that proteins labeled with the isotopically heavy and normal forms of the reagent, respectively, precisely co-migrate in 2DE gels and that the ratio of protein abundance can be accurately determined from the relative mass spectrometric signal intensities of the heavy and normal forms of labeled, cysteine-containing peptides. Both identification and accurate quantification are therefore achieved in a simple single-step analysis. This method also reliably identifies and quantifies each protein in spots containing multiple different polypeptides and identifies and quantifies each isoform or post-translationally modified form of a protein that migrates to different positions. As the method is based on post-isolation chemical labeling of proteins, it is compatible with most if not all protein samples that can subsequently be separated and analyzed by 2DE.

EXPERIMENTAL PROCEDURES

Labeling Standard Proteins—The standard proteins used were bovine serum albumin (BSA), chicken ovalbumin (OVAL), bovine β-lactoglobulin (LACB), bovine α-lactalbumin (LCA), or superoxide dismutase (Sigma). Labeling was performed as described previously (14), with slight modifications. 1 mg of each protein was dissolved in 1 ml of Solution A (0.05% (w/v) SDS, 6 M urea, 5 mM EDTA, and 50 mM Tris buffer, pH 8.3). Specified amounts of each protein solution were mixed together (as described for each case under "Results"), and aliquots of 100 μg of total protein mixture (100 μl) were labeled. Disulfide bonds were reduced with 5 mM tributyl phosphine (Sigma) for 10 min and then 60 μg of the isotopically normal or heavy form of ICAT™ reagent (Applied Biosystems, Framingham, MA) was added. This amount of ICAT™ reagent (105 nmol) represented an ~1 mM concentration and a 3 to 1 molar ratio over total protein sulfhydryls assuming an average molecular protein mass of 30 kDa and 8 cysteine residues per protein. The reaction mixtures were incubated for 2 h at room temperature in an Eppendorf tube in the dark. The reactions were stopped by adding DTT to a final concentration of 10 mM. The light and heavy ICAT™ reagent-labeled protein mixtures were combined prior to 2DE separation. Incomplete labeling of proteins was carried out under the conditions described above except that the sample was diluted 10-fold, and urea was omitted.

Preparation and Labeling Yeast Protein Extracts—Yeast cells (*Saccharomyces cerevisiae* strain BWG1-7A) were grown until mid-log phase in yeast/peptone/dextrose medium containing either 2% (w/v) glucose or 2% (w/v) galactose as carbon source and harvested by centrifugation. Glucose- and galactose-grown cells were lysed separately in lysis buffer containing 1% (w/v) SDS, 50 mM Tris, pH 8.0, 5 mM EDTA, 1 mM phenylmethylsulfonyl fluoride after lyticase treatment to prepare spheroblasts according to standard procedures (15). An aliquot of each of the resulting protein extracts was used to determine the protein concentration using a commercial protein assay kit (Bio-Rad). To the remainder of the samples DTT was added to a final concentration of 100 mM, and the solution was boiled for 5 min. 300 μg of total protein from each extract was precipitated with a cold ethanol:acetone:acetic acid (50:50:0.1 (v/v/v)) solution, and the precipitate was washed once with 70% ethanol in water. The protein pellet was re-suspended in 50 μl of Solution A, and the disulfide bonds were reduced by the addition of tributyl phosphine to a final concentration of 5 mM and incubation for 10 min at room temperature. 100 μg of ICAT™ reagent (~3 mM final concentration) was added to the reaction solution. After 2 h the reaction was stopped by adding DTT to a final concentration of 10 mM.

For the purpose of comparing 2DE gel patterns of each sample (glucose- or galactose-grown) the labeled proteins were again precipitated as described above, re-dissolved, and separated by 2DE. For concurrent electrophoresis of both samples on the same gel and subsequent quantification by MS, equal amounts of the two labeled extracts were combined, precipitated as described above, re-dissolved, and separated by 2DE.

Two-dimensional Gel Electrophoresis—Approximately 20 μg of standard protein mixture or 300 μg of yeast protein extract were solubilized in 340 μl of sample buffer containing 8 M urea, 4% (w/v) CHAPS, 2% (v/v) carrier ampholytes, pH 3–10, 70 mM DTT, and 0.001% (w/v) bromphenol blue. Except for CHAPS, which was from Sigma, all reagents were from Amersham Biosciences, Inc. Samples were applied to immobilized pH gradient strips with a non-linear separation range of pH 3–10 (catalog number 17-1235-01; Amersham Biosciences, Inc.). After a 10-h rehydration, isoelectric focusing was carried out, at 20 °C, for 1 h at 500 V, for an additional hour at 1000 V, and then for 10 h at 8000 V in an IPGphor apparatus (Amersham Biosciences, Inc.) maintaining a limiting current of 50 μA per strip. First dimension strips were subjected to the standard reduction and alkylation steps prior to second-dimension electrophoresis. Strips were soaked for 10 min in a solution containing 50 mM Tris-HCl, pH 6.8, 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, and 2% (w/v) DTT and for an additional 10 min in the same solvent containing 2.5% (w/v) iodoacetamide instead of DTT. Second-dimension electrophoresis (SDS-PAGE) was performed on an SE 600 system connected to a MultiTemp II refrigerating system (both from Amersham Biosciences, Inc.). After laying the strip on the top of a 12.5% polyacrylamide gel and sealing it with agarose, electrophoresis was carried out for 1 h at 90 V, at which time a constant amperage of 30 mA per gel was applied until the dye front reached the lower end of the gel. Proteins were detected by a silver nitrate staining protocol adapted from Ref. 16.

Protein Identification and Quantitation by Matrix-assisted Laser Desorption Ionization (MALDI) Quadrupole Time-of-Flight (TOF) MS—Peptides were generated and extracted from the gel-separated proteins following established in-gel tryptic digestion protocols (17). The peptides were analyzed on a QSTAR™ Pulsar quadrupole TOF tandem mass spectrometer (Applied Biosystems/MDS Sciex, Toronto, Canada), which was equipped with a MALDI ion source. Prior to

application to the sample plate the samples were purified using C18 ZipTips (Millipore) and mixed at a 1:1 ratio with a solution of 16% (w/v) 2,5-dihydrobenzoic acid in acetonitrile:water (25:75 v/v).

Identification of the standard proteins was performed by submitting its known sequences to the Peptidemass tool (18) and manually comparing the predicted peptide masses of each standard protein to the measured masses in the mass spectrum. For identification of the proteins from yeast extracts, the measured mass of the tryptic peptides were searched against *S. cerevisiae* entries from Swiss-Prot using the MS-Fit program (University of California, San Francisco; prospector.ucsf.edu/) (19). In this program, it is already possible to consider ICAT™ labeling of cysteines for calculating the predicted mass of the peptides for each protein. The observed pI and MW of the identified spots in the 2DE map were compared with the theoretical pI and MW to confirm identifications.

For quantifying the relative abundance of a protein, the peak height of the monoisotopic peak of the light ICAT™-labeled peptide was divided by the peak height of the monoisotopic peak of the heavy labeled form of the peptide. When more than one cysteine-containing peptide was detected for a particular protein, the abundance ratio was calculated from the peak pair with the highest signal intensity. In cases in which the shape of the isotope distributions of the light and heavy ICAT™-labeled peptides differed (*i.e.* asymmetric shape of the clusters) we suspected the presence of contaminating peaks and did not use such peak pairs for quantification. Instead, the peak pair with the next highest signal intensity was used.

RESULTS

Principle of the Method—The method for quantitative protein profiling that we describe in this paper is schematically illustrated in Fig. 1. It is based on the separation of proteins labeled with ICAT™ reagents by 2DE and their identification and quantification by mass spectrometry. The proteins contained in two separate samples are first labeled with the isotopically normal (non-deuterated) or heavy forms (deuterated), respectively, of ICAT™ reagents according to established protocols (2, 14).[2] The two samples are then combined and concurrently separated by 2DE in the same gel, and the separated proteins are detected by an MS compatible staining protocol. The proteins migrating to specific spots are enzymatically digested in the gel matrix, and the resulting peptides are extracted and analyzed by mass spectrometry. Protein identification is achieved either by peptide mass mapping in a single-stage mass spectrometer (21, 22) or by collision-induced dissociation of selected peptides and sequence data base searching (23). The ratio of abundance of the protein(s) in the spot analyzed is determined by the ratio of signal intensities for the isotopically normal and heavy forms of a specific, tagged peptide.

Migration of ICAT™ Reagent-labeled Proteins in 2DE—To determine whether and how labeling with the ICAT™ reagent affected protein migration during 2DE a mixture of equal amounts of LACB and LCA was labeled with the isotopically normal form of ICAT™ reagent and separated by 2DE. The 2DE profile of the labeled proteins was compared with the profile obtained from an unlabeled but otherwise identical sample separated under the same conditions. Fig. 2, *A* and *B* show the profiles obtained from the unlabeled and labeled samples, respectively. It is apparent that ICAT™ reagent labeling decreased the electrophoretic mobility of both proteins. The observed mobility decrease was consistent with the mass of 442 Da added by each ICAT™ reagent molecule. For LCA, which contains eight cysteines, the calculated molecular mass increased from 14186 to 17722 Da, and the apparent molecular mass increased from 14 to ~18 kDa. For LACB, which contains five cysteines, the calculated molecular mass increased from 18281 to 20491 Da, and the apparent molecular mass increased from 18 to ~21 kDa. In contrast, no shift in pI was apparent as a consequence of ICAT™ labeling. The observation that the pI of the acidic proteins LCA and LACB is unaffected is consistent with the fact that the ICAT™ reagent molecule is uncharged and, by reacting with sulfhydryl groups that are uncharged at acidic pH values, does not alter the solution charge state of the protein. For basic proteins a slight change in pI is expected to occur after ICAT™ reagent labeling, because the negative charge of sulfhydryl groups at pH values above the pK of sulfhydryls will be neutralized by the labeling reaction.

We have demonstrated previously that incomplete labeling of proteins with ICAT™ reagents results in a "fuzzy" banding pattern in SDS-PAGE and reported optimized labeling protocols for complete labeling of all the cysteines in a reduced protein (14). Fig. 2*C* shows the 2DE profile of partially labeled LCA and LACB if a non-optimized labeling protocol was used. It is evident that incomplete protein labeling is easily recognized by the characteristic ladder spot pattern and that the well defined spots achieved by complete labeling are essential for the success of the method.

Co-migration by 2DE of Proteins Labeled with Isotopically Normal and Heavy ICAT™ Reagents—To determine whether proteins labeled with the different isotopic forms of the ICAT™ reagent migrated to identical coordinates in 2DE gels, samples of LCA were labeled with the isotopically normal or heavy form of the reagent. The proteins were combined at a d0:d8 (light ICAT™: heavy ICAT™) ratio of 11:1, and the mixed sample was separated by 2DE. The resulting spot, detected by silver staining, was divided into four quadrants, and the protein in each quadrant was in-gel digested and analyzed mass spectrometrically. Fig. 3*A* shows the gel spot, and the four quadrants and Fig. 3*B* detail the d0:d8 ratios used for quantification. The expected d0:d8 ratio was 11.00. The ratio calculated by averaging the observed values over all the quadrants was 11.72. The highest deviation from this observed mean value detected in quadrant number 3 was 11.30, corresponding to a 3.6% difference.

We chose LCA for this experiment, because the prevalence of cysteine residues in this protein would be expected to

[2] T. Nadler, K. Parker, B. Wagenfeld, R. Lotti, B. Purkayastha, S. Daniels, W. Stanick, S. Pillai, J. N. Marchese, and G. Vella (2001) Optimization of a protocol for preparing protein samples with an isotope-coded affinity tag (ICAT™) reagent, submitted for publication.

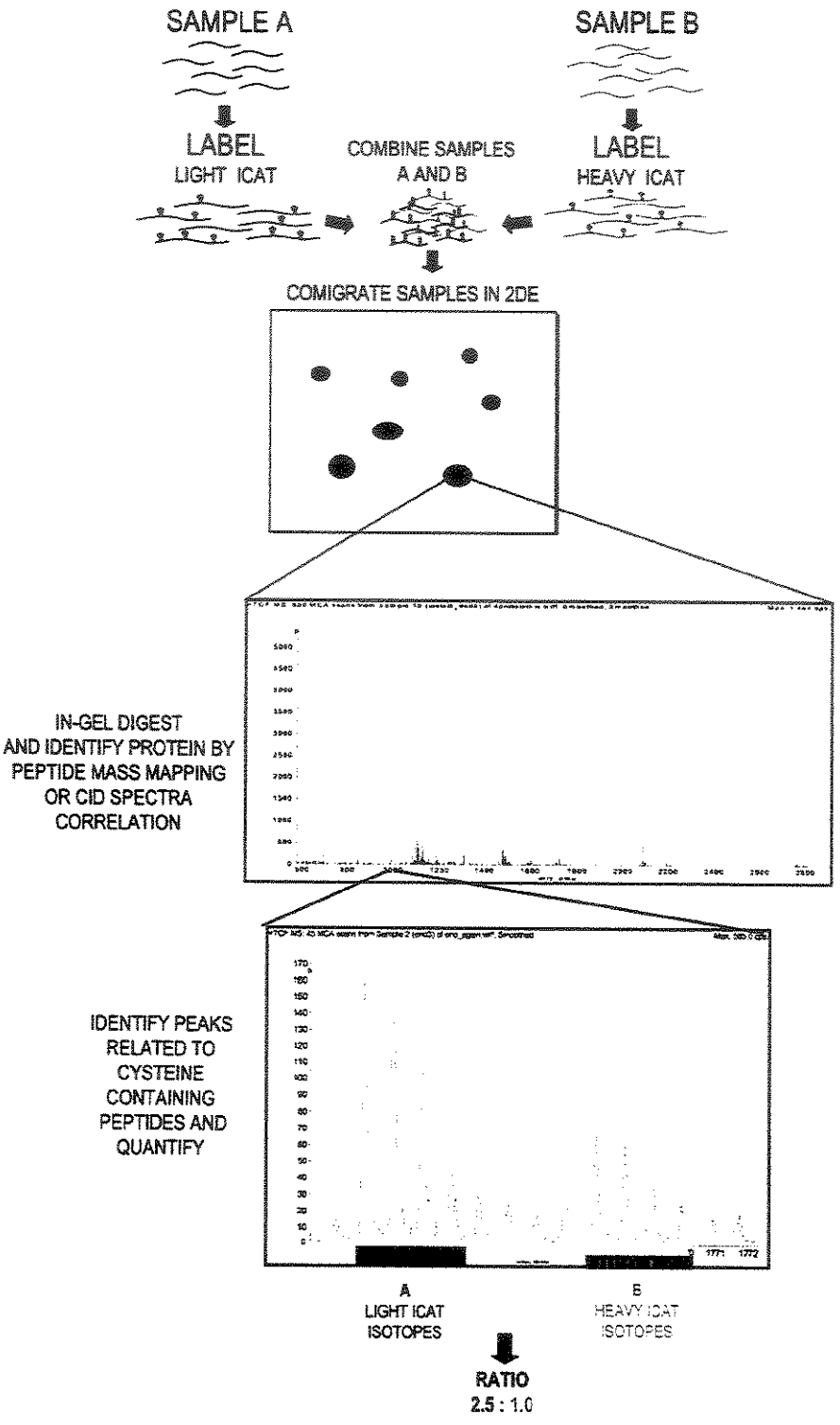


FIG. 1. **Schematic illustration of the proposed method for quantitative protein profiling.** The proteins contained in two separate samples are first labeled with the isotopically normal (non-deuterated) or heavy forms (deuterated), respectively, of ICAT™ reagents. The two samples are then combined and concurrently separated by 2DE in the same gel, and the separated proteins are detected by MS-compatible stains. The proteins migrating to specific spots are enzymatically digested in the gel matrix, and the resulting peptides are extracted and analyzed by mass spectrometry. Protein identification is achieved either by peptide mass mapping in a single-stage mass spectrometer or by collision-induced dissociation of selected peptides and sequence data base searching. The ratio of abundance of the protein(s) in the spot analyzed is determined by the ratio of signal intensities for the isotopically normal and heavy forms of a specific, tagged peptide.

exaggerate isotope effects on electrophoretic migration. At eight cysteine residues the heavy ICAT™ reagent-labeled protein contained 64 deuterium atoms. Moreover, the pI (4.8) and molecular mass (14.1 kDa) of LCA position this protein in an area of high resolution in the gel used. These results indicate that proteins labeled with differentially isotopically

labeled forms of the ICAT™ reagents precisely co-migrate by 2DE.

Accuracy of Quantification—To evaluate the accuracy of quantification of the method we prepared and analyzed two protein mixtures containing the same five proteins at different quantities. The composition of the samples is indicated in Table I. One sample was labeled with the isotopically normal and the other with the heavy form of the ICAT™ reagent, and the samples were combined and separated by 2DE. Protein spots were detected by silver staining and in-gel digested with trypsin, and the resulting peptides were identified and quantified by mass spectrometry.

The concise shape of the protein spots in the 2D electropherogram and the absence of vertical "streaking" suggested that quantitative labeling had been achieved (Fig. 4*A*). The multiple spots (with same MW and different pI) observed for LACB, OVAL, and BSA were also observed in the 2DE map of the unlabeled protein mixture (data not shown) and are therefore unrelated to the labeling with ICAT™ reagent. The identity of each protein spot was confirmed by peptide mass fingerprinting. For each protein, the predicted masses of the labeled, cysteine-containing peptides were also calculated and used to identify the corresponding peaks in the mass spectra. Fig. 4*B* shows the peptide mass spectrum of superoxide dismutase and expansions of the peak areas containing the signals for cysteine-containing peptides. Each peptide appeared as two isotope envelopes with an 8.04-Da mass difference, corresponding to the non-deuterated and deuter-

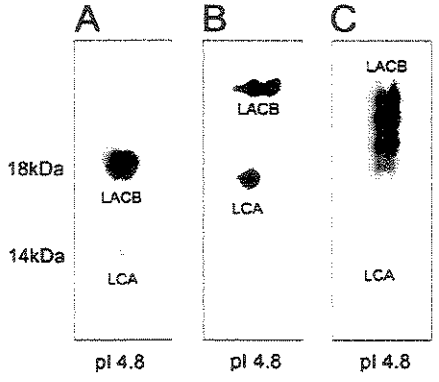


FIG. 2. **Effect of ICAT™ reagent labeling on 2DE protein migration.** 2DE profiles of unlabeled (*A*), completely labeled (*B*), and partially labeled (*C*) β-lactoglobulin (*LACB*) and α-lactalbumin (*LCA*) are shown.

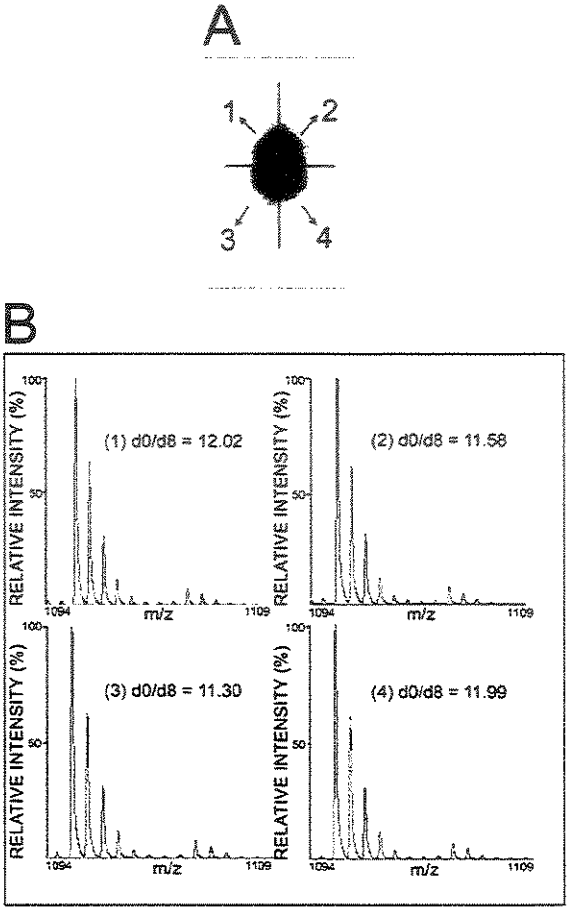


FIG. 3. **Co-migration of isotopically normal and heavy ICAT™ reagent-labeled LCA by 2DE.** Aliquots of LCA were labeled with either the isotopically normal or heavy ICAT™ reagent, respectively, combined at an 11:1 ratio and separated by 2DE. The protein was detected by silver staining, and four quadrants of the spot were dissected and separately trypsinized, and the extracted peptides were analyzed by MALDI-TOF mass spectrometry. The orientation and numbering of the sections are indicated in *A*. The observed isotopic ratios for each quadrant are indicated in *B*.

TABLE I

Composition of each of the standard protein mixtures used for the experiment shown in Fig. 5

After separately labeling with the indicated form of ICAT™ reagent the two mixtures were combined and concurrently separated by 2DE. The expected quantitative ratios, the quantitative ratios that were experimentally determined, and the calculated error are also displayed.

Protein	Mixture 1 (d0-labeled)	Mixture 2 (d8-labeled)	Expected ratio	Observed ratio	Error
	μg		*d0:d8*		%
Superoxide dismutase (SOD)	34	17	2.00:1.00	2.02:1.00	1
LCA	44	4	11.00:1.00	11.72:1.00	7.92
LACB	9	20	1.00:2.22	1.00:2.12	4.4
OVAL	7	35	1.00:5.00	1.00:5.5	10
BSA	6	24	1.00:4.00	1.00:3.57	12

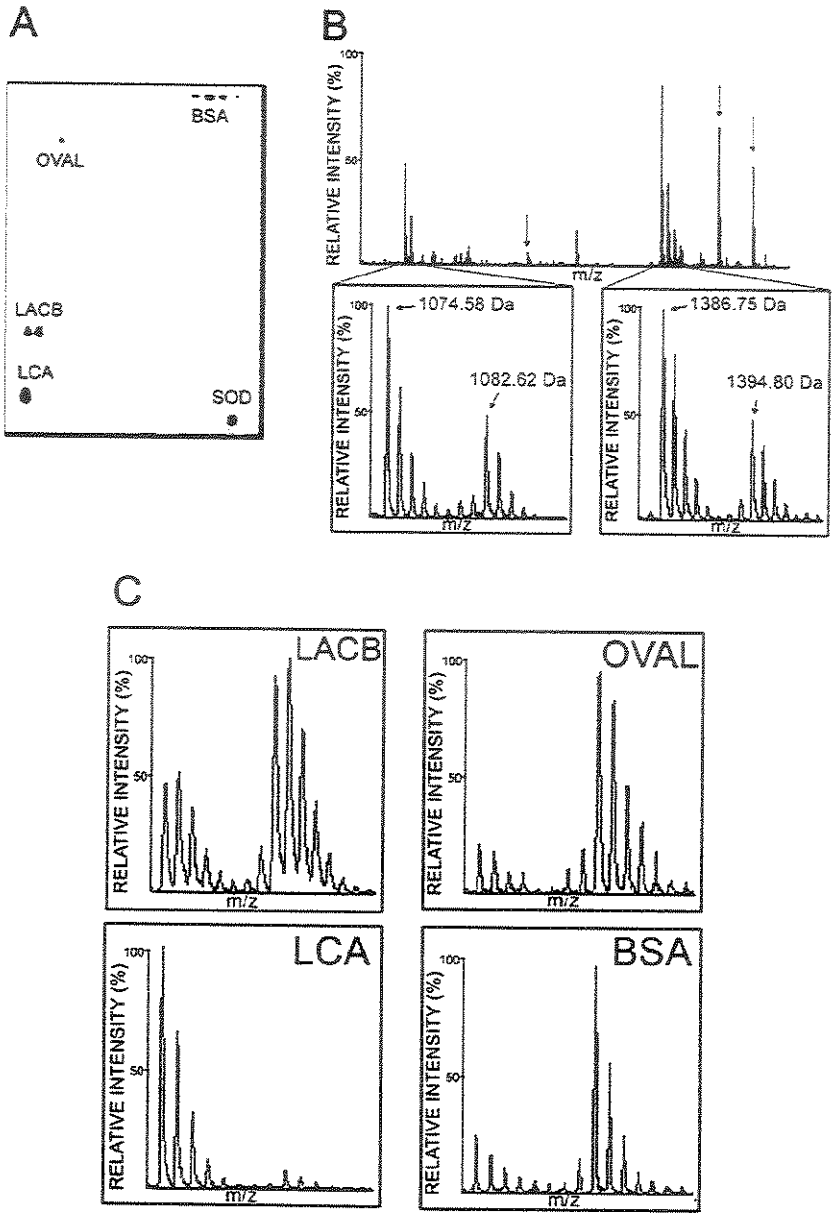


FIG. 4. **Measurement of the relative abundances of proteins labeled with isotopically normal and heavy ICAT™ reagents.** Two protein mixtures containing known amounts of the five proteins BSA, OVAL, LACB, LCA, and superoxide (*SOD*) dismutase were prepared (for composition of the samples refer to Table I). The mixtures were labeled with the isotopically normal or heavy form of the ICAT™ reagent, combined, and separated by 2DE. Spots were detected by silver staining, and the tryptic peptides from individual spots were analyzed by MALDI-TOF MS. *A*, 2DE gel electropherogram of the combined sample mixture. *B*, identification and quantification of superoxide dismutase. The *arrows* in the peptide mass spectrum indicate the non-cysteine-containing peptides used for protein identification by peptide mass fingerprinting. The masses of the two ICAT™ reagent-labeled (cysteine-containing) peptides were calculated and used to identify the corresponding peaks in the MS scan. The respective spectra are expanded in the *zoomed squares*. *C*, peak pairs of cysteine-containing peptides used for quantification of the other four proteins present in the sample. The signal intensities of the monoisotopic peaks were used for the calculation of the observed abundance ratios indicated in Table I.

ated forms of the peptides. Fig. 4C shows the peaks of cysteine-containing peptides used for quantification of the other four proteins in the mixture. The signal intensities of the monoisotopic peaks were used to calculate the observed abundance ratios. The measured ratios are indicated and related to the expected respective ratios in Table I. The discrepancy between the expected and measured values was less than 12% for all the proteins in the mixture. These results demonstrate the possibility of using ICAT™ reagent labeling for accurate quantification of proteins separated by 2DE.

Quantification of Changes in Protein Abundance in Yeast Induced by Metabolic Shift—Protein expression in yeast is known to be highly affected by the type of carbon source used. To assess the ability of the method of this paper to identify proteins in complex mixtures and to quantify changes in their abundance, we investigated changes in the protein profile of *S. cerevisiae* induced by a shift from glucose to galactose in the carbon source. A protein extract from cells grown in glucose-containing medium was labeled with the isotopically normal ICAT™ reagent, and an extract from yeast grown in galactose-containing medium was labeled with the heavy ICAT™ reagent. The samples were either run individually (see Fig. 5, *A* for glucose and *B* for galactose) or combined (Fig. 6) in 2DE gels. No indication for partial protein

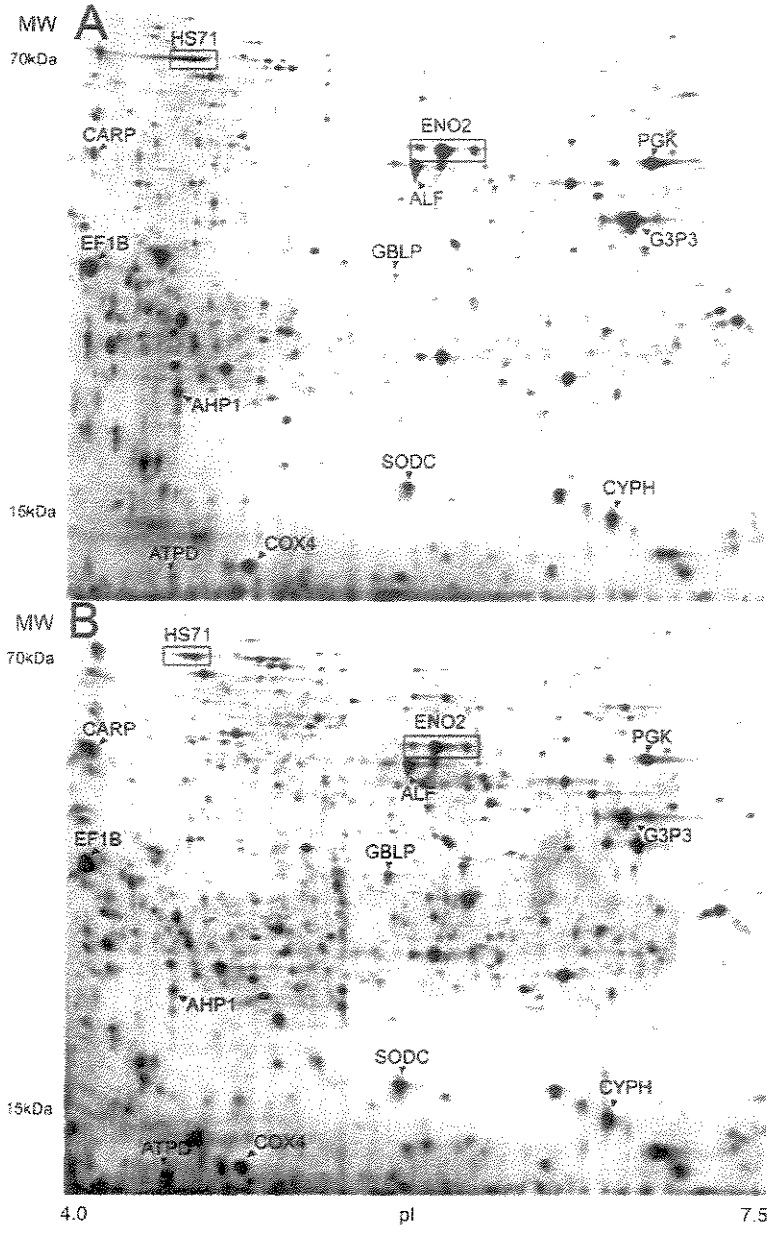


FIG. 5. **2DE profile of yeast protein extract labeled with ICAT™ reagent.** Total protein extracts from yeast grown in glucose (*A*) or galactose (*B*) were labeled with isotopically normal or heavy ICAT™ reagent, respectively, and separated by 2DE. Proteins were detected by silver staining. The spots containing proteins that were subsequently identified and quantified are labeled with the names of the identified proteins. The protein names are defined in Table II.

labeling was detectable. Moreover, the quality of the 2DE gels was comparable with that of 2DE gels of unlabeled yeast extracts (data not shown), indicating that ICAT™ labeling of proteins does not interfere with the separation of complex protein mixtures.

Thirteen spots from the 2DE map presented in Fig. 6 were in-gel digested and submitted to identification and quantification by mass spectrometry. The results are summarized in Table II. The abundance ratios indicated were calculated from the intensities of the most intense ICAT™ reagent-labeled peptides. The expression level of metabolic enzymes such as adolase, phosphoglycerate kinase, and glyceraldehyde-3-phosphate dehydrogenase 3 was observed to be lower in the galactose-grown yeast, a result consistent with results obtained from microarray experiments (24). In contrast, mitochondrial proteins such as COX4 and ATPD were found to be de-repressed in the absence of glucose, also consistent with the mRNA data.

Quantitative Analysis of Post-translational Modifications—Post-translational modifications frequently change protein pI and/or apparent MW of proteins and are therefore easily detected by 2DE. To assess the potential of the method to detect quantitative changes in the spot pattern of differentially modified proteins and therefore to quantify induced changes

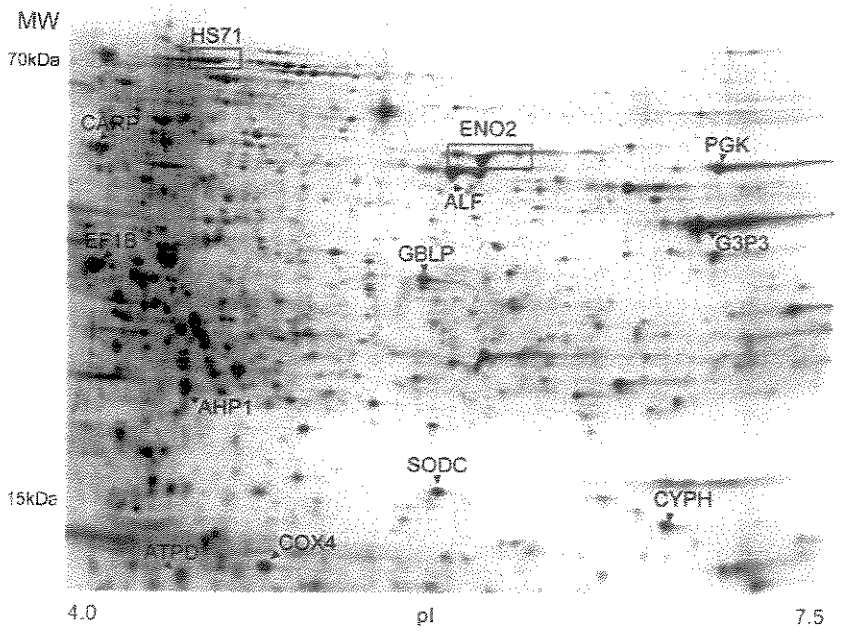


FIG. 6. **Co-migration of the protein samples shown in Fig. 5.** Extracts from yeast cells grown in glucose (light ICAT™ reagent-labeled) or galactose (heavy ICAT™ reagent-labeled) were combined and concurrently separated by 2DE. Proteins were detected by silver staining. The spots containing proteins that were subsequently identified and quantified are labeled with the names of the identified proteins. The protein names are defined in Table II.

TABLE II

Quantification of the relative abundance levels of proteins from yeast cells grown in glucose or galactose, respectively

Protein (Swiss-Prot gene name)		Description	Observed ratio (glucose d0:galactose d8[a])
AHP1	yeast	Peroxisomal alkyl hydroperoxide reductase	1.88:1.00
ENO2	yeast	Enolase 2	2.39:1.00
ALF	yeast	Aldolase	2.31:1.00
PGK	yeast	Phosphoglycerate kinase	2.2:1.00
G3P3	yeast	Glyceraldehyde-3-phosphate dehydrogenase 3	1.53:1.00
EF1B	yeast	Elongation factor 1-β	1.03:1.00
CYPH	yeast	Peptidylprolyl *cis-trans* isomerase	1.16:1.00
CARP	yeast	Saccharopepsin (precursor)	1.00:1.17
ATPD	yeast	ATP synthase Δ chain, mitochondrial (precursor)	1.00:6.58
SODC	yeast	Superoxide dismutase (Cu-Zn)	1.00:1.33
GBLP	yeast	Guanine nucleotide-binding protein β-subunit-like protein	1.21:1.00
COX4	yeast	Cytochrome c oxidase polypeptide IV, mitochondrial (precursor)	1.00:3.10
HS71	yeast	Heat shock protein SSA1	1.00:1.05

[a] d0, light ICAT™-labeled; d8, heavy ICAT™-labeled.

in protein modification we mass spectrometrically analyzed individual spots in the spot pattern of ENO2 and HSP70, proteins known to be phosphorylated (25, 26). Results are shown in Fig. 7. The three ENO2 spots showed a consistent glucose:galactose ratio of 2.4:1. In contrast, the ratio for the two HSP70 spots analyzed varied significantly (2.7:1 for the more acidic spot; 1:1 for the more basic spot) indicating that the modification causing the acidic shift is increased if the cells are grown in glucose. Although the observed spot pattern was consistent with protein phosphorylation, no phosphorylated peptide was detected directly by MS. For the positive identification of phosphorylated peptides auxiliary methods such as metabolic [^{32}P] or [^{33}P] radiolabeling or immunodetection (27, 28) might need to be employed. These results indicate that the method can detect quantitative changes in protein modification or processing, provided that the different protein forms can be separated by 2DE.

DISCUSSION

We describe a new strategy for quantitative protein profiling that is based on the separation of proteins labeled with ICAT™ reagents by 2DE and their identification and quantification by mass spectrometry. The method is based on the observation that proteins labeled with isotopically different ICAT™ reagents precisely co-migrate during 2DE and that therefore two or more isotopically encoded samples can be separated concurrently in the same gel. The ICAT™-labeled proteins can be separated by 2DE at a resolution comparable

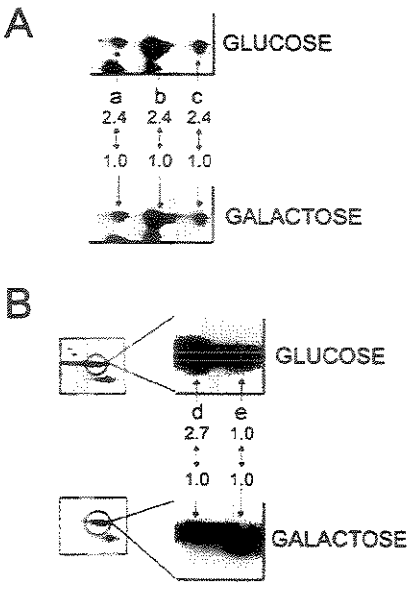


FIG. 7. **Quantification of separated charge isoforms.** Protein spots detected in the gel shown differing in pI but containing the same protein species were separately analyzed by MALDI-TOF MS, and each charge isoform was separately quantified. *A*, three charge isoforms of ENO2 show a consistent ratio of abundance. *B*, two charge isoforms (*d* and *e*) of HS71 show significantly different ratios of abundance.

with unlabeled protein samples. We show that the method has a quantitation accuracy of better than 20% and that different isoforms or the differentially post-translationally modified forms of a protein can be identified individually and quantified. Furthermore, different proteins migrating to the same 2DE coordinates can also be identified and quantified individually.

For separating ICAT™ reagent-labeled proteins at high resolution it was critical that the labeling reaction proceeded to completion. Incomplete labeling was evident from vertically streaking spots. Based on the absence of detectable vertical streaks, the labeling procedures we and others have optimized previously (14)[2] achieved near quantitative labeling even for very complex samples. ICAT™ reagent concentrations in excess of 1 mM (which results in a molar excess of reagent over to the sulfhydryl groups), the addition of urea, and the maintenance of protein solubility by the addition of SDS were shown to significantly improve the labeling reaction.

Langen *et al.* (20) and Oda *et al.* (13) have used metabolic stable isotope labeling of proteins prior to 2DE, also for the purpose of quantifying the proteins by mass spectrometry. Metabolic stable isotope labeling, although simple and effective, is only applicable to protein samples from cells that can be cultured in isotopically enriched or depleted media and is therefore essentially restricted to microbial species. Furthermore, the number of heavy isotopes added by metabolic labeling is sequence-specific. The mass difference between pairs of signals representing the heavy and normal form of a specific peptide is therefore variable, complicating protein identification and quantification. In contrast, the ICAT™ reagent labeling method is based on post-isolation chemical tagging of proteins and is therefore compatible with essentially any protein sample that contains cysteine. Furthermore, with the exception of the cysteine-containing peptides, peptide masses remain unchanged. Protein identification by peptide mass fingerprinting, the detection of post-translationally or otherwise modified peptides in the sample, and the detection of the peptides tagged with the ICAT™ labeling reagent is therefore a straightforward operation. Because cysteine is a relatively rare amino acid, the positive identification of cysteine-tagged peptides in a peptide mixture provides a strong constraint for sequence data base searching (29). Cysteine-containing peptides are easily identified by the presence of the different isotopic forms. Data base search programs such as MS-fit (19) already include ICAT™ reagent labeling as a possible cysteine modification. The use of post-isolation isotopic protein labeling prior to 2DE/MS analysis therefore presents a simpler and more general alternative to metabolic stable isotope labeling.

Recently, an elegant method was introduced (11) and evaluated (12) in which proteins are covalently labeled with fluorescent dyes prior to separation by 2DE. Two structurally similar dyes with different spectral properties were used to label the proteins in two samples with a specific color, the combined samples were concurrently separated by 2DE, and the ratio of abundance for spot was determined by spectral analysis of the emitted fluorescent light. This method also effectively eliminates the problem of electrophoretic variation between gels, experiments demonstrating accurate quantification have validated the technique (11, 12), and fluorescent detection is potentially extremely sensitive. A drawback of the current implementation is that the epsilon amino groups of the abundant amino acid lysine are labeled with relatively hydrophobic dyes, thus reducing the solubility of extensively labeled proteins. To avoid protein precipitation in the gel, minimal labeling is performed. This incomplete labeling procedure limits the sensitivity of the method and makes equal quantitative labeling between two samples more difficult. Another problem is that the labeled and unlabeled proteins of the same species migrate to different coordinates in the 2DE gel, generating a rather complex 2DE profile.

In contrast, the method described in this paper is based on complete protein labeling. Consequently, all of the protein molecules of a particular species are concentrated in the same spot. Furthermore, the present method offers the possibility of individually quantifying multiple proteins migrating to the same spot, a feature that is impossible with the fluorescent or other imaging techniques. We believe that the present method should represent an attractive alternative to the fluorescent labeling method, particularly in cases in which mass spectrometric analysis of the separated proteins is attempted.

In addition to being a constraining factor in sequence data

base searching cysteine-specific protein labeling has other beneficial features. As cysteine is relatively rare no extreme shifts in SDS-PAGE migration are expected after labeling. We also did not notice a decrease in protein solubility after ICAT™ reagent labeling, probably because the amino acids with charged side chains (lysine, arginine, glutamic acid, and aspartic acid) are not affected by the labeling reaction and because the reagent used is not very hydrophobic. It was also of practical importance that all the cysteines were reacted with ICAT™ reagent molecules, thus eliminating the problem of cysteine re-oxidation during the IEF run and contributing to maintain the proteins in a desirable denatured state. In most 2DE protocols, cysteine reduction is achieved by including high DTT concentrations to the IEF separation solution. But during the IEF run, DTT may acquire a negative charge and migrate to the anode, depleting reducing power on the basic regions. Moreover, DTT accumulated in the acidic region may migrate in the second-dimension run and cause artifacts in the 2DE profile. Labeling proteins with ICAT™ prior to IEF separation should avoid the necessity of using DTT in the IEF run. The subsequent reduction/alkylation step prior to the second-dimension run can also be eliminated.

The described technique has several limitations. The ICAT™ reagent method using current reagents only labels cysteine-containing proteins. For the S. *cerevisiae* proteome ~92% of the proteins contain at least one cysteine residue. Although any other protein would also be identified by the method, accurate quantification relies on the detection of at least one cysteine-containing peptide in the mass spectrum. Reagents with different specificities that are under development will eliminate this limitation. In the described method protein quantification is achieved in the mass spectrometer on individual spots. Therefore, for determining differentially expressed spots it is either necessary to run three gels per sample, one for each of the samples separately and one for the combined sample, or to analyze all the spots present in the gel containing the combined samples. With the increasing availability of automated, high throughput robotic sample-processing systems (4–6) and the development of methods for parallel isolation and digestion of 2DE-separated proteins (30, 31) the latter option is increasingly becoming feasible.

Concluding Remarks—Significant advances in protein identification have been accomplished over the last ten years through the use of mass spectrometry, and large-scale or even proteome-wide protein identification is now a common practice. The descriptive information obtained from such protein cataloguing projects can be significantly enhanced if the quantity and changes thereof can be determined precisely for each protein in a sample. The strategy described in this paper provides a simple and effective tool by which accurate protein quantitation can be achieved in proteome-wide experiments. The method also allows the determination of the absolute 2DE and mass spectrometry, two techniques established in most proteomics laboratories. The method also lends itself to determine the absolute amounts of specific proteins, if isotopically labeled, calibrated internal standards are being used and for the selective, quantitative analysis of selected protein spots, a task that might be very useful for medical diagnostic purposes. We therefore anticipate that this method will find wide application in the field of quantitative proteomics.

Acknowledgments—The support and critical input from Drs. Ken Parker and George Vella (Applied Biosystems, Framingham, MA) are gratefully acknowledged.

* This work was supported in part by NCI, National Institutes of Health Grant 1R33CA84698 (to R. A.) and by a fellowship from Fundação de Amparo à Pesquisa do Estado de São Paulo (to M. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Inst. for Systems Biology, 4225 Roosevelt Way, Suite 200, Seattle, WA 98105. Tel.: 206-732-1204; Fax: 206-732-1254; E-mail: raebersold@systemsbiology.org.

REFERENCES

1. Aebersold, R., and Goodlett, D. (2001) Mass spectrometry in proteomics. *Chem. Rev.* **101,** 269–295
2. Gygi, S. P., Rist, B., Gerber, S. A., Frantisek, T., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17,** 994–999
3. Griffin, T. J., Gygi, S. P., Rist, B., Aebersold, R., Loboda, A., Jilkine, A., Ens, W., and Standing, K. G. (2001) Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer. *Anal. Chem.* **73,** 978–986
4. Traini, M., Gooley, A. A., Ou, K., Wilkins, M. R., Tonella, L., Sanchez, J.-C., Hochstrasser, D. F., and Williams, K. L. (1998) Toward an automated approach for protein identification in proteome projects. *Electrophoresis* **19,** 1941–1949
5. Houthaeve, T., Gausepohl, H., Ashman, K., Nilson, T., and Mann, M. (1997) Automated protein preparation techniques using a digest robot. *J. Protein Chem.* **16,** 343–348
6. Wilkins, M. R. (2001) in *Proteomics: from Protein Sequence to Function* (Pennington, S. R., and Dunn, M. J., eds) pp. 171–192, Bios, Oxford
7. Patton, W. F. (2000) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* **21,** 1123–1144
8. Patton, W. F. (2001) in *Proteomics: from Protein Sequence to Function* (Pennington, S. R., and Dunn, M. J., eds) pp. 31–59, Bios, Oxford
9. Lopez, M. F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M., and Patton, W. F. (2000) A comparison of silver stain and SYPRO ruby protein gel stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* **21,** 3673–3683
10. Steinberg, T. H., Jones, L. J., Haugland, R. P., and Singer, V. L. (1996) SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Anal. Biochem.* **239,** 223–237
11. Unlu, M., Morgan, M. E., and Minden, J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18,** 2071–2077
12. Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **1,** 377–396
13. Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **12,** 6591–6596
14. Smolka, M. B., Zhou, H., Purkayastha, B., and Aebersold, R. (2001) Opti-

mization of the isotope-coded affinity tag labeling procedure for quantitative proteome analysis. *Anal. Biochem.* **297,** 25–31
15. Dunn, B., and Wobbe, C. R. (1987) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) Vol. 1, pp. 13.1–13.9, John Wiley & Sons, Inc., New York
16. Blum, H., Beier, H., and Gross, H. J. (1987) *Electrophoresis* **8,** 93–99
17. Schevechenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68,** 850–858
18. Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J.-C., Hochstrasser, D. F., and Appel, R. D. (1997) Detailed peptide characterization using PEPTIDEMASS-a World-Wide Web accessible tool. *Electrophoresis* **18,** 403–408
19. Clauser K. R., Baker P. R., and Burlingame A. L. (1999) Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **71,** 2871–2882
20. Langen, H., Fountoulakis, M., Evers, S., Wipf, B., and Berndt, P. (1998) in *Proceedings from the 3rd Siena 2-D Electrophoresis Meeting, Siena, Italy, August 31–September 3, 1998* (Hochstrasser, D. F., Bini, L., and Pallini, V., eds) p. 58, ICES, Siena, Italy
21. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 5011–5015
22. Pappin, D. J. (1997) Peptide mass fingerprinting using MALDI-TOF mass spectrometry. *Methods Mol. Biol.* **64,** 165–173
23. Yates, J. R., III, Eng, J. K., McCormack, A. L., and Schieltz D. (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67,** 1426–1436
24. Roth, F. P., Hughes, J. D., Estep, P. W., and Church, G. M. (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.* **16,** 939–945
25. Zhou, H., Watts, J. D., and Aebersold, R. (2001) A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19,** 375–378
26. Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., and Garrels, J. I. (1999) A sampling of the yeast proteome. *Mol. Cell. Biol.* **19,** 7357–7368
27. Larsen, M. R., Sorensen, G. L., Fey, S. J., Larsen, P. M., and Roepstorff, P. (2001) Phospho-proteomics: evaluation of the use of enzymatic dephosphorylation and differential mass spectrometric peptide mass mapping for site-specific phosphorylation assignment in proteins separated by gel electrophoresis. *Proteomics* **1,** 223–238
28. Kaufmann, H., Bailey, J. E., and Fussenegger, M. (2001) Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Proteomics* **1,** 194–199
29. Goodlett, D. R., Bruce, J. E., Anderson, G. A., Rist, B., Pasa-Tolic, L., Fiehn, O., Smith, R. D., and Aebersold, R. (2000) Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal Chem.* **72,** 1112–1118
30. Binz, P. A., Muller, M., Walther, D., Bienvenut, W. V., Gras, R., Hoogland, C., Bouchet, G., Gasteiger, E., Fabbretti, R., Gay, S., Palagi, P., Wilkins, M. R., Rouge, V., Tonella, L., Paesano, S., Rossellat, G., Karmime, A., Bairoch, A., Sanchez, J. C., Appel, R. D., and Hochstrasser, D. F. (1999) A molecular scanner to automate proteomic research and to display proteome images. *Anal. Chem.* **71,** 4981–4988
31. Bienvenut, W. V., Sanchez, J. C., Karmime, A., Rouge, V., Rose, K., Binz, P. A., and Hochstrasser, D. F. (1999) Toward a clinical molecular scanner for proteome research: parallel protein chemical processing before and during Western blot. *Anal. Chem.* **71,** 4800–4807

2. **Artigo Publicado:** Otimização da Técnica de Rotulação com Etiquetas Isotopicamente Marcadas para Análise Quantitativa de Proteomas

Artigo publicado na revista Analytical Biochemistry

Smolka MB, Zhou H, Purkayastha S, Aebersold R. **Optimization of the Isotope-Coded Affinity Tag Labeling Procedure for Quantitative Proteome Analysis.** *Anal Biochem* 2001 Oct 1;297(1):25-31

Analytical Biochemistry 297, 25–31 (2001)
doi:10.1006/abio.2001.5318, available online at http://www.idealibrary.com on IDEAL®

Optimization of the Isotope-Coded Affinity Tag-Labeling Procedure for Quantitative Proteome Analysis

Marcus B. Smolka,* Huilin Zhou,† Subhasish Purkayastha,‡ and Ruedi Aebersold†,[1]

*Departamento de Bioquímica, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Sao Paulo, Brazil;
†Institute for Systems Biology, 4225 Roosevelt Way, Suite 200, Seattle, Washington 98105; and
‡Applied Biosystems, Inc., 500 Old Connecticut Path, Framingham, Massachusetts 01701

Received April 18, 2001; published online September 12, 2001

The combination of isotope coded affinity tag (ICAT) reagents and tandem mass spectrometry constitutes a new method for quantitative proteomics. It involves the site-specific, covalent labeling of proteins with isotopically normal or heavy ICAT reagents, proteolysis of the combined, labeled protein mixture, followed by the isolation and mass spectrometric analysis of the labeled peptides. The method critically depends on labeling protocols that are specific, quantitative, general, robust, and reproducible. Here we describe the systematic evaluation of important parameters of the labeling protocol and describe optimized labeling conditions. The tested factors include the ICAT reagent concentration, the influence of the protein, SDS, and urea concentrations on the labeling reaction, and the reaction time. We demonstrate that using the optimized conditions specific and quantitative labeling was achieved on standard proteins as well as in complex protein mixtures such as a yeast cell lysate. © 2001 Academic Press

The combination of protein labeling with isotope-coded affinity tag (ICAT)[2] reagents and the mass spectrometric analysis of isolated, tagged peptides derived by proteolysis of the labeled proteins is a new approach to quantitative proteomics (1). The method allows the concurrent identification of the proteins in a complex mixture and the accurate quantification of the difference in abundance for each protein present in two or more protein samples.

The ICAT reagents consist of three functional components (Fig. 1). The first is a thiol reactive group that, if applied to peptides, is selective for the sulfhydryl groups in the side chain of reduced cysteines. The second component is an ethylene glycol linker group that occurs in a deuterated (isotopically heavy) or non-deuterated (isotopically normal) form and provides the basis for accurate quantification. The third component, a biotin group provides an affinity tag for the selective isolation of the tagged peptides from complex peptide mixtures via avidin affinity chromatography. In the course of an experiment the reduced cysteine residues of the proteins in the two samples to be compared are labeled with the isotopically heavy or normal reagent, respectively. The two samples are then combined, digested with trypsin, and the tagged peptides isolated by avidin affinity chromatography and analyzed by MS/MS, thus determining the amino acid sequence and the relative abundance of each protein in the samples being compared. The sequence of a peptide is determined by correlating the CID spectrum obtained in a tandem mass spectrometer with a sequence database, using specific software tools exemplified by the Sequest program (2). The relative abundance of a peptide and therefore the relative abundance of the protein from which the peptide originated is determined by the ratio of signal intensities of the isotopically normal and heavy forms of a peptide.

A critical step in the procedure is the protein labeling reaction in which the sulfhydryl groups of peptides are coupled to the ICAT reagents. Complete and specific labeling is essential for maximizing the recovery of the tagged peptides and for accurate quantification. Addi-

[1] To whom correspondence should be addressed. Fax: (206) 732-1254. E-mail: raebersold@systemsbiology.org.

[2] Abbreviations used: ICAT, isotope coded affinity tags; MS/MS, tandem mass spectrometer/try; CID, collision-induced dissociation; BSA, bovine serum albumin; TBP, tributyl phosphine; DTT, dithiothreitol; YPD, yeast peptone dextrose; 2DE, two-dimensional gel electrophoresis; Chaps, 3-([3-cholamidopropyl]dimethylammonio)-1-propane-sulfonate; IPG, immobilized pH gradient; TOF, time-of-flight; MALDI, matrix assisted laser desorption ionization; DHB, 2,5-dihydrobenzoic acid.

0003-2697/01 $35.00

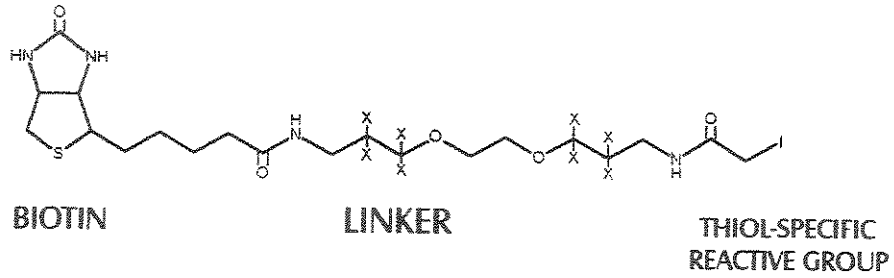


FIG. 1. Structure of the ICAT reagent. It consists of three elements: a thiol-specific reactive group that binds to cysteines residues; a deuterated (heavy reagent, with deuterium in the indicated X position) or nondeuterated (light reagent, with hydrogen in the indicated X position) linker; and a biotin tag used for affinity isolation of the ICAT labeled peptides.

tionally, the reaction should be fast to accelerate the process and to minimize detrimental side reactions such as proteolysis by proteases present in the sample. It also should be compatible with the presence of protein denaturants such as urea and SDS that are required to denature the proteins and to keep them in solution, and, for economical reasons, the amount of ICAT reagent used should be minimized.

In this article we report on the systematic evaluation of the parameters critical for the success of the labeling reaction and present an optimized protocol. Specifically, we investigated the ICAT reagent concentration and its molar excess required to achieve quantitative labeling, the effects of the presence of SDS and urea on the reaction, and the reaction time course.

MATERIALS AND METHODS

Labeling Standard Proteins

The standard proteins used for this study were BSA, chicken ovalbumin, bovine β-lactoglobulin, and bovine α-lactalbumin (Sigma). Unless stated, all reactions were performed as follows: Twenty micrograms of protein was dissolved in 50 μl of solution A (0.05% (w/v) SDS, 5 mM EDTA, and 50 mM Tris buffer, pH 8.3). Disulfide bonds were reduced with 5 mM TBP (Sigma) for 10 min, after which the specified amount of nondeuterated ICAT reagent (Applied Biosystems) was added. The reaction mixture was incubated for 2 h at room temperature in an Eppendorf tube protected from light exposure. The reactions were stopped by adding DTT to a final concentration of 10 mM.

Labeling Yeast Protein Extract

Yeast *Saccharomyces cerevisiae* strain (BWG1-7A) cells were grown till mid-log phase in YPD medium containing 2% (w/v) glucose as carbon source and harvested by centrifugation. Cells were lysed in lysis buffer containing 6 M urea, 1% (w/v) SDS, 100 mM DTT, 50 mM Tris, 5 mM EDTA, 1 mM phenylmethylsulfonyl fluoride after lyticase treatment to prepare spheroblasts according to standard procedures (3). Three hundred micrograms of total protein were precipitated with a cold ethanol:acetone:acetic acid (50:50:0.1 v/v/v/) solution and the precipitate was washed once with 70% ethanol in water. The protein pellet was resuspended in 50 μl of solution A containing 6 M urea. Disulfide bonds were reduced by the addition of TBP to 5 mM final concentration and incubation for 10 min at room temperature. One hundred micrograms of ICAT reagent (3 mM final concentration) was added to the reaction solution. After 2 h the reaction was stopped by adding DTT to a final concentration of 10 mM. Proteins were again precipitated as described above, redissolved, and separated by 2DE as described below.

SDS–PAGE and Two-Dimensional Gel Electrophoresis

To check for labeling efficiency, standard protein samples were subjected to SDS–PAGE (12.5% gels) after reaction and then stained with Coomassie blue to detect the mass shift caused by the addition of the ICAT reagent. Because total yeast protein extracts were too complex to detect the ICAT-induced mobility shift in a single dimension, with these samples labeling efficiency was monitored by 2DE. Approximately 300 μg of labeled or unlabeled yeast protein extracts were solubilized with 340 μl of sample buffer containing 8 M urea, 4% (w/v) Chaps, 2% (v/v) carrier ampholytes, pH 3–10, 70 mM DTT, and 0.001% (w/v) Bromophenol blue (except for Chaps that was from Sigma, all reagents were from Amersham Pharmacia Biotech). Samples were applied to IPG strips with a nonlinear separation range of pH 3–10 (Catalog No. 17-1235-01, Amersham Pharmacia Biotech). After a 10-h rehydration, isoelectric focusing was carried out, at 20°C, for 1 h at 500 V, for 1 h more at 1000 V, and then for 10 h at 8000 V in an IPGphor apparatus (Amersham Pharmacia Biotech) with a limiting current of 50 μA per strip. Both first-dimension strips from labeled and unlabeled samples were subjected to the standard reduction and alkylation steps. Strips were soaked for 10 min in a solution containing 50 mM Tris–HCl (pH 6.8), 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, and 2% (w/v)

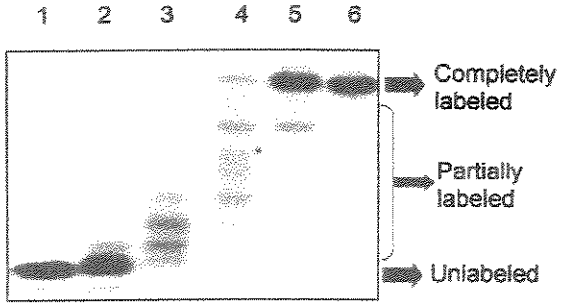


FIG. 2. SDS–PAGE (12.5%) migration profile of unlabeled, partially labeled and completely labeled α-lactalbumin, an 8-cysteine-containing protein. α-Lactalbumin was labeled with (1) 0, (2) 0.05, (3) 0.15, (4) 0.30, (5) 0.60, (6) 1.20 mM of ICAT. As more ICAT reagent is present in the reaction solution, more cysteine residues get labeled and a slower electrophoretic migration is observed. The band from Lane 6 was *in-gel* digested and submitted to MS analysis (Fig. 3).

DTT, and for an additional 10 min in the same solvent containing 2.5% (w/v) iodoacetamide instead of DTT. Second-dimension electrophoresis (SDS–PAGE) was performed on a SE-600 system connected to a Multitemp II refrigerating system (both from Amersham Pharmacia Biotech). After laying the strip on the top of a 12.5% polyacrylamide gel and sealing it with agarose, running was carried out for 1 h at 90 V and then it was applied a constant amperage of 30 mA per gel until front reached the lower end of the gel. Proteins were detected by a standard silver nitrate staining protocol (4).

Mass Spectrometry

Peptides were generated from the gel-separated proteins following established *in gel* tryptic digestion protocols (5). The extracted peptides were analyzed on a QSTAR Pulsar quadrupole TOF tandem mass spectrometer (PE Sciex, Toronto, Canada) that was equipped with a MALDI ion source. Prior to application to the sample plate the peptide samples were purified using C18 ZipTips (Milipore) and mixed at a 1:1 ratio with a solution of 16% (w/v) DHB in acetonitrile:water (25:75, v/v).

RESULTS

SDS–PAGE as an Assay to Monitor Efficiency of Labeling Reaction

We first established whether SDS–PAGE was suitable for the separation and detection of unlabeled and partially and completely labeled forms of a particular protein. As each ICAT reagent molecule bound to a protein adds 442.2 Da to the protein molecular mass, we expected that the differentially labeled forms of that protein could be resolved by SDS–PAGE and that the slowest migrating band would contain the fully labeled protein. Samples of α-lactalbumin which contains eight cysteine residues were subjected to labeling using different reagent concentrations and the products were analyzed by SDS–PAGE. Results are shown in Fig. 2. We observed a banding pattern that is consistent with the separation of unlabeled and partially and completely labeled α-lactalbumin. To confirm that the slowest migrating band indeed contained the completely labeled protein, it was subjected to *in gel* digestion and the recovered peptides were mass spectrometrically analyzed. The masses of the expected tryptic peptides are indicated in Table 1 and the MS results are shown in Fig. 3. All the cysteine-containing peptides were observed only at a mass that corresponded to the labeled form (442.2 Da per cysteine added mass value). These data indicate that SDS–PAGE is a suitable assay for detecting differentially labeled forms of a protein and that complete labeling can be achieved under suitable reaction conditions.

TABLE 1

Theoretical and Measured Masses and Number of Cysteines of the Tryptic Peptides from α-Lactalbumin

Peptide No.	Position on protein	Calculated mass of peptide (Da)	No. of cysteines	Calculated mass of labeled peptide (Da)	Measured mass of peptide (Da)[a]
1	36–77	4654.15	1	5096.35	—
2	82–98	1889.77	2	2774.17	2774.19
3	99–112	1642.73	1	2084.93	2084.93
4	118–127	1200.65	0	—	1200.64
5	134–141	1034.50	1	1476.70	1476.69
6	25–29	653.31	1	1095.51	1095.50
7	128–133	650.32	1	1092.52	1092.53
8	20–24	618.35	0	—	—
9	78–81	549.29	1	991.49	991.50

Note. Masses for labeled peptides were computed by adding 442.2 Da for each cysteine present in the peptide sequence. The number of the peptide refers to the peak number on Fig. 3.

[a] Masses observed on mass spectrum (Fig. 3).

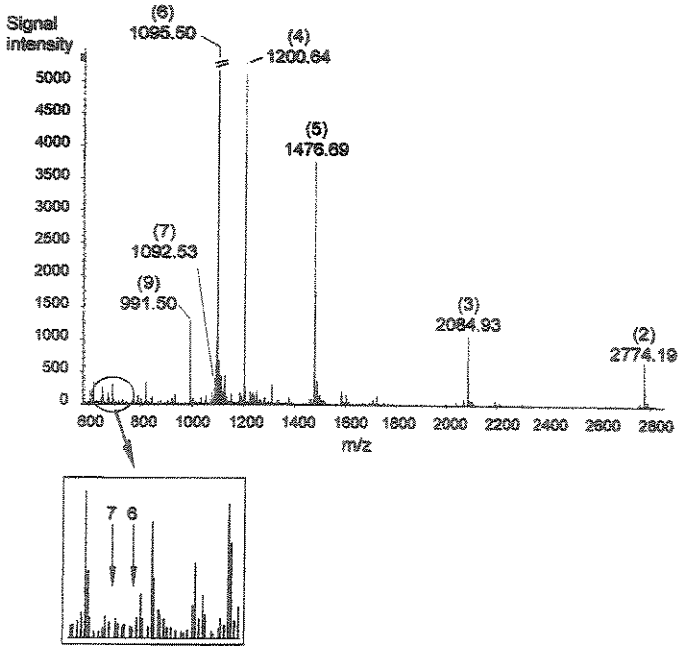


FIG. 3. Mass spectrum of the tryptic peptides from the completely labeled α-lactalbumin band shown in Fig. 2. All the cysteine containing peptides are present with a 442.2-Da per cysteine added mass value, resulting from the ICAT coupling. None of the cysteine-containing peptides was found with the unlabeled mass value. The mass range shown is from 600 to 2800 Da, so peptide 1 (see Table 1) was the only labeled peptide that could not be analyzed. (Inset) An amplification of the region from 620 to 700 Da; the arrows indicate the expected position of peptides 6 and 7 in the unlabeled form. None of the peptides could be found at the mass value relative to the acrylamide modified cysteine version (71.04 Da added mass value per cysteine).

Effects of Reagent Concentration and Reagent to Sulfhydryl Molar Ratio on Labeling Reaction

The four proteins, BSA, ovalbumin, β-lactoglobulin, and α-lactalbumin, were labeled using different reagent concentrations and reagent to sulhydryl molar ratios and the reaction mixtures were separated by SDS–PAGE to monitor the efficiency of the labeling reactions. For each sample the protein in the band with the slowest electrophoretic mobility was digested and the resulting peptides were analyzed by MS to confirm that complete labeling had been achieved (data not shown). The reagent concentration ranged from 0 to 3 mM. The data are summarized in Fig. 4. Essentially, in a 2-h reaction time, complete labeling of the proteins tested was achieved at an ICAT reagent concentration of 3 mM, which represented a 20-fold molar excess of reagent over the total free SH groups present. We also performed labeling reactions at a 50:1 reagent to substrate molar ratio. In this experiment, the same amount of protein was used as above but the sample was diluted to a reaction volume of 1 mL, resulting in an ICAT reagent concentration of a 0.4 mM. Under these conditions incomplete labeling was achieved (data not shown). These results show that complete ICAT labeling is dependent upon a minimal ICAT concentration of approximately 3 mM for the solvent condition used in this experiment and that the ICAT reagent concentration is more important than the molar excess of reagent over substrate.

Effect of SDS and Urea on the Labeling Reaction

To test the effect of the protein denaturing agents on the labeling reaction, the proteins BSA, ovalbumin, β-lactoglobulin, and α-lactalbumin were labeled in the presence of high urea and SDS concentration, respectively, and the reaction mixtures were separated by SDS–PAGE. In these experiments (shown in Fig. 5), except for the presence of the specified denaturant in the labelling solution, all the other conditions (such as protein and reagent concentration, time of reaction, and molar ratio of reagent to sulfhydryl) were strictly the same as described in Fig. 4. The corresponding lanes in all figures are comparable in order to facilitate evaluation of the effect of the denaturant. Figure 5A shows that in the presence of 6 M urea labeling was already complete at a 1.2 mM concentration of ICAT

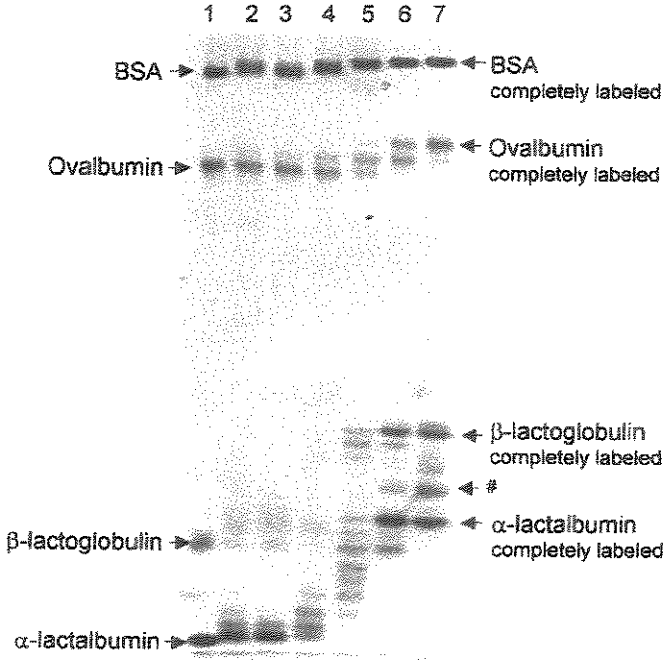


FIG. 4. SDS–PAGE analysis of labeling efficiency in different ICAT concentrations. The four standard proteins were labeled for 2 h with (1) 0, (2) 0.05, (3) 0.15, (4) 0.30, (5) 0.60, (6) 1.20, or (7) 3.00 mM of ICAT in a reaction solution containing 0.05% SDS, 5 mM EDTA, 50 mM Tris buffer, pH 8.3, and 5 mM TBP. *, Bands of completely labeled proteins. #, MS analysis of this band showed that it is a α-lactalbumin form, probably the precursor normally present at small quantities in the commercially available version.

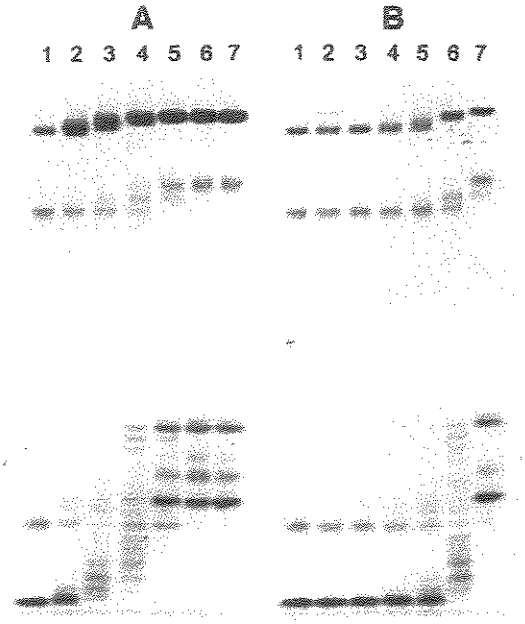


FIG. 5. SDS–PAGE analysis of labeling efficiency in different ICAT concentrations using 6 M urea (A) or 0.5% SDS (B) as an additive to the reaction solution described in Fig. 3. The four standard proteins were labeled for 2 h with (1) 0, (2) 0.05, (3) 0.15, (4) 0.30, (5) 0.60, (6) 1.20, or (7) 3.00 mM of ICAT.

(corresponding to an 8 reagent to sulfhydryl molar ratio) instead of a 3 mM shown in Fig. 4, indicating that urea improved the labeling reaction, presumably by making the sulfhydryl groups more accessible. In the presence of 0.5% SDS complete labeling of BSA, β-lactoglobulin, and α-lactalbumin was achieved at a 3 mM ICAT reagent concentration in otherwise unchanged conditions, whereas ovalbumin was incompletely labeled (Fig. 5B). We furthermore observed that, with 0.5% SDS, any partial labeling of cysteines occurred only at an ICAT concentration higher than 1.2 mM (Fig. 5B, lane 6), where in solution A with 0.05% SDS (no urea), partial labeling already took place with 0.3 mM of ICAT (Fig. 4, lane 4). These results indicate that high concentrations of SDS compromise the labeling reaction, presumably by preventing access of the reagent to the sulfhydryl groups. For these reasons, SDS concentrations higher than 0.5% were not tested. If the experimental design requires high SDS concentrations for initial protein solubilization, the sample might be subsequently diluted to a SDS concentration better suited for ICAT labeling.

Reaction Time Course

In order to evaluate how fast complete labeling was achieved under optimized reaction conditions, we performed time-course experiments using β-Lactoglobulin as the substrate protein. The reactions were carried out in buffer A (see Materials and Methods) added with 6 M urea and 0.05% SDS and at a 1 mM ICAT reagent concentration, which resulted in a 5.5-fold molar excess of reagent over sulfhydryl groups. Results shown in Fig. 6A show that after 10 min most of the cysteinyl residues were labeled and that after 1 h the reaction was essentially complete.

Effect of the Protein Concentration on the Labeling Reaction

As shown above the labeling reactions are more efficient at high ICAT concentrations, suggesting that the volume of the reaction solution should be minimized. For a constant amount of protein, a reduction in the reaction volume implies that the protein concentration will be higher. To detect any negative effects potentially caused by high protein concentrations we determined the efficiency of the labeling reaction at increasing substrate concentration. β-Lactoglobulin was dissolved at concentrations ranging from 0.1 to 3.6 μg/μL in 0.05% SDS, 6 M urea, 5 mM EDTA, and 50 mM Tris buffer, pH 8.3, reduced and reacted with ICAT reagent at a 1 mM for 2 h. The samples were analyzed by SDS–PAGE and data are shown in Fig. 6B. Even at a protein concentration of 3.6 μg/μL, which

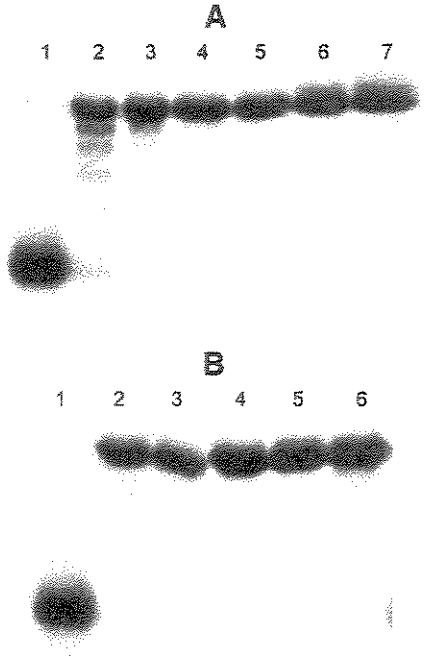


FIG. 6. SDS–PAGE analysis of labeling efficiency in (A) different reaction times and (B) different protein concentrations. In all cases, β-lactoglobulin was labeled in a reaction solution containing 6 M urea, 0.05% SDS, 5 mM EDTA, 50 mM Tris buffer, pH 8.3, and 5 mM TBP. (A) Reactions were performed for (2) 10, (3) 30, and (4) 60 min and (5) 2, (6) 6, and (7) 24 h. (B) Protein concentrations were (2) 0.1, (3) 0.2, (4) 0.9, (4), 1.8, and 3.6 μg/μL. In both experiments, lane 1 represents an unlabeled β-lactoglobulin as a control.

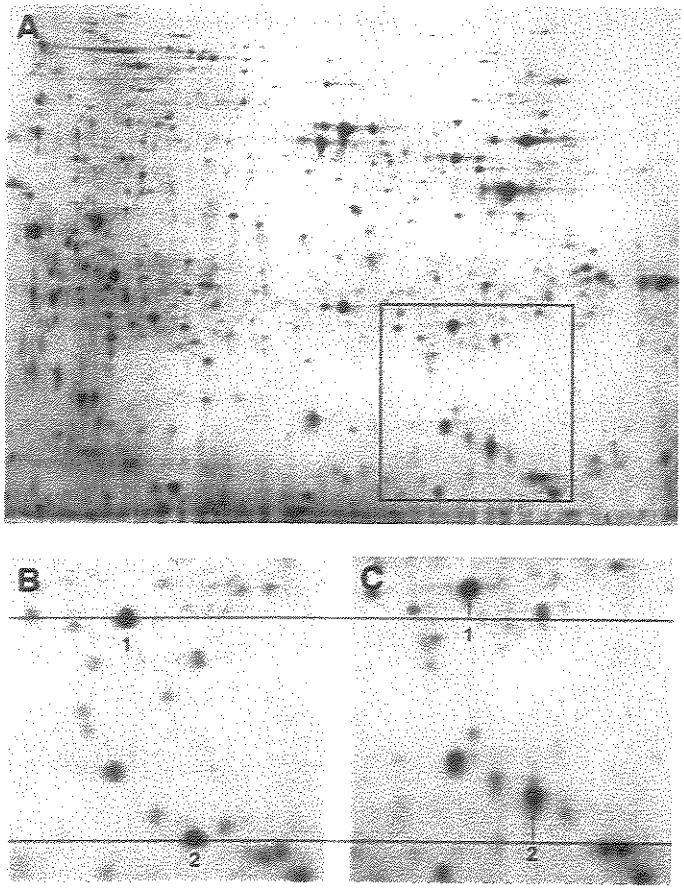


FIG. 7. Analysis of labeling efficiency of total yeast protein extract. (A) 2D-PAGE profile of ICAT-labeled yeast proteins. Partial labeling profile (fuzzy bands as shown in Fig. 3) could not be observed. (B and C) Comparison of the region indicated by the rectangle in A from unlabeled and labeled sample, respectively. The position of non-cysteine-containing protein spots were used to align both gels. 1 and 2 are examples of complete shifting spots.

constitutes a 1:1 molar ratio of ICAT reagent to sulfhydryls, essentially complete labeling was achieved. This indicates that efficient labeling can be achieved at high sulfhydryl concentration even when the ICAT reagent is not in large molar excess.

Labeling of Proteins in Complex Mixtures

To test the suitability of the labeling conditions optimized above on single proteins for the application to complex protein mixtures, a total yeast protein extract was labeled. Three hundred micrograms of the precipitated yeast protein sample was dissolved in 50 μL of reaction solution containing 6 M urea, 0.05% (w/v) SDS, 5 mM EDTA, 50 mM Tris buffer, pH 8.3, and 5 mM TBP and labeled with ICAT reagent at a concentration of 3 mM for 2 h. Considering that proteins from yeast show, on average, a mass of 40 kDa and contain seven cysteine residues, the amount of ICAT reagent used (175 nmol) represented an approximately threefold molar excess over free sulfhydryl groups. The proteins were separated by 2DE and visualized by silver staining. No indication of partial labeling (fuzzy spots) was observed (Fig. 7A). The efficiency of the labeling reaction is further illustrated in Figs. 7B and 7C, in which a section of the 2DE gel was used to compare the protein profiles of the unlabeled and labeled samples, respectively. The clear shifts to higher molecular weights for all the protein spots in the segment shown suggests that complete labeling was achieved. Furthermore, we detected no substantial difference in the isoelectric points (pI) of labeled and unlabeled proteins of the same species represented in the 2DE profile, and the number of spots detected in the gel segments containing the labeled and unlabeled proteins was virtually the same. These results suggest that little or no side reaction of the ICAT reagent occurred with amino acids with charged side chains and that the presence of urea in the labeling solution did not lead to detectable protein carbamylation. The apparent shift to higher molecular weight and absence in change in pI of proteins were also observed for all other regions of the 2D gels of labeled and unlabeled proteins (not shown).

DISCUSSION

In this study we have systematically examined the conditions required to achieve complete labeling of proteins using the ICAT reagents. Specifically, the following variables were tested and the following observations were made: (i) Varying the concentration of ICAT reagent we determined that, as expected, the rate of the reaction was concentration dependent and that a minimal concentration of 3 mM was sufficient to completely label proteins from standard and complex mixtures provided that this reagent concentration represented an excess of reagent over substrate. (ii) Varying the type and concentration of the protein denaturing agents used, we found that urea improves and SDS compromises the reaction of ICAT reagents. This may be due to the fact that SDS as a detergent binds to proteins and thus interferes with the accessibility of ICAT molecules to the sulfhydryls. Urea, in contrast, denatures proteins without binding to them, making sulfhydryl residues more accessible. (iii) Investigation of the reaction time indicated the minimal time required to achieve complete labeling. Minimal reaction time should be used in order to minimize the occurrence of side reactions.

We found that optimal labeling occurred in 6 M urea, 0.05% (w/v) SDS, 50 mM Tris (pH 8.3), 5 mM EDTA, and 5 mM TBP, where ICAT reagent was present at a minimal concentration of 1 mM as well as at a molar excess of minimally 1.1:1 over free sulfhydryls. It is well known that reduction and alkylation of cysteines

are more efficient in pH between 8 and 9 (6), so buffering is an important factor.

In the optimized condition, complete labeling of proteins was shown to be fast (essentially complete in 60 min). Therefore, the reaction was quenched with an excess of DTT at the indicated time to prevent side reactions between the ICAT reagents and the side chains of other amino acid residues such as histidine and lysine. Moreover, although TBP is known to react with alkylating agents such as the ICAT reagent (7), it appears that this reaction is much slower than the reaction of the reagents with sulfhydryls, since a larger molar excess of TBP present in the reaction solution did not interfere with ICAT labeling of proteins. We thus expect that the addition of an excess of DTT after completion of the labeling reaction will inactivate residual ICAT reagent and therefore limit the adduct formation with TBP. The observation that the reaction takes place under denaturing condition at a relatively short reaction time also prevents undesirable proteolysis. This is of particular importance in studies on cell or tissue lysates which are frequently rich in denaturation-resistant proteases. We furthermore showed that the amount of ICAT reagent can be reduced, as long as its concentration is above 1 mM. Reducing the reaction volume to accommodate these concentration constraints can therefore provide considerable saving in reagent cost.

The efficiency of the labeling reactions was assayed by observing the electrophoretic mobility of differentially labeled protein species by SDS–PAGE. We chose 12.5% total acrylamide gels because their high resolution in the 10- to 40-kDa region allowed us to better observe the mobility shifts for α-lactalbumin and β-lactoglobulin. Not surprisingly, the observed mobility shift for BSA (36 cysteines) was indeed smaller than that of α-lactalbumin (8 cysteins). The bands containing the completely labeled proteins had virtually the same specific staining intensity as the band of the completely unlabeled protein. This was expected since Coomassie blue staining depends on the side chains of basic amino acids, particularly the epsilon amino group of lysine and these are unaffected by the ICAT chemistry. Incompletely labeled proteins were detected as multiple bands with a sometimes "fuzzy" appearance (see Fig. 2, lanes 3 and 4). The presence of these bands of intermediate mobility and the staining intensity of the band containing the completely labeled protein were therefore used as criteria to assess the efficiency of the reaction.

To validate the optimized labeling protocol for a complex protein mixture we applied it to the labeling of the proteins in a total protein extract from the yeast *S. cerevisiae* and assayed by 2DE. It should be noted that the pI of basic proteins is expected to change after labeling. This was, however, not detected in the experiment. Differential migration should only be observed for proteins with pI over 8 (close to the pK_a of cysteine sulfhydryl group). These proteins are not well resolved with the 3–10 (nonlinear pH gradient) IPG strip that was used. By monitoring the pI of proteins with pI lower than 8, it was possible to analyze unspecific labeling of charged amino acids. Unspecific labeling would not be apparent in basic proteins as it would not be possible to differentiate between specific and unspecific labeling. We concluded that complete and cysteine specific labeling was achieved with minimal side reactions.

In this report, we show that specific and quantitative labeling of proteins by ICAT reagents can be achieved and we describe optimized labeling conditions. This should be useful for the application of the ICAT technique to quantitative proteome analyses.

ACKNOWLEDGMENTS

This work was supported by grants from the Merck Genome Research Institute (MGRI) and the National (USA) Cancer Institute to Ruedi Aebersold and by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via a fellowship to Marcus Smolka. We thank all the members of Aebersold's lab for the useful discussions.

REFERENCES

1. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17,** 994–999.
2. Ducret, A., Van Oostveen, I., Eng, J. K., Yates, J. R., 3rd, and Aebersold, R. (1998) High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci.* **7**(3), 706–719.
3. Moore, D. D., *et al.* (Eds.) (1987) Current Protocols in Molecular Biology, Wiley, New York.
4. Blum, H., Beier, H., and Gross, H. J. (1987) *Electrophoresis* **8,** 93–99.
5. Schevechenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68,** 850–858.
6. Yan, J. X., Sanchez, J. C., Rouge, V., Williams, K. L., and Hochstrasser, D. F. (1999) Modified immobilized pH gradient gel strip equilibration procedure in SWISS-2DPAGE protocols. *Electrophoresis* **20,** 723–726.
7. Getz, E. B., Xiao, M., Chakrabarty, T., Cooke R., and Selvin, P. R. (1999) A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal. Biochem.* **273,** 73–80.

Conclusões Gerais

Capítulo 1: Análise Proteômica da *Xylella fastidiosa*.

A presente tese apresentou a detecção e identificação das proteínas mais expressas pela linhagem 9a5c da bactéria *Xylella fastidiosa* causadora da clorose variegada do citros, crescida em meio de cultura. Ao todo, os produtos de 142 genes diferentes foram identificados por espectrometria de massas.

A técnica de eletroforese 2DE foi estabelecida com a máxima resolução, reprodutibilidade e sensibilidade possível por otimizações das etapas de preparação de amostra, isoeletrofocalização (primeira dimensão), SDS-PAGE (segunda dimensão) e visualização (coloração por nitrato de prata ou coomassie blue). O método de "peptide mass fingerprint" para identificação de proteínas por espectrometria de massas também foi estabelecido e otimizado para máxima sensibilidade, possibilitando a identificação de proteínas coradas pela coloração de nitrato de prata, presentes em quantidades inferiores a 500 fentomoles.

A técnica de SDS-PAGE em gel contendo 9 % de poliacrilamida possibilitou que dezenas de proteínas com massa molecular acima de 100 kDa, não detectadas antes por 2DE, pudessem ser detectadas e identificadas.

Acreditamos que proteínas potencialmente envolvidas com mecanismos de colonização e patogenicidade foram mapeadas. A porina OprF foi detectada como sendo a mais abundante proteína de membrana externa e devido ao fato do xilema ser um ambiente extremamente hiposmótico, sua alta abundância deve ser fator crucial para a *X. fastidiosa* sobreviver na planta.

Foi detectada a co-expressão de proteínas putativamente associadas a três diferentes sistemas de adesão (fimbria tipo IV, adesina MrkD e fibrilas de superfície do tipo Hsf). As adesinas MrkD e Hsf foram detectadas formando complexos multiméricos de alta massa molecular (acima de 300 kDa) resistentes à desnaturação por SDS e DTT. Maiores análises devem ser feitas para se obter uma detalhada caracterização estrutural e funcional destas estruturas adesivas.

Trinta proteínas presentes no meio extracelular foram detectadas e identificadas. Várias destas proteínas estão relacionadas a mecanismos de defesa, nutrição, homeostase de ferro e toxicidade, sugerindo que o espaço intercelular de colônias de *X. fastidiosa* seja um microambiente multifuncional importante para sobrevivência e

patogênese da bactéria. Foram encontradas proteínas com reconhecida atividade citotóxica, como a toxina RTX. Outras 16 proteínas sem função conhecida também foram identificadas na fração extracelular.

Proteínas detectadas no extrato celular total como isoformas com diferentes pIs e que também foram detectadas na fração extracelular, estavam majoritariamente em sua forma mais negativa quando secretadas. Isto sugere a existência de um mecanismo pós-traducional de regulação de secreção de proteínas que envolva fosforilação.

A utilização de meio sólido para crescimento das bactérias se mostrou eficaz para extração de proteínas extracelulares, principalmente porque elas são concentradas no espaço intercelular das colônias, permitindo que consideráveis quantidades sejam recuperadas, diferentemente de crescimento em meio líquido, onde proteínas se difundem no meio de cultura.

Análise comparativa do perfil protéico da bactéria crescida como colônias isoladas em meio sólido de cultura com o da bactéria formando biofilme em meio líquido de cultura indicou a expressão específica de uma cisteíno protease (XF0156), uma lipase (XF2151) e uma isoforma da subunidade da fímbria do tipo IV (XF2539) na condição de formação de biofilme. Maiores estudos funcionais destas três proteínas serão importantes para esclarecer o mecanismo de adesão e formação de biofilme.

A subunidade fimA da fímbria tipo IV pode ser expressa por diferentes genes em diferentes condições de crescimento. A XF2539 é expressa no crescimento em meio líquido (formação de biofilme) e a XF2542 no crescimento em meio sólido, sugerindo a existência de diferentes mecanismos de regulação de expressão de cada isoforma.

Foi possível detectar a ocorrência de valores extremamente baixos de *codon bias* em genes altamente expressos. É proposto que isto seja a causa da natureza fastidiosa da *X. fastidiosa*. O conhecimento de que esta bactéria não possue um mecanismo otimizado para utilização preferencial de codons para a síntese protéica pode contribuir para entender vários aspectos biológicos deste patógeno. Em última análise, esta "falha" metabólica no processo de síntese de proteínas pode ser usado em estratégias de controle da doença.

O banco de dados foi desenvolvido e deverá auxiliar pesquisadores de outras instituições em análises de expressão de proteínas da *X. fastidiosa*.

O Apêndice da presente tese traz um capítulo de livro detalhando todos os procedimentos da 2DE e identificação de proteínas por PMF.

Capítulo 2: Desenvolvimento de Novo Método para Análise Quantitativa de Proteomas por Espectroemetria de Massas

Foi desenvolvido um novo método de análise quantitativa de proteomas baseado na combinação da 2DE, marcação de proteínas com reagentes do tipo ICAT e espectrometria de massas.

Foi mostrado que marcação isotópica de proteínas com reagentes ICAT é compatível com 2DE e quantificação precisa de proteínas por espectrometria de massas. O método apresentou precisão de quantificação melhor que 20 %. Ainda, foi mostrada sua aplicabilidade para análise quantitativa de proteínas em amostras complexas, como o proteoma de *S. cerevisae*. O método se mostrou particularmente útil para a análise quantitativa e caracterização estrutural de diferentes formas processadas ou pós-traducionalmente modificadas de uma proteína e espera-se que tenha vasta aplicação em pesquisas proteômicas.

Um artigo sobre otimização do procedimento de marcação das proteínas com o reagente ICAT mostrou que é possível atingir marcação completa das cisteínas de proteínas presentes em amostras complexas. Para isso, importantes fatores como concentração de SDS, uréia, proteína e reagente ICAT devem ser ajustados precisamente. A capacidade de realizar marcações completas é de fundamental importância para o sucesso da aplicação do ICAT para análises quantitativas de proteomas.

Perspectivas

Este trabalho representa apenas o início da análise proteômica da *X. fastidiosa*. É necessário um estudo mais detalhado das proteínas apontadas como potenciais alvos. A expressão destas proteínas em sistema recombinante (por exemplo, *E. coli*) será fundamental para viabilizar estudos de caracterização estrutural (por cristalografia) e de atividade bioquímica. Além disso, a atual capacidade de transformar geneticamente a *X. fastidiosa* possibilita que um gene de interesse seja retirado e sua funcionalidade analisada por comparação das características fenotípicas da bactéria mutante (sem o gene) com as de uma bactéria selvagem. Isto será importante para comprovação do envolvimento das proteínas com a capacidade de colonização e patogenicidade.

Colaborações com pesquisadores de diferentes instituiçoes de pesquisa com vasta experiência no estudo da CVC e da *X. fastidiosa* foram parte integrada ao presente projeto de doutorado. Experimentos abordando diferentes aspectos da biologia da bactéria já estão em fase adiantada (Figura 17) e devem resultar na definição de mais alvos potenciais e em importantes contribuições para o entendimento dos mecanismos de patogenicidade.

O novo método de análise quantitativa de alta precisão apresentado nesta tese irá possibilitar que pequenas alterações em abundância das proteínas possam ser monitoradas nos modelos em estudo, contribuindo para um estudo sistemático mais global e apurado das expressões diferenciais.

A possibilidade de utilizar reagentes ICAT com diferentes quantidades de deutérios (por exemplo, quatro e doze) possibilitará que mais de duas amostras sejam analisadas simultaneamente no mesmo gel. Estudos já estão sendo feitos neste sentido (Figura 18).

FIGURA 17: Organograma de colaborações já em andamento. Estes grupos fazem parte do Programa Genoma Funcional da FAPESP.

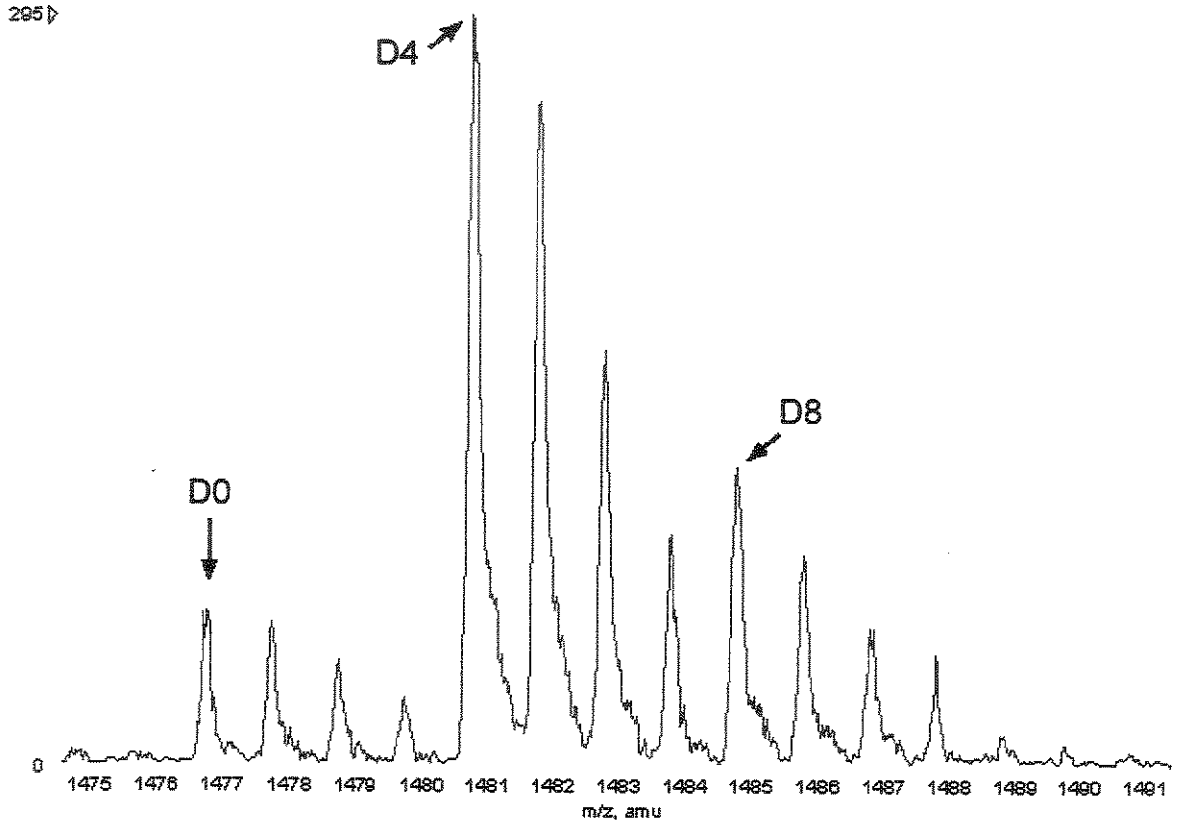


FIGURA 18: Espectro de massas de peptídio marcado com três formas isotópicas (0, 4 e 8 deutérios) do reagente ICAT. A utilização de três formas isotópicas permite que três amostras diferentes sejam comparadas simultaneamente. Espera-se que até cinco ou seis formas istópicas possam ser feitas para o ICAT, o que permitiria a análise simultânea de cinco ou seis amostras diferentes. A quantificação pelo sinal espectrométrico indica este peptídio está na proporção 1 : 5 : 2.

Apêndice:

Capítulo de Livro:

Análise Proteômica Utilizando Eletroforese de Duas Dimensões e Espectrometria de Massas

Capitulo de livro a ser publicado pela Editora Humana Press

Marcus B. Smolka and Ruedi Aebersold **Proteome Analysis Using Two Dimensional Gel Electrophoresis and Mass Spectrometry**

Proteome Analysis Using Two Dimensional Gel Electrophoresis and Mass Spectrometry

Marcus B. Smolka[1] and Ruedi Aebersold[2]

1 Laboratório de Química de Proteínas, Departamento de Bioquímica, Universidade Estadual de Campinas – UNICAMP, Campinas, SP, Brazil

2 The Institute for Systems Biology, Seattle, WA

Contact: Ruedi Aebersold, Institute for Systems Biology, 4225 Roosevelt Way NE. Seattle, WA 98105, USA.

Phone: 206 732 1204

Fax: 206 732 1255

Email: raebersold@systemsbiology.org

ABBREVIATIONS

2DE	two-dimensional polyacrylamide gel electrophoresis
MS	mass spectrometry
IEF	isoelectric focusing
pI	isoelectric point
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
MW	molecular weight
MALDI	matrix-assisted laser desorption-ionization
TOF	time-of-flight
LC	liquid chromatography
ESI	electrospray ionization
MS/MS	tandem mass spectrometry
CID	collision induced dissociation
PMSF	phenylmethylsulphonyl fluoride
EDTA	ethylenediaminetetraacetic acid
SDS	sodium dodecyl sulfate
CHAPS	3-[3-cholamidopropyl) dimethylammonio]-1-propanesulphonate
NP-40	Nonidet P-40
DTT	dithiothreitol
DTE	dithioerythritol
BPB	bromophenol blue
IPG	immobilized pH gradient
CBB	coomassie briliant blue
APS	ammonium persulfate
TEMED	N,N,N',N'-tetramethylethylenediamine
EST	expressed sequence tag
MALDI-TOF MS	MALDI-TOF based mass spectrometry

1. Introduction

A central element of proteomics is the intent to systematically identify and quantify the proteins expressed by a cell or tissue or those present in complex samples such as body fluids. The most commonly applied technology for this task is the combination of high-resolution two-dimensional polyacrylamide gel electrophoresis (2DE) for the separation, detection and quantification of proteins, and mass spectrometry for the identification of the separated proteins. Whereas the 2DE technique was already established in the 1970's (1,2), most of the protein identification procedures suitable to identify proteins separated by 2DE rapidly, conclusively and sensitively were only developed relatively recently (3). The keys to these protein identification methods were developments in mass spectrometry (MS), the availability of whole genome sequences or other large sequence databases and the development of software tools for the correlation of data obtained by mass spectrometry with sequence databases. Large-scale protein identification/quantification and proteome comparisons by differential 2DE analysis are now important tools to detect and identify proteins that are involved in th control of biological processes or whose abundance is controlled by them. Typica applications include the search for molecular disease markers, the search for markers indicating the response of a cell or tissue to external stress, perturbations or insults, the identification of molecular markers that define cell types and differentiate cellular states and, most ambitiously, the reconstruction of biochemical processes based from protein profiles.

2DE combines two acrylamide-gel based separation techniques (Figure 1). In the first dimension separation is achieved by isoelectric focusing (IEF) in which proteins are separated by their isoelectric point (pI). In the second dimension separation is achieved by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) that separates proteins based on their molecular weight (MW). The separated proteins are

then stained directly in the gel resulting in a two-dimensional array of spots, each one usually containing multiple copies of a homogeneous protein. The amount of protein migrating to each spot is quantified based on the staining intensity of the respective spot. Usually the gel image is digitized and densitometrically analyzed using computer programs. Different gel patterns representing different proteomes, e.g. the proteins expressed by a cell in different metabolic states, can thus be compared for the purpose of detecting differentially expressed proteins.

For protein identification (Figures 2 and 3), spots from the 2DE gels are excised and in-gel digested with a specific protease. The resulting fragments are extracted and analyzed by one of a variety of mass spectrometric methods (3,4). The most commonly applied methods are matrix assisted laser desorption ionization (MALDI) time-of-flight (TOF) and liquid chromatography (LC) electrospray ionization (ESI) tandem mass spectrometry (MS/MS). In both methods protein identities are determined by comparing the experimentally obtained data with the contents of sequence databases, although the MS/MS method in principle also permits the explicit determination of peptide sequences from the obtained collision-induced dissociation (CID) spectra. In relation to other techniques such as protein sequencing by the Edman degradation (5) and protein identification by the amino acid composition (6), mass spectrometry identifies proteins faster and at higher sensitivity.

Even though 2DE with its potential to separate hundreds to thousands of proteins is currently the most commonly used technique for the separation of complex protein samples, it also has some significant limitations. Specific classes of proteins including very acidic (<3.5), very basic (>9.0), small (<6000Da), very hydrophobic and very large (>120000Da) proteins, and proteins of low abundance are generally not detectable with the currently available separation and visualization methods (7,8). For the analysis of those proteins that are visible 2DE is however a powerful separation

method because all the proteins in the sample are concurrently separated, information concerning the pI, MW and relative abundance can be obtained for all the detected proteins, and the separated proteins can be submitted to post separation analysis. Furthermore, the presence of post-translational modifications is usually apparent based on the altered electrophoretic mobility caused by the modification.

In this chapter, we describe and discuss the experimental protocols for performing 2DE and MS based protein identification.

2. Sample preparation

Before proteins can be separated and analyzed, they first have to be extracted from cells or tissues and solubilized. The sample preparation procedures are of critical importance for the success of the experiment and are somewhat sample specific. Generally the protocols involve disruption of the cells or tissue and extraction of the proteins into a buffered solution that also contains strong protein solubilizing agents and protease inhibitors. To obtain reproducible and accurate results care has to be taken to minimize proteolysis and unwanted reactions such as oxidation and carbamylation. The protein concentration of the sample solution should be determined to allow for optimal loading of the gel and to assure that equal amounts of protein are analyzed in cases in which quantitative comparisons of the 2DE patterns are attempted. The preferred staining methods for protein quantitation are the enhanced alkaline copper (9), bicinchoninic acid (10) and coomassie blue (11). Protein assay kits are available from Bio-Rad (Richmond, CA), Pierce (Rockford, IL) and Sigma (St. Louis, MO).

1.1. Cell and tissue disruption methods

The objective of the disruption methods is to extract all the proteins from cells or tissue and to expose them as quickly as possible to a solubilization solution that minimizes unwanted reactions such as proteolysis and protein modifications. Table 1 lists some of the commonly used disruption methods. The methods have to be adapted to the nature of the sample being analyzed. Some samples such as strongly connected tissues require strong disruption methods that combine mechanical homogenization and strong solubilizing solutions, whereas some bacteria and many cultured eukaryotic cells can be lysed and extracted by exposure to solutions containing detergents. To avoid irreproducible or selective protein extraction, the disruption should be complete and uniform.

Generally, disruption is already performed in a solubilizing solution so that proteins are immediately extracted, solubilized and denatured. Performing the disruption in strongly denaturing conditions has the added advantage that proteases which are abundant in cells and tissues are also denatured and that proteolysis is therefore minimized. Since residual protease activity might persist in spite of the strongly denaturing conditions, the use of protease inhibitors is highly recommended. The most abundant proteases present are usually serine proteases or metalloproteases. Phenylmethylsulphonyl fluoride (PMSF) is commonly used to irreversibly inhibit serine proteases and ethylenediaminetetraacetic acid (EDTA) to inhibit metalloproteases by chelation of free metal ions. Buffering the extraction solution at pH over 8.5 and keeping the sample on ice during the extraction and solubilization steps can also reduce protease activity.

Some samples can have high contents of undesirable contaminants such as nucleic acids, sugars and lipids. These molecules usually interfere with the 2DE

separation and should be removed prior to gel electrophoresis, e.g. by protein precipitation (16,17).

1.2. Protein solubilization and denaturation

Ideally, protein solubilization and denaturation methods should bring all the proteins into solution and destabilize their tertiary structures. The better the proteins are solubilized and denatured, the better 2DE will be able to efficiently resolve them. In spite of recent advances in the development of solubilizing reagents compatible with 2DE (18,19) it may be difficult or impossible to find solubilization conditions that are also compatible with IEF for some hydrophobic proteins (mainly membrane proteins). Therefore, if total cell lysates are analyzed by 2DE such proteins tend to be underrepresented or completely absent (20-22).

Solubilization and denaturation of proteins can be performed using high concentrations of detergents, reducing agents and heat. Numerous detergents and denaturants have been used for protein solubilization. However, since the first dimension separation (IEF) is based on the intrinsic charge of the denatured protein in solution, anionic or cationic detergents that attach themselves to the protein such as the very effective solubilizing agent sodium dodecyl sulfate (SDS) are incompatible with 2DE. Therefore, 2DE is commonly performed in the presence of neutral or zwiterionic detergents. These include 3-[3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), Nonidet P-40 (NP-40) and Triton X-100. The use of limited amounts of SDS to initially solubilize the proteins is still compatible with 2DE as long as the anionic detergent is displaced in the gel by a neutral detergent (see in the suggested sample preparation method). Reducing agents such as dithiothreitol (DTT) or dithioerythritol (DTE) are also frequently used in high concentrations (50-100mM) to disassociate

disulfide linked protein complexes and to improve protein solubilization and denaturation.

Based on our experience the combination of a high SDS concentration (1%), the presence of 100mM DTT and heating of the sample to 95°C for 5 minutes achieves a high degree of protein solubilization/denaturation for most samples. The protein concentration should be in excess of 10 μg/μL, so that the sample can be diluted to a final SDS concentration of 0.05% or lower prior to IEF, while an appropriate protein concentration is maintained. DTT is rapidly oxidized in solution and reduction of disulphide bonds is reversible (especially at alkaline pH and at elevated temperature). Therefore, even if DTT was present during protein solubilization, an excess of fresh DTT should be added immediately prior to applying the sample to the first dimension gel.

We recommend that the sample be diluted in a solution containing 8M urea, 4% CHAPS, 70mM DTT and 0.005% Bromophenol blue (BPB) prior to IEF. Furthermore, carrier ampholytes (in concentrations from 0.5 to 2%) should be included in the sample, even if IEF is performed in immobilized pH gradient (IPG) gels, to help in the initial migration of the proteins along the pH gradient. The proper ampholyte concentration should be empirically determined for each sample. This is best achieved by monitoring the current through the gel during electrophoresis. With increasing ampholyte concentration the electrical current through the IEF strip will also increase. If too much current is passing through the strip, Ohm heating will limit the maximal voltage that can be applied and therefore interfere with the performance of the IEF separation. An optimal ampholyte concentration will allow the voltage to rise to close to its maximum while good protein solubility and initial protein migration are maintained (refer to section 2).

For solubilization procedures that include heating of the sample, particularly at neutral or basic pH and for long term sample storage, urea should be avoided. Cyanate

ions formed under these conditions will carbamylate proteins, thus altering their pI. We therefore recommend to add urea immediately prior to IEF.

Suggested protocol: The following protocol represents a basic sample preparation procedure that may have to be adapted for each sample for optimal performance:

- **Lysis:** lyse sample in 20mM Tris (pH 8.8), 1% SDS, 5mM EDTA and 1mM PMSF. If homogenization is needed, add SDS after disruption to avoid foam formation. Let sample incubate for 5 minutes for the PMSF to irreversibly inactivate serine proteases. DTT and thiol reagents should be added after protease inactivation because their presence interferes with PMSF protease inactivation. All the procedure should be performed on ice;
- **Protein assay:** the volume of the lysis solution used in the first step should be adapted to result in a protein concentration of approximately 10μg/μL. **NOTE:** DTT strongly interferes with most protein assays, so an aliquot for measuring the protein concentration should be removed prior to adding the DTT;
- **Reduction:** add DTT to a final concentration of 100mM. The combination of high SDS and DTT concentrations helps in protein solubilization/denaturation and protease inhibition. At this stage, samples can be stored at -20°C for several months;
- **Dilution into IEF rehydration solution:** for the analysis of complex protein mixtures, generally 150μg of protein (for silver stained gels) should be applied to an 18cm IPG strip. For general survey gels the nonlinear gradient available in strips designated IPG 3-10NL is preferred. Add the proper volume of sample (if sample is 10μg/μL, add 15μL) to 350μL of a solution containing 8M urea, 4% CHAPS, 70mM DTT, 2% ampholytes (in the pH range corresponding to the pH range of the IPG strip) and 0.005% BPB. **NOTE:** sample needs to be diluted at

least 20 fold to bring the SDS concentration below 0.05%. Let solution incubate for 30 minutes before loading to IEF gel. For gels stained with coomassie brilliant blue (CBB) a protein load of approximately 500μg is recommended. In this case the protein concentration should be increased and/or the SDS concentration should be decreased to comply with the maximal SDS tolerance of the IEF gel. Should this be impossible (e.g. due to limited protein solubility) the proteins should be precipitated and redissolved in a solvent compatible with IEF.

3. First dimension: isoelectric focusing (IEF) using IPG strips

Isoelectric focusing is an electrophoretic technique for separating proteins based on their isoelectric point. Separation is performed in IPG strips, which are made of a polyacrylamide gel matrix that is covalently modified with acidic and basic groups that form an immobilized pH gradient over the length of the gel (23,24). After loading the sample to the gel and applying a voltage, proteins migrate to the pH region corresponding to their pI. At this position the sum of positive and negative charges is zero and the protein does not experience a net electrophoretic force and stops migration.

Dehydrated IPG strips are commercially available in many different sizes and covering different pH ranges. Strips containing broad pH gradients (for example, covering pH 3-10) are preferred for experiments in which as many proteins as possible are going to be surveyed, whereas strips containing a narrow pH range (for example, 4.5-5.5 and 5.5-6.7) are preferred in experiments in which maximal resolution in a particular pH range is required. We recommend the protein sample to be loaded by rehydrating the gel in the sample solution (in-gel sample application). In this method the proteins are soaked into the gel as it swells during rehydration. This is technically

simple, allows high protein quantities to be loaded (25) and prevents high local protein concentrations at the beginning of the electrophoretic process. Electrophoresis is performed in instruments that can generate high voltage (up to 8000 Volts) and have an efficient cooling system. Efficient cooling to a constant temperature is necessary to prevent overheating of the gel during the initial electrophoretic phase in which relatively high currents flow and to prevent gel-to-gel variation in protein mobility due to the temperature dependence of the pI.

Electrical parameters should to be closely monitored and controlled during a run. When using the IPGphor system (or a similar system) the current is usually limited to 50μA per strip while the voltage is slowly raised from 100V to 8000V. If a sample contains too high concentrations of salt, ampholytes or other substances that increase conductivity, the maximum voltage may not be reached because the limiting current of 50μA will be reached at a lower voltage. If the voltage reached at the end of a run is too low (lower than 4000Volts) band sharpness and resolution are compromised. Therefore, the source of conductivity should be identified to improve electrophoretic performance. If the excess conductivity is caused by a high salt concentration in the sample the proteins should be precipitated and re-dissolved in a solution of lower conductivity prior to application to the gel.

Suggested protocol: Table 3 lists the reagents and equipment required for performing IEF separation:

- **In-gel sample loading:** (sample loading + strip rehydration) for 18cm strips, 350μL of the final rehydration/sample solution should be loaded per strip. Immediately before loading, spin sample at 14000 rpm in an Eppendorf centrifuge to pellet any insoluble material. Use the supernatant and let the strip rehydrate in the IPGphor unit for at least 10 hours at 20°C;

- **Running conditions:** running conditions are not only dependent on the type of sample but also on the amount of sample applied. As a general condition (that might need to be optimized in later experiments with the same type of sample) we recommend an 80kVh run as described in Table 4. If preparative gels containing more than 500μg of total protein are run, the electrophoresis time should be increased to 150kVh and the initial steps at lower voltage should be extended. At the end of the run, strips can be stored at -70°C until the second dimension is run.

4. Second dimension: SDS-PAGE

In SDS-PAGE, proteins are separated according to their molecular weight. Proteins are coated with SDS and subjected to electrophoretic migration through a polyacrylamide matrix. The anionic detergent masks the intrinsic electrical charges of the proteins. Therefore, all proteins have the same density of negative charges. The electrophoretic mobility is restricted by the pore size of the polyacrylamide matrix in a manner that proteins migrate at a velocity that is proportional to their size. The most common SDS-PAGE system is the one described by Laemmli in 1970 (26). It uses a discontinuous buffer system of Tris-glycine.

For 2DE the whole IPG strip with the proteins separated by their pI is incubated in equilibration buffer containing SDS, urea, glycerol and reducing agent. The strip is further treated with a second equilibration buffer of similar composition, except that the reducing agent is replaced by iodoacetamide to alkylate the thiol groups of the reduced proteins, thus preventing their reoxidation. As the reduction and alkylation reactions are more efficient at alkaline pH, Tris is used to maintain a pH of 8.8.

The polyacrylamide matrix used to fractionate the proteins can be of homogeneous or gradient pore size. Homogeneous gels are easier to prepare, but show good resolution over a more narrow range of MW (for example, 12% gels have good resolution in the MW range of 10 to 70kDa). Gradient gels show increased separation performance over a wider range of MW but they are more difficult to prepare. To achieve a high degree of gel-to-gel reproducibility gradient gels are preferably prepared using computer assisted pumping systems. We have found it difficult to achieve satisfactory gel-to-gel reproducibility by pouring gradient gels with gravity-based systems.

The width of the gel has to be equivalent to the length of the IPG strip being used. Regarding its vertical size, bigger gels usually result in improved separation. In contrast to 1D SDS-PAGE the use of a stacking gel is usually omitted in 2DE because the polyacrylamide gel from the IPG strip promotes protein stacking.

Suggested protocol: the materials and solutions required are listed in tables 5 and 6, respectively:

- **Gel preparation:** second dimension gels should be prepared a day in advance:
  1) Clean glass plates with strong detergent and hot water and then assemble gel sandwiches according to the manufacturer's instructions for a particular gel system;
  2) Prepare gel solution (Table 7);
  3) Degas gel solution for 10 minutes. Oxygen inhibits polymerization. Degassing is therefore required for uniform, reproducible and fast gel polymerization;
  4) Prepare APS solution (solution 3);
  5) Polymerization: at this stage is important the gel solution to be at room temperature. Add 400µL of solution 3 and 15µL of TEMED to gel solution.

Slowly pour gel solution in the assembled gel sandwich up to 1cm below the top of the gel plate and cover the unpolymerized gel solution with 400μL of 50% butanol in water. Let gel polymerize for 2 hours and then exchange butanol with distilled water. **NOTE:** polymerization is affected by many different factors including temperature, presence of oxygen, amount of the catalysts APS and TEMED, pH, concentration of monomer and cross-linker, so these conditions should be tightly controlled in order to achieve good quality and reproducible polymerization;

- **Strip equilibration:**

1) Before run, incubate strips for 10 minutes in the first equilibrating solution (solution 6; reduces disulfide bonds and coats proteins in SDS) and for an additional 10 minutes in the second equilibrating solution (solution 7; alkylates cysteines and coats proteins with SDS);
2) Put equilibrated strip over the previously prepared gel and cover with melted agarose solution (solution 8) heated to 70ºC. This is to keep the strip in contact with the upper surface of the gel. For 18cm strips we use 14x16cm slab gels and cut 1 cm off from both ends of the IEF gel strip;

- **Electrophoretic run:**

1) Fill upper and lower chambers of the electrophoresis unit with running buffer (solution 5). Upper buffer (500mL) has to be new for each run as the electrolyte concentration changes during electrophoresis. Lower buffer (around 4 liters) can be re-used for up to 20 runs if after every run 500mL of it is replaced with 500mL of the upper buffer just used. This is because Tris migrates to the cathode (in the upper buffer) during the run;

2) Optionally, a cooling unit can be used to maintain the temperature of the gels constant during the run, improving reproducibility. We recommend to perform electrophoresis at 15ºC;
3) Start run with 90V (fixed) for 30 minutes and then with constant amperage of 30mA/gel (free voltage) until the blue font reaches the bottom of the gel (approximately 3 to 4 hours). At the end of run, soak gels for 5 minutes in distilled water (for removal of excess SDS) and fix them overnight in methanol 40% and acetic acid 10% prior to silver staining, or directly in the CBB staining solution.

5. Visualization and quantification

5.1. Staining methods

A number of different methods for the detection of proteins separated by polyacrylamide gel electrophoresis have been used. These include the detection of radiolabeled proteins by fluorography or autoradiography, staining of proteins post separation with dyes and fluorescent molecules and the covalent dye-labeling of proteins prior to electrophoresis (27,28). Among these the most common methods are CBB and silver staining. Both are simple to perform, inexpensive and compatible with subsequent MS analysis of the stained proteins, provided that suitable protocols are being used. Silver staining is one of the most sensitive staining methods (1-10 ng/spot), but is difficult to use for accurate quantification because the amount of silver bound is protein specific, the dynamic range is limited and because the staining intensity is strongly dependent on the development time, making quantitative comparisons from gel to gel difficult (29-31). Common silver staining procedures use sensitization reagents such as formaldehyde

and glutaraldehyde that may promote chemical protein modifications and/or cross-linking reactions between proteins and the acrylamide matrix and are therefore unsuitable for the detection of proteins prior to MS analysis. Modified silver staining procedures specifically adapted for MS analysis of the detected proteins have been described and should be followed (32-34). CBB staining is MS compatible and fairly quantitative in the protein concentration range of 100-1000 ng/spot, the sensitivity of detection is however 10-50 times lower than that of silver staining (27,35). Figure 4 shows the image of two silver stained 2DE gels prepared as described in the protocols.

Fluorescent and radioactive labeling methods are highly sensitive and quantitative over a wide range of protein abundance. These tend to be more expensive and, in the case of radioactive labeling, limited to samples for which radiolabeling (metabolic or post isolation) can be performed. A series of recent papers have described the use of fluorescent stains such as SYPRO Orange, SYPRO Red and SYPRO Ruby for the detection of gel separated proteins by non-covalent staining (27,36,37). A method for differential fluorescence gel electrophoresis has also been described. In this method proteins are covalently labeled prior to electrophoresis with one of two fluorescent dyes (e. g., Cy™2 and Cy™5). Two samples, one of each being labeled with either dye are then combined and separated in the same gel (38,39). As proteins labeled with either dye co-migrate in the gel the total light emitted from each spot is composed from the light emitted from the two dyes. The spectral properties of each detected spot indicate the relative contributions of each dye and therefore indicate the relative abundance of the two combined samples. While this method is attractive because it effectively eliminates electrophoretic variation for the two co-separated samples, the method needs to be further refined to achieve the sensitivity and reproducibility required for routine use. One of the major drawbacks of the current implementation is that the epsilon amino groups of the abundant amino acid lysine are

labeled with relatively hydrophobic dyes, thus reducing the solubility of extensively labeled proteins. To avoid protein precipitation in the gel, minimal labeling is performed, limiting the sensitivity of the method and making equal quantitative labeling between two samples difficult.

Suggested protocol for CBB staining: tables 8 and 9 list the reagents, equipment and solutions required for CBB staining. The procedure described is for two 14 X 16 X 0.15 cm gels. If the stained proteins are intended for subsequent MS analysis, gels need to be handled with gloves to avoid keratin contamination:

- **Staining:** after SDS-PAGE, soak gels in 300mL of distilled water for 5 minutes. This removes excess SDS and is necessary to avoid the interference of SDS with CBB binding to the protein. Removal of excess SDS also makes the staining solution re-usable. The gels are then soaked in 300mL of solution 9 for at least 3 hours but not for more than 24 hours. Because staining solution contains methanol and acetic acid, it also promotes protein fixation and the staining time should be limited.
- **Destaining:** remove staining solution, quickly wash the gel with distilled water to remove excess of CBB, and destain gels by soaking them 3 times in 300mL of destaining solution (solution 10). Each destaining step should be for 30 minutes and performed under constant agitation. Then, soak gels in solution 11 for 30 minutes, followed by as many changes in solution 10 as deemed necessary to achieve the required contrast. To completely clear the background, incubate gels overnight in solution 11.

Suggested protocol for silver staining: tables 10 and 11 list the reagents, equipment and solutions required for silver staining. The procedure is adapted from Shevchenko et al., 1996 (32). The critical factor that distinguishes these protocols from other commonly

used silver staining protocols is that a reduced amount of formaldehyde is used to avoid protein fixation. The volumes indicated are suitable for staining two 14 X 16 X 0.15 cm gels. Silver staining is highly affected by contaminants, so all the reagents and water should be as pure as possible. It is important to use extremely clean glass trays (we recommend washing them with 50% nitric acid prior to staining) and, if proteins are intended for subsequent MS analysis, gels need to be handled with gloves to avoid keratin contamination:

- **Fix:** after SDS-PAGE, soak gels in 300mL of distilled water for 5 minutes and then in 300mL of solution 12 for 30 minutes;
- **Wash 1:** wash the gels with 4 changes of 400mL distilled water, each wash being 15 minutes. All the acid should be removed prior to the next step;
- **Oxidizer:** incubate the gels for exactly 2 minutes in solution 13;
- **Wash 2:** wash the gels with 3 changes of 400mL distilled water, each wash being for 30 seconds;
- **Silver incubation:** incubate gels in 300mL of solution 14 for 25 minutes;
- **Wash 3:** wash the gels with 3 changes of 400mL distilled water, each wash being 60 seconds;
- **Develop:** incubate gels in 300mL of solution 15 until spots are stained. Stop developing before background gets too dark;
- **Stop:** soak gels in solution 16 for 10 minutes to stop developing;
- **Storage:** preserve gels in distilled water at 4°C until spots are picked for MS analysis. It is advantageous to picked the spots as soon after staining as possible to minimize spot diffusion. We recommend to cut out the spots and store them individually in Eppendorf tubes at -20°C.

5.2. Image analysis

For image analysis, the gels are digitized using a scanner and the generated files (usually in TIFF or similar file format) are analyzed in specialized programs. While individual programs vary a great deal they are generally able to detect and densitometrically quantify the spots and to calculate the pI and MW of the proteins contained in them. Most programs also have the ability to subtract staining background and to compensate, at least to some extent for local warping in the gel patterns. Comparative image analysis of different gels is used to detect differentially expressed proteins. The determination of observed pI and MW values is important because these parameters are useful for the confirmation of the protein identity determined by MS or other methods (see below). Many programs for image analysis are well developed and support a wide variety of analyses. The success of image analysis depends, however, strongly on the quality and reproducibility of the imaged gel patterns. Suggested scanners and image analysis programs are listed in Table 12.

6. Protein identification by mass spectrometry

The identification of the proteins contained 2DE spots is a key step in proteome analysis. It transforms 2DE from a purely descriptive technique into a powerful integrated research tool for the structural and functional analysis of biological processes. Identification is either achieved by *de novo* sequencing or by correlating experimental data obtained from a protein with the contents of sequence (protein, cDNA, genomic DNA, expressed sequence tag (EST)) databases. Different types of data have been used to identify proteins by sequence database searching. Currently, most of these data are being generated by one of a variety of MS techniques (3). The methods have in common that the protein(s) to be identified are first fragmented, typically by tryptic

digestion, and that the peptide fragments rather than the intact protein is subjected to MS. Protein identification is either achieved via the information obtained collectively from a number of peptides derived from the same protein (MS based techniques) or from a single peptide (MS/MS techniques). The identification of proteins using the accurate mass of a number of peptides derived from the same protein is called peptide-mass finger printing or peptide mass mapping (Figures 2 and 5). Software for peptide mass mapping was developed independently by several groups (40-44). The masses for peptide mass mapping are generally determined by MALDI-TOF MS. Peptide mass mapping is a simple and popular method that has gained widespread acceptance. The identification of proteins by the information obtained from a single peptide requires further fragmentation of a tryptic peptide in the collision cell of a tandem mass spectrometer for the generation of sequence specific mass spectra (Figure 3). Software for the correlation of the resulting CID with sequence databases has been pioneered by Eng and Yates (45) and several software tools for that purpose are now available. CID spectra are generated by a variety of tandem mass spectrometers, including triple quadrupole instruments, ion trap mass spectrometers and more recently in quadrupole time-of-flight instruments. Typically these instruments operate with an ESI ion source, although recently tandem mass spectrometers equipped with MALDI ion sources have also been described. Protein identification via MS/MS is now a very mature, powerful and popular technique. It is unique in its ability to identify proteins in samples containing multiple different proteins such as proteins co-migrating to a spot in a 2D gel. Table 13 contains a list of some of the search programs available for protein identification using peptide mass or CID data.

In practice, the protein identification process starts by cutting out individual spots from the 2DE gel. The proteins in the spots are then subjected to in-gel digestion. For smaller scale applications this is typically done manually, for large-scale projects robotic

workstations have been developed. After digestion, peptides are extracted from the gel matrix and contaminants such as residual detergents, salts, excess dye etc. are removed to avoid interference with MS analysis. For MALDI-TOF MS this sample purification is typically done on small reverse phase columns or ZipTips™, small reverse-phase columns packed directly into pipette tips. For ESI-MS/MS peptide purification is frequently achieved by on-line capillary chromatography or on-line capillary electrophoresis.

In the mass spectrometer, the peptides are ionized, separated according to their mass-to-charge ratio (m/z) and then detected. In different types of mass spectrometers the ionization and separation processes are accomplished by different physical principles. All instruments generate mass spectra that are typically represented as a two-dimensional plot displaying the m/z value on the x-axis and relative ion signal intensity (normalized to the most abundant ion in the spectrum) on the y-axis (Figure 5A). In spectra generated in high-resolution instruments each peptide is apparent as an isotope cluster (Figure 5B), a consequence of the fact that 1.1% of the naturally existing carbon is the isotopically heavy ^{13}C. Because peptides contain a high fraction of carbon atoms a significant fraction of the peptide ions will contain one ^{13}C (and therefore a mass value increased by 1mass unit compared to the ^{12}C monoisotopic ion), or two ^{13}C atoms (and therefore a mass value increased by 2 mass units compared to the ^{12}C monoisotopic ion). The x-axis scale of the spectrum is user-defined and usually set between 500-3000 Da, the mass region most commonly scanned for peptide MS. Although many peptides may have a mass of less than 500 Da, peaks related to the matrix (in MALDI-TOF) and small contaminant molecules pollute the spectrum in this mass region.

A detailed description of the mass spectrometers used for peptide fingerprinting or for protein identification via CID spectra of single peptides exceeds the scope of this

chapter. Detailed description of the types of mass spectrometers commonly used for peptide MS, the strengths and limitations of the peptide fingerprinting and MS/MS methods and detailed descriptions of the database search algorithms can be found in recent reviews (3-4, 46). In the following we describe protocols suitable for the isolation of peptides from proteins separated by 2DE and their purification into an MS compatible state.

6.1. In-gel digestion of proteins

Trypsin, which catalyzes specific cleavage C-terminal to the basic amino acids Lys and Arg is the preferred proteolytic agent for MS analysis. Because these amino acids are abundant in proteins most of the resulting peptides will have a mass under 3000 Da. In a typical protocol, the gel slice containing the spot to be analyzed is first dehydrated with acetonitrile and dried to completion in a speedvac. The solution containing trypsin is then added to the dried gel piece so that the trypsin enters the gel matrix during rehydration and digests the protein in the gel. When analyzing sub-picomolar amounts of protein, it is important to perform efficient digestion with a minimal amount of proteolytic enzyme to avoid excessive levels of contamination due to autodigested protease. Peptide extraction is achieved by passive elution of the peptides into solution.

Suggested protocol: tables 14 and 15 list the reagents/equipment and solutions required for the procedures. The method is adapted from Shevchenko, 1996 (32). Use gloves during the whole procedure to minimize keratin contamination:

- **Pick spot:** using a clean razor-blade, cut the exact region of the gel containing the protein to be analyzed, slice it into pieces of approximately 1 mm3 and place the

pieces in a clean 0.5mL eppendorf tube. Note that cutting a region larger than the spot itself will require an extra amount of trypsin solution to rehydrate the gel and thus contribute the level of contamination.

- **Gel dehydration:** wash gel pieces 2 times for 10 minutes each in acetonitrile. The gel will dehydrate and shrink. This helps to remove small contaminant molecules and facilitates subsequent drying of the gel pieces by vacuum centrifugation. Complete dehydration of the gel in a Speedvac vaccum concentrator usually requires approximately 20 minutes;
- **Trypsin digestion:** add sufficient amount of trypsin solution (solution 17) to cover the gel pieces. A volume of 30µL is sufficient for a medium size spot. Let gel rehydrate for 45 minutes on ice and, if necessary add more of solution 17. Seal eppendorf tubes with parafilm to avoid evaporation and incubate them overnight on a 37°C water bath;
- **Peptide extraction:** the next day, quickly spin sample tubes in eppendorf centrifuge and add 30µL of solution 18. After 10 minutes the supernatant is removed and placed into a clean sample tube. The gel pieces are added with 40µL of solution 19 to gel and extracted for 20 minutes. The supernatant is removed again and pooled with the first supernatant. This extraction is repeated two more times and all the supernatants are pooled. Finally, the sample placed in a in a vacuum concentrator to reduce the acetonitrile concentration. To avoid excessive sample loss due to adherence to the plastic vessel, complete drying should be avoided.

6.2. Peptide purification prior to MS

Peptide samples generated by in-gel digestion should be cleaned prior to MS to avoid interference by contaminants also present in the sample. Sample clean-up is particularly

important if small (sub-microgram) amounts of protein are being processed. As peptides are relatively hydrophobic and the elution solvents are compatible with MS, reverse-phase chromatography is a preferred method for sample clean-up. For on-line LC-MS/MS or CE-MS/MS sample clean-up and peptide separation are carried out in the same operation. Due to the complexities of these methods they are not further described here and more specialized reports should be consulted. For MALDI-TOF MS peptide samples are frequently purified by batch reverse-phase chromatography on small columns packed into pipette tips (ZipTip™).

Suggested protocol: tables 16 and 17 list the reagents/equipment and solutions required for performing ZipTip™ sample cleaning:

- **Sample cleaning:** first, ZipTip™ should be conditioned by two washes in 20µL of solution 21 followed by 3 washes in 20µL of solution 20 (use a 20µL micropipette). Quickly (to avoid the C18 material from drying) bind peptides by aspirating and releasing the sample solution 3 times through the ZipTip™ (note that if more then 10% of acetonitrile is present in the sample solution, peptides may not bind) and then wash the immobilized peptide sample 3 times with 20µL of solution 20 to remove unbound contaminants. Finally, elute peptides with the desired volume of solution 21 (minimum of 6µL).

6.3. Identification by peptide-mass finger printing - MALDI-TOF instruments

Protein identification by peptide mass fingerprinting is simple, fast and sensitive and is usually performed on MALDI-TOF mass spectrometers. Many institutions support MALDI-TOF instruments in service facilities or as "walk-up" instruments. In contrast, protein identification via sequence database searching with CID spectra requires more

complex MS/MS instruments, and a description of their operation is beyond the scope of this chapter.

For a protein to be identified by peptide mass finger printing it is critical that the peptide masses are measured at high accuracy (<50ppm). Highly accurate mass measurements minimize constrain the database search and therefore minimize the chance for false positive protein identification and also reduce the number of peptide masses that are required to identify a protein. Modern TOF mass analyzers are able the measure peptide masses at a mass accuracy approaching 10 ppm. MALDI-TOF instruments, as well as the more complex and expensive Q-TOF and TOF-TOF instruments are therefore suitable choices for peptide mass finger printing experiments.

MALDI-TOF instruments are simple and easy to use. First, the sample is mixed with an aromatic matrix compound, usually 2,5-dihydroxy benzoic acid (DHB), α-cyano-4-hydroxybenzoic acid or 3,5 dimethoxy sinapic acid, on a metallic support (MALDI plate). The matrix promotes the formation of peptide containing crystals as it dries and promotes ionization after irradiation of the sample/matrix crystals with a laser pulse. Ionization is achieved in a vacuum chamber by applying short laser pulses to the crystals. Mostly singly charged peptide ions are generated which are then accelerated in an electric field and gated to fly through a field-free, evacuated flight tube. At the end of the flight tube the ions are detected. The mass of the ions is calculated from the amount of time they take to fly through the TOF tube and hit the detector (time-of-flight), the length of the flight tube and the magnitude of the accelerating potential. If the length of the flight tube and the accelerating potential are kept constant, there is a simple relationship between peptide mass and flight time. Once the instrument is calibrated samples can be analyzed quickly and in rapid succession. Up to 100 samples can be spotted in one MALDI plate and analyzed within a few hours.

Some recent advances have greatly improved the resolution and mass accuracy of MALDI-TOF instruments. These include the implementation of delayed extraction and/or reflectrons (ion mirror) both of which improve resolution by correcting the initial spread of kinetic energies, an inherent characteristic of the MALDI ionization process (4). Some instruments are also able to perform fragmentation of selected peptides by a process known as post-source decay (PSD) to generate additional structural information. Figure 5A shows the mass spectrum obtained from a peptide mass finger printing experiment.

Suggested protocol: tables 18 and 19 list the reagents, equipment and solutions required for MALDI-TOF mass mapping. Protein identification by peptide mass mapping relies on the availability of a suitable sequence database. Access to such a database either locally or through the Web is therefore a precondition for successful peptide mass fingerprinting experiments:

- **Sample spotting:** there are many ways for spotting a sample onto the sample plate. We suggest to spot 0.5µL of matrix solution (solution 22) directly on MALDI plate and then overlay it with 0.5µL of the ZipTip™ cleaned sample solution. If working with sub-picomole quantities of sample, we suggest to concentrate sample solution to approximately 1µL, so half of the sample can be applied in a single spot. Proper crystallization is extremely important for generating good quality data. If sample crystals look too different from the crystals formed by pure matrix, it probably means that there are contaminants in the sample solution that affect matrix crystallization;
- **Instrument adjustment:** first, using peptide digest mixtures from standard proteins, determine the laser power that gives highest intensity (for the expected peptide peaks) with least amount of background noise. Use the mass of expected peptides to mass calibrate the instrument;

- **Data acquisition:** acquire data in the range of 500 to 3000 Da and choose to accumulate data, so the spectra from each shot is added up. Start shooting sample, preferentially the edges of the crystals, where it tends to give more signal. Shoot as many laser pulses as necessary to obtain a good quality spectrum, but note that sample is being partially consumed in each shot. As the sample is being consumed higher noise levels are apparent. In the generated mass list, eliminate the masses of known trypsin and keratin peptides prior to submission of the data to the database search algorithm;
- **Database search:** submit the peptide mass list for database search using any of the appropriated programs listed in Table 13. To constrain the database search select the approximate range of pI and MW of the parent protein being analyzed in the program parameter menu. The values are apparent from the protein spot position in the 2D gel. If the obtained matching score is too low to be considered conclusive we suggest to submit the candidate sequences to PeptideMass (http://ca.expasy.org/tools/peptide-mass), a proteomic tool that cleaves protein sequences with a chosen enzyme, and computes the masses of the generated peptides. It can also return the mass of peptides known to carry posttranslational modifications, and highlight peptides the masses of which may be affected by database conflicts, isoforms or splicing variants. Comparison of these predicted information with the crude mass spectrum is usually useful to select the right candidate and validate the identification (47).

7. Concluding remarks

We have described the steps required for proteome analysis using 2DE and MS. From sample preparation to protein identification, the suggested protocols are proven but may

have to be optimized for specific cases. There are other emerging techniques, such as those based on stable isotope labeling, that may complement the 2DE approach and provide accurate quantitative data. The next few years are anticipated to see further acceleration of proteomics technology and applications.

References

(1) Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik **26**, 231–243.

(2) O'Farrel, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007–4021.

(3) Aebersold, R. and Goodlett, D. (2001) Mass spectrometry in Proteomics. *Chemical Reviews* **101**, 269–295.

(4) Corthals, G. L., Patterson, S. D. and Aebersold, R. (1999) Identification of proteins by mass spectrometry, in *Proteome Research: 2D Gel Electrophoresis and Detection Methods* (Rabilloud, T., ed.), Springer, New York, 1999, pp. 197–231.

(5) Aebersold R. H., Leavitt J., Saavedra R. A., Hood L. E. and Kent S. B. (1987) Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc Natl Acad Sci U S A* **84**, 6970–6974.

(6) Wilkins M. R., Pasquali C., Appel R. D., Ou K., Golaz O., Sanchez J. C., Yan J. X., Gooley A. A., Hughes G., Humphery-Smith I., Williams K. L. and Hochstrasser D. F. (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology* **14**, 61–65.

(7) Urquhart, B. L., Cordwell, S. J. and Humphery-Smith, I. (1998) Comparison of predicted and observed properties of proteins encoded in the genome of Mycobacterium tuberculosis H37Rv. *Biochem. Biophys. Res. Commun.* **253**, 70–79.

(8) Gygi S. P., Corthals G. L., Zhang Y., Rochon Y. and Aebersold R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* **15**, 9390–9395.

(9) Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951) Protein measurement with folin phenol reagent. *J. Biol. Chem.* **193**, 265–275.

(10) Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76–85.

(11) Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.

(12) Dignam, J. D. (1990) Preparation of extracts from higher eukaryotes. *Methods Enzymol.* **182**, 194–203.

(13) Gegenheimer, P. (1990) Preparation of extracts from plants. *Methods Enzymol* **182**, 174–193.

(14) Jazwinski, S. M. (1990) Preparation of extracts from yeast. *Methods Enzymol* **182**, 154–174.

(15) Cull, M. and McHenry, C. S. (1990) Preparation of extracts from prokaryotes. *Methods Enzymol* 182, 147–153.

(16) Rabilloud, T. (1996) Solubilization of proteins for electrophoretic analyses. *Electrophoresis* 17, 813–829.

(17) Englard, S. and Seifter, S. (1990) Precipitation techniques. *Methods Enzymol* 182, 285–300.

(18) Rabilloud, T. (1999) Solubilization of proteins in 2-D electrophoresis. An outline. *Methods Mol Biol* 112, 9–19.

(19) Rabilloud, T. (1998) Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. *Electrophoresis* 19, 758–760.

(20) Wilkins, M. R., Gasteiger, E., Sanchez, J. C., Bairoch A. and Hochstrasser D. F. (1998) Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number. *Electrophoresis* 19,1501–1505.

(21) Zewert, T. E. and Harrington, M. G. (1992) Polyethyleneglycol methacrylate 200 as an electrophoresis matrix in hydroorganic solvents. *Electrophoresis* 13, 824–831.

(22) Zewert, T. E. and Harrington, M. G. (1993) Protein electrophoresis. *Curr Opin Biotechnol* 4, 3–8.

(23) Bjellkvist, B., Ek, K., Righetti, P. G., Gianazza, E., Görg, A., Westermeier, R. and Postel, W. (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J. Biochem Biophys Methods* **6**, 317–339.

(24) Gorg, A., Postel W. and Gunther S. (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **9**, 531–546.

(25) Sanchez, J. C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R. and Hochstrasser D. F. (1997) Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients. *Electrophoresis* **18**, 324–327.

(26) Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

(27) Patton, W. F. (2000) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* **21**, 1123–1144.

(28) Patton, W. F. (2001) Detecting proteins in polyacrylamide gels and electroblot membranes, in *Proteomics: from protein sequence to function* (Pennington, S. R. and Dunn, M. J., eds.), Bios, Oxford, pp. 31-59.

(29) Rabilloud, T. (1999) Silver staining of 2-D electrophoresis gels. *Methods Mol. Biol.* **112**, 297--305.

(30) Rabilloud, T. (1992) A comparison between low background silver diammine and silver nitrate protein stains. *Electrophoresis* **13**, 429--439.

(31) Rabilloud, T. (1990) Mechanisms of protein silver staining in polyacrylamide gels: a 10-year synthesis. *Electrophoresis* **11**, 785--794.

(32) Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850--858.

(33) Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T. and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466--469.

(34) Yan, J. X., Wait, R., Berkelman, T., Harry, R. A., Westbrook, J. A., Wheeler, C. H. and Dunn, M. J. (2000) A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry. *Electrophoresis* **21**, 3666--3672.

(35) Dunn, M. J. (1987) Two-dimensional polyacrylamide gel elctrophoresis, in *Advances in Electrophoresis* (Chrambach, A., Dunn, M. J. and Radola, B. J., eds.), Vol. 1, VCH, Weinheim, pp. 1-109.

(36) Lopez, M. F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M. and Patton, W. F. (2000) A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* **21**, 3673--3683.

(37) Steinberg, T. H., Jones, L. J., Haugland, R. P. and Singer, V. L. (1996) SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Anal Biochem.* **239**, 223--237.

(38) Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young J., Pognan, F., Hawkins, E., Currie, I. and Davison M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **1**, 377--396.

(39) Unlu, M., Morgan, M. E. and Minden, J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071--2077.

(40) Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C. and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A.* **90**, 5011-5015.

(41) Pappin, D. J. (1997) Peptide mass fingerprinting using MALDI-TOF mass spectrometry. *Methods Mol Biol.* **64**, 165--173.

(42) Mann, M., Hojrup, P. and Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* **22**, 338–345.

(43) Yates, J. R. 3rd, Speicher, S., Griffin, P. R. and Hunkapiller T. (1993) Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* **214**, 397–408.

(44) James, P., Quadroni, M., Carafoli, E. and Gonnet G. (1993) Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun.* **195**, 58–64.

(45) Yates, J. R. 3rd, Eng, J. K., McCormack, A. L. and Schieltz D. (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67**, 1426–1436.

(46) Patterson, S. D. and Aebersold, A. (1995) Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* **16**, 1791—1814.

(47) Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J. C., Hochstrasser, D. F. and Appel, R.D. (1997) Detailed peptide characterisation using PEPTIDEMASS - a World-Wide Web accessible tool. *Electrophoresis* **18**, 403–408.

Table 1. Disruption methods.

Type of sample	**Disruption method**
Hard tissues (e. g. muscle, liver, plant seeds)	Mechanical homogenization (12,13).
Tissue culture and blood cells	Sonication, detergent lysis or osmotic lysis (12).
Yeast cells	Lyticase treatment and detergent lysis or glass beads (14).
Bacterial cells	Lysozyme treatment and/or detergent lysis (15).

Table 2. Reagents for sample preparation.

Reagent	Comments	Manufacturer
CHAPS	Zwitterionic detergent effective for protein solubilization.	Sigma, Bio-Rad and APB*
DTT	Reducing agent. Store at -20°C in desicator. Rapidly oxidized in contact with air, always make fresh solutions from desicator stored powder.	Sigma, Bio-Rad and APB
Carrier ampholytes	Buffering molecules that improve separation and sample solubility.	APB
Ultrapure urea	Solubilizes and denatures proteins by disrupting hydrogen bonds. Impure urea contains cyanate ions that may cause protein carbamylation, thus changing the pI of proteins. Only highest purity urea should be used and solutions should be made fresh.	APB
Bromophenol blue	Dye used to monitor the beginning of IEF run and to be sure that current is passing thorugh the strip.	Sigma, Bio-Rad and APB
SDS	Anionic detergent used for protein solubilization.	Sigma and Bio-Rad
PMSF	Inhibitor of serine protease. Very toxic and unstable in water. Prepare a 100mM stock solution in propanol and keep in -20°C.	Sigma
EDTA	Inhibitor of metalloproteases. Prepare a 100mM stock solution and keep at room temperature.	Sigma

*Amersham Pharmacia Biotech

Table 3. Materials required for performing IEF.

Reagent / Equipment	Comments	Manufacturer
Immobilized pH gradient (IPG) strips	Polyacrylamide matrix containing co-polymerized buffering groups that make-up the pH gradient. We recommend to use 18cm strips.	APB
Mineral Oil	Inert oil used to cover the IPG strips during rehydration and electrophoresis. It avoids drying of the strip, urea crystalization and difusion of CO_2 that may react with the proteins.	APB (sold as IPG cover fluid)
IEF electrophoretic unit (IPGphor)	This instrument is capable of reaching 8000V and has a Peltier system for improved temperature control. Rehydration and IEF run are performed sequentially without user's intervention.	APB

Table 4. Running conditions for analytical IEF. Temperature should be set to 20°C and the maximum current to 50μA per strip.

Voltage (V)	Duration (hours)
100	1
500	1
4000	5
8000	Until it reaches 80kVh

Table 5. Materials for second dimension.

Reagent / Equipment	Comments	Manufacturer
Acrylamide : Bis-acrylamide (30%:0.8%) solution	Buy it as ready solution to avoid contact with acrylamide powder that is potentialy more hazardous.	Bio-Rad
Tris	Buffering component.	Bio-Rad
Ammonium Persulfate (APS)	Polymerization catalyst. Store in dessicator.	Bio-Rad and Sigma
N,N,N',N'-tetramethylethylenediamine (TEMED)	Polymerization catalyst.	Sigma, Bio-Rad and APB
Agarose	Use low electroendosmosis and high resistance agarose.	Bio-Rad, Gibco and APB
Vertical electrophoresis unit system	We suggest the use of 14X16cm plates and 1.5cm thickness spacers.	APB (model SE600)
Temperature controlled water circulator (optional)	Allows internal temperature control of electrophoretic run.	APB (Multitemp)
Power Supply	It sohuld be able to generate until 600 Volts and 200mA.	APB and Bio-Rad

Table 6. Solutions for second dimension.

Solution	Preparation / comments
(1) Acrylamide : Bis-acrylamide	Ready to use solution. Be careful, acrylamide is neurotoxic, do not mouth pipette solutions.
(2) 1M Tris pH 8.8	Adjust pH with concentrated HCl. For storage, keep at 4°C for 1month at maximum.
(3) APS 10%	Make fresh right before polymerization.
(4) Temed	Ready to use
(5) Running buffer (10X concentrated) -250mM Tris, 192mM Glycine and 1% SDS	Do not adjust pH, it should already be around 8.4. Dilute 1:10 prior to use. For storage, keep at 4°C for 1 month at maximum.
(6) Equilibrating solution 1 -50mM Tris-HCl (pH 8.8), 6M urea, 30% glycerol, 2% SDS and 2%DTT	Prepare stock solution without urea and DTT, and add them right before use. Reduces dissulfide bonds and coats IEF separated proteins in SDS.
(7) Equilibrating solution 2 -50mM Tris-HCl (pH 8.8), 6M urea, 30% glycerol, 2% SDS and 2.5% iodoacetamide	Prepare stock solution without urea and iodoacetamide, add them right before use. Alkylates cysteines and coats protein with SDS.
(8) Agarose solution -0.1% agarose in 10X diluted running buffer solution.	Melt agarose and keep at -20°C in 10mL aliquots.

Table 7. Recipe for preparing two 12.5% acrylamide gels (14 X 16 X 0.15cm).

Solution	Volume
Solution 1	33,2mL
Solution 2	29,8mL
Distilled water	17mL

For polymerization, add 400µL of solution 3 and 10µL of solution 4.

Table 8. Materials for CBB staining.

Reagent / Equipment	Manufacturer
Coomassie Brilliant Blue R-250	Sigma
Acetic acid	local supplier
Methanol	local supplier
Plastic or glass tray, 20 X 20 X 5cm	local supplier
Orbital shaker	local supplier

Table 9. Solutions for CBB staining.

Solution	Preparation / comments
(9) Staining solution: - 0.1% CBB R250 in 50% methanol and 10% acetic acid	Dissolve CBB in pure methanol and stir for 10 minutes. Add distilled water and, at last, acetic acid. CBB is not properly solubilzed under extreme acidic condition so acetic acid should be added only after dye is completely dissolved.
(10) Destaining solution: - 50% methanol and 10% acetic acid	For re-utilization, clear solution by passing it through a funnel containing activated charcoal.
(11) 10X diluted destaining solution: - 5% methanol and 1% acetic acid	Dilute 10 times solution 10. Used to clear background. Because methanol dehydrates and shrinks the gels, CBB is more difficult to remove from shrunk gel. We propose to soak gels in low methanol containing solution to quickly rehydrate gels and thus to facilitate CBB removal.

Table 10. Materials for silver staining.

Reagent / Equipment	Manufacturer
Silver nitrate - $AgNO_3$	Sigma
Acetic acid	local supplier
Sodium thiosulphate - $Na_2S_2O_3$	Merck
Formaldehyde - HCOH	Merck
Methanol	local supplier
Glass tray, 20 X 20 X 5cm	local supplier
Orbital shaker	local supplier

Table 11. Solutions for silver staining.

Solution	Preparation / comments
(12) Fixing solution: - 50% methanol and 10% acetic acid.	
(13) Sodium thiosulphate solution: - 0.2g $Na_2S_2O_3$/L	Make fresh.
(14) Silver solution: - 0.2g $AgNO_3$/100mL	Make fresh.
(15) Developing solution: - 3 g Na_2CO_3, 50µL HCOH and 2mL of solution 13/100mL	Solution 13 must be freshly made and HCOH is added just prior to use.
(16) Stopping solution: - 10% acetic acid	Developing reaction is stopped by lowering the pH.

Table 12. Materials for image acquisition and analysis.

Reagent / Equipment	Comments	Manufacturer
Scanner	Although it is possible to use common desktop scanners, specialized laser scanners offer higher resolution and may provide more reliable densitometric analysis and also allow acquisition of fluorescence patterns.	APB (Storm™)
Software for image analysis	Detailed manuals are provided by manufacturer.	APB (Image Master 2D Elite™) and Bio-Rad (Melanie™)

Table 13. Database search programs for protein identification using MS data.

Programs	Internet address
Peptide mass data search	
MS-Fit (Prospector)	prospector.ucsf.edu/ucsfhtml3.4/msfit.htm
MOWSE	www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse/
PeptideSearch	www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage
Profound	prowl.rockefeller.edu/cgi-bin/ProFound
MultiIdent	ca.expasy.org/tools/multiident/multiident2.html
Mascot	www.matrixscience.com/cgi/index.pl?page=../home.html
CID data search	
SEQUEST	thompson.mbt.washington.edu/sequest/
MS-Tag (Prospector)	prospector.ucsf.edu/ucsfhtml3.4/mstagfd.htm
PeptideSearch	www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage
Mascot	www.matrixscience.com/cgi/index.pl?page=../home.html

Table 14. Materials for in-gel digestion.

Reagent / Equipment	Comments	Manufacturer
Acetonitrile – CH_3CN		Pierce
Ammonium Bicarbonate – NH_4HCO_3	Volatile buffer used to keep proper pH (around 8-9) for trypsin activity.	Sigma
Formic acid		Sigma
Sequencing grade modified trypsin	This trypsin is alkylated to reduce autolysis.	Promega

Table 15. Solutions for in-gel digestion.

Solution	Preparation / comments
(17) Trypsin solution: - 12.5ng trypsin/µL of 50mM NH_4HCO_3	A concentrated stock is prepared in acidic solution to minimize autolysis during storage. Right before usage, dilute sample in the NH_4HCO_3 buffer. Using pH strips, make sure the pH is basic after dilution.
(18) Extraction buffer 1: - 20mM NH_4HCO_3	Store at room temperature.
(19) Extraction buffer 2: - 5% formic acid and 50% CH_3CN	Acetonitrile shrinks gel piece, helping to extract peptides.

Table 16. Materials for sample cleaning.

Reagent / Equipment	Comments	Manufacturer
Acetonitrile – CH_3CN		Pierce
Trifluoracetic acid (TFA) – $C_2HF_3O_2$		Sigma
C18 ZipTip™	C18 reverse phase material is more appropriate for peptide samples than C4.	Millipore

Table 17. Solutions for ZipTip™ sample cleaning.

Solution	Preparation / comments
(20) Equilibrating solution: - 0.1% TFA	TFA is higly corrosive, handle it carefully in the hood, wearing proteoctive clothing.
(21) Eluting solution: - 0.1% TFA and 60% CH_3CN	TFA is higly corrosive, handle it carefully in the hood, wearing protective clothing.

Table 18. Materials for protein identification by peptide-mass finger printing.

Reagent / Equipment	Comments	Manufacturer
2,5-dihydroxy benzoic acid (DHB)		Sigma
MALDI-TOF mass spectrometer		Applied Biosystems (Voyoger DE-Pro)
Acetonitrile		Pierce

Table 19. Solutions for protein identification by peptide-mass finger printing.

Solution	Preparation / comments
(22) Matrix solution: - 160mg DHB/1mL of 25% CH_3CN	Make it fresh.

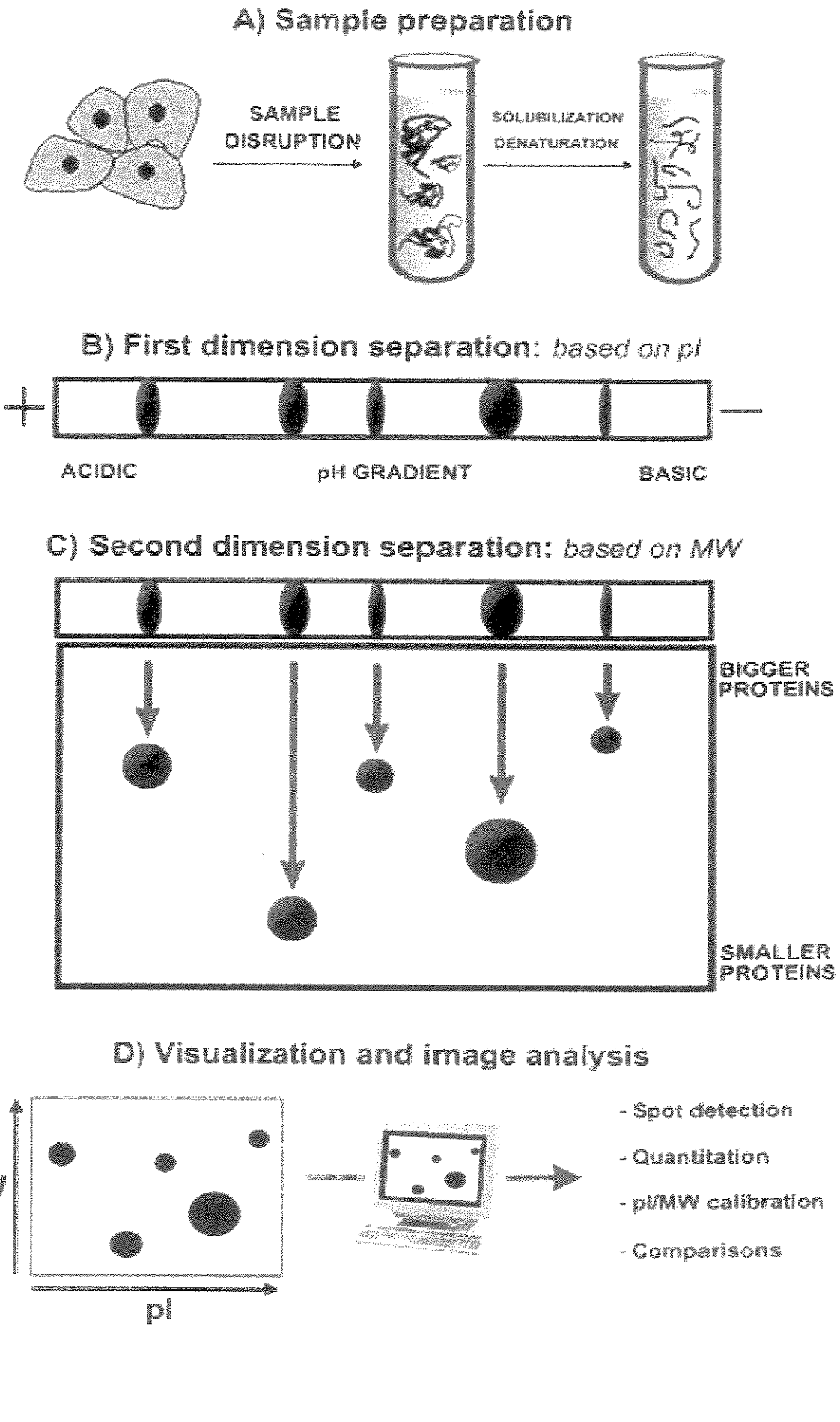

A) Sample preparation
SAMPLE DISRUPTION
SOLUBILIZATION DENATURATION
B) First dimension separation: based on pI
+
—
ACIDIC
pH GRADIENT
BASIC
C) Second dimension separation: based on MW
BIGGER PROTEINS
SMALLER PROTEINS
D) Visualization and image analysis
MW
pI
- Spot detection
- Quantitation
- pI/MW calibration
- Comparisons

FIGURE 1: Steps of the 2DE procedure. **(A)** Sample preparation; proteins are first extracted from cells or tissues and solubilized. **(B)** First dimension separation is achieved by IEF, in which proteins are separated according to their isoelectric point. **(C)** Second dimension separation is achieved by SDS-PAGE, that separates proteins according to their molecular weight. **(D)** The separated proteins are then stained directly in the gel resulting in a two-dimensional array of spots, each one usually containing multiple copies of a homogeneous protein. The amount of protein migrating to each spot is quantified based on the staining intensity of the respective spot. Usually the gel image is digitized and densitometrically analyzed using computer programs.

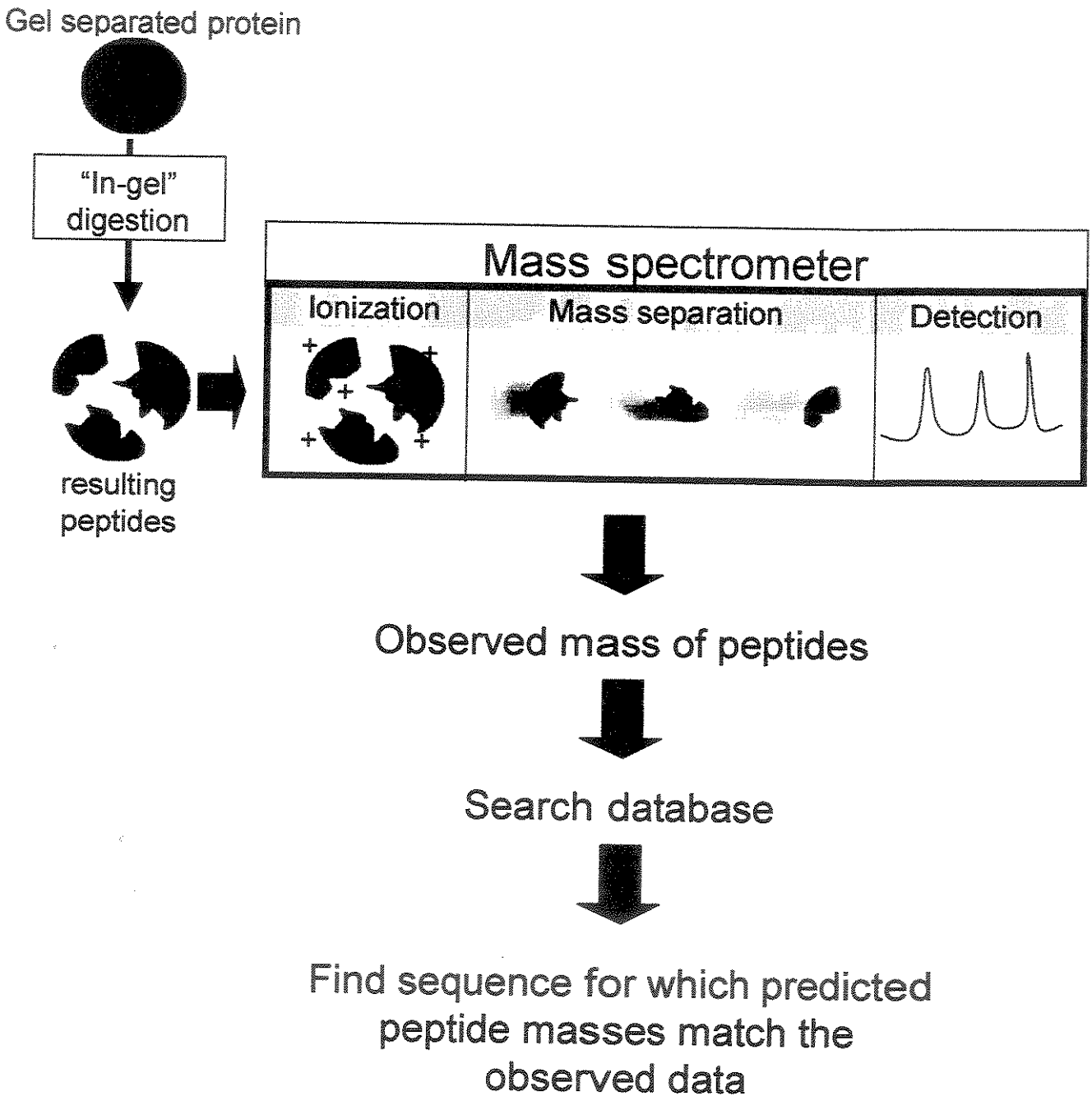

Gel separated protein
"In-gel" digestion
resulting peptides
Mass spectrometer
Ionization
Mass separation
Detection
Observed mass of peptides
Search database
Find sequence for which predicted peptide masses match the observed data

FIGURE 2: Protein identification by peptide mass fingerprinting. 2DE separated proteins are in gel digested with trypsin, the masses of resulting peptides are determined (commonly by MALDI-TOF MS) and used to identify protein by sequence database searching.

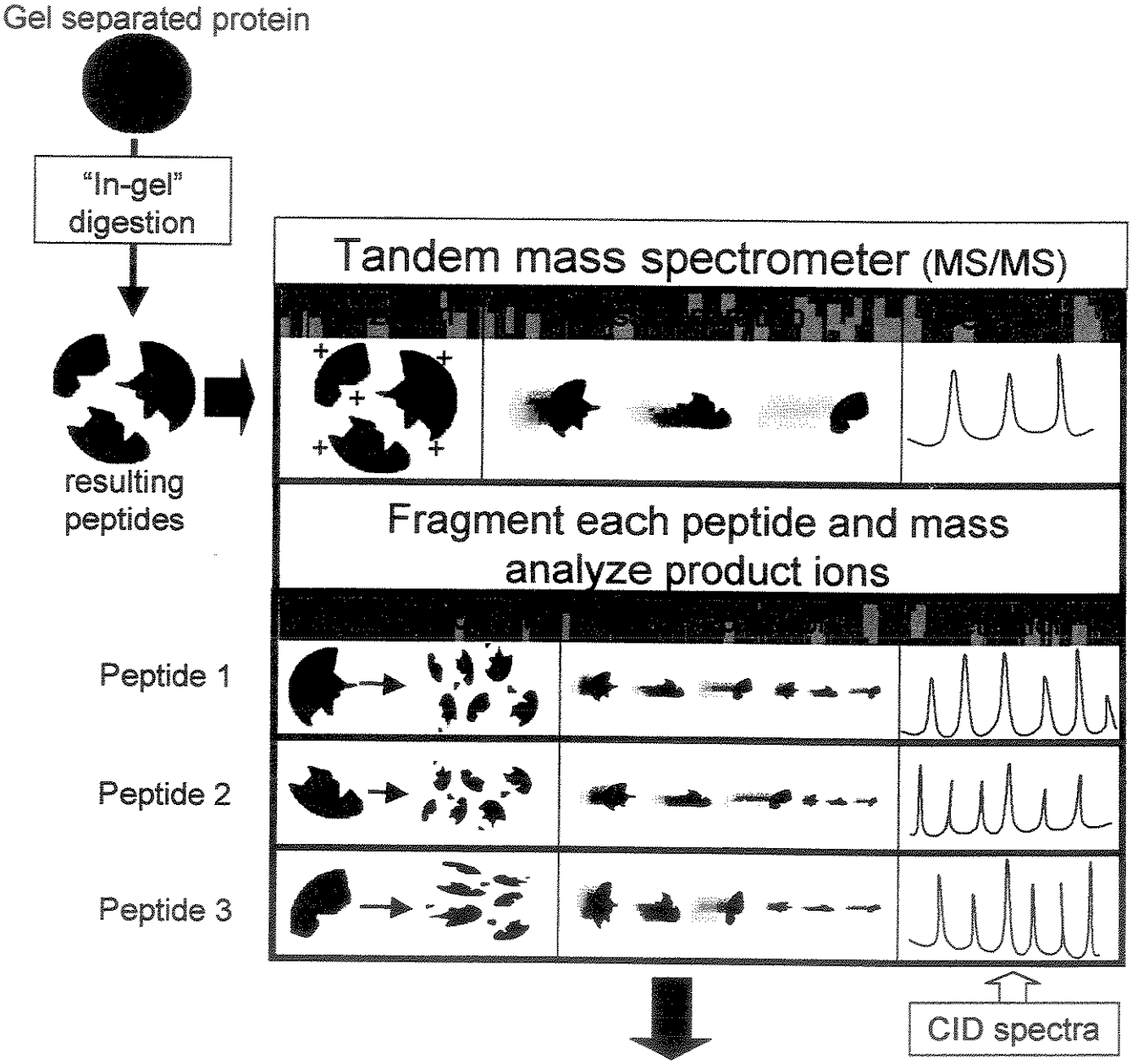

Gel separated protein
"In-gel" digestion
resulting peptides
Tandem mass spectrometer (MS/MS)
Fragment each peptide and mass analyze product ions
Peptide 1
Peptide 2
Peptide 3
CID spectra
Observed fragment ion masses from peptides 1, 2 and 3
Search database (each peptide separately)
Find peptide sequences for which predicted fragment ion masses match the observed data

FIGURE 3: Protein identification using tandem mass spectrometry. 2DE separated proteins are digested with trypsin and resulting peptides are mass analyzed and further fragmented in the collision cell of a tandem mass spectrometer for the generation of sequence specific fragment ion mass spectra (CID spectra). In principle, CID spectrum of one peptide is usually sufficient for protein identification.

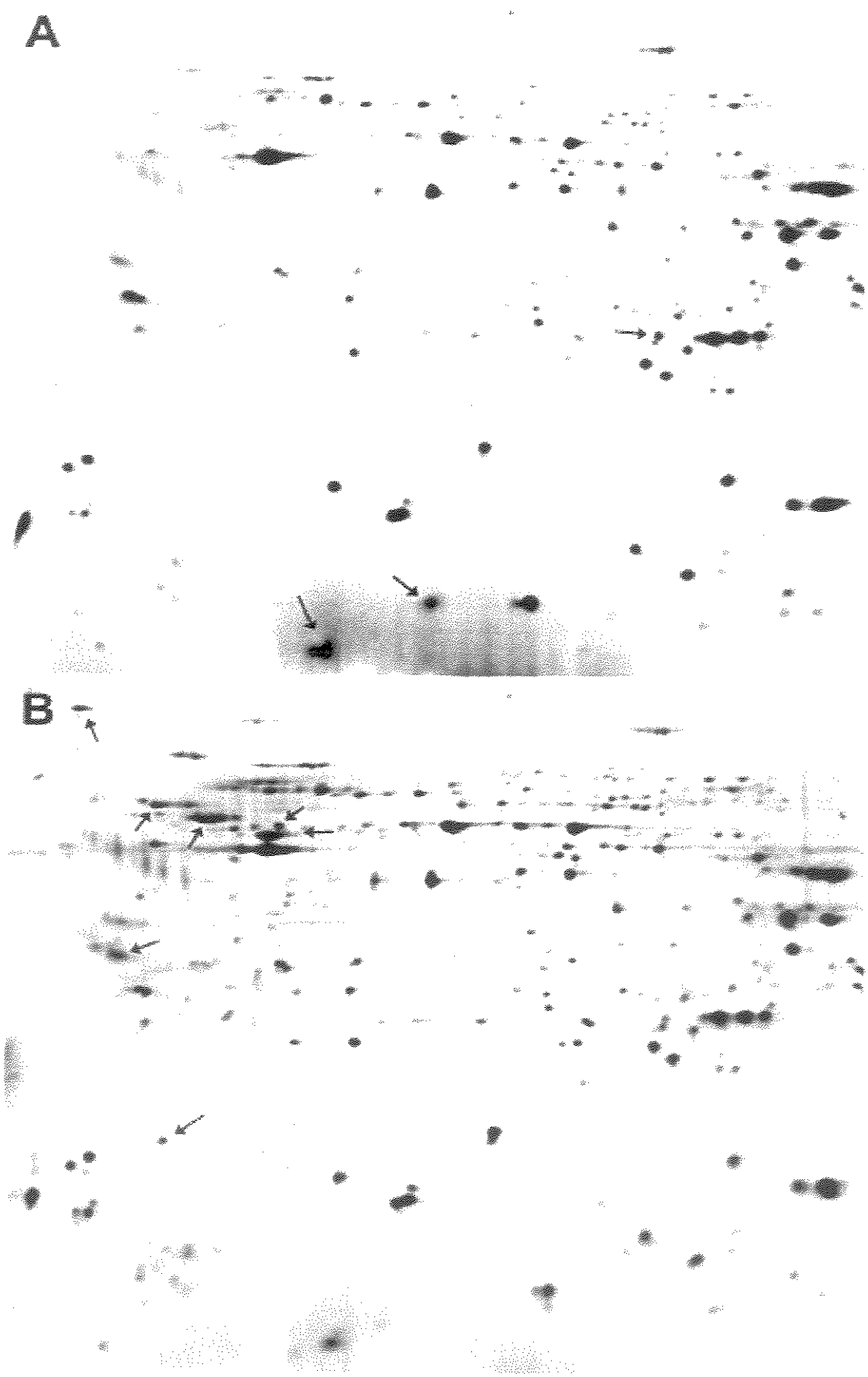
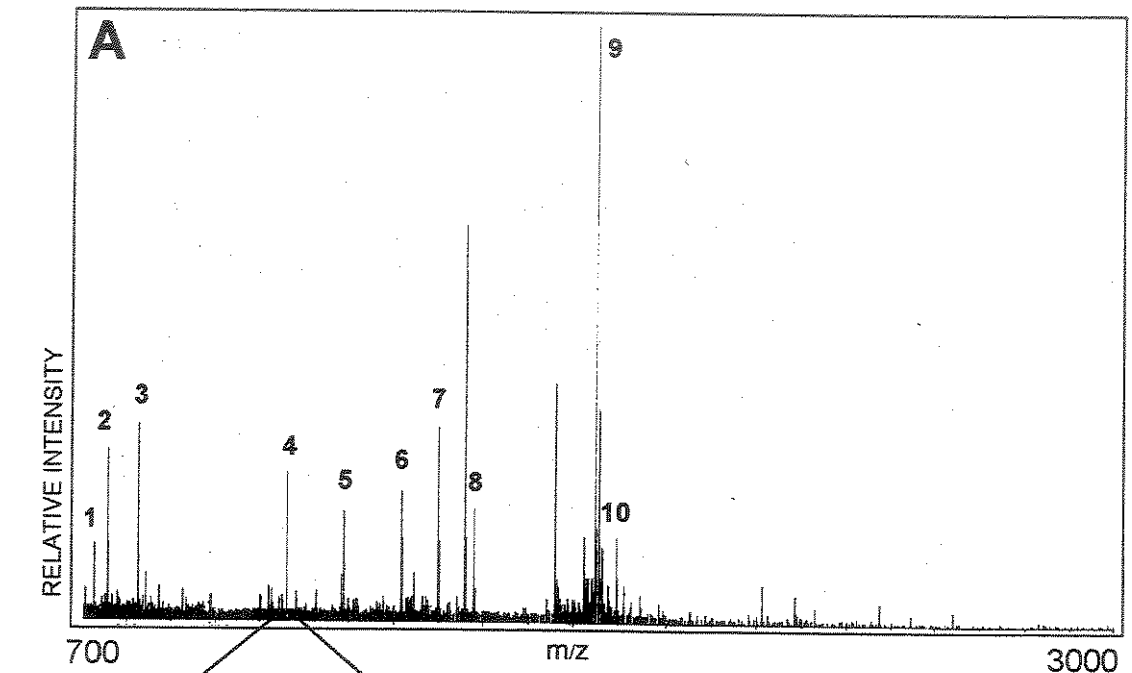
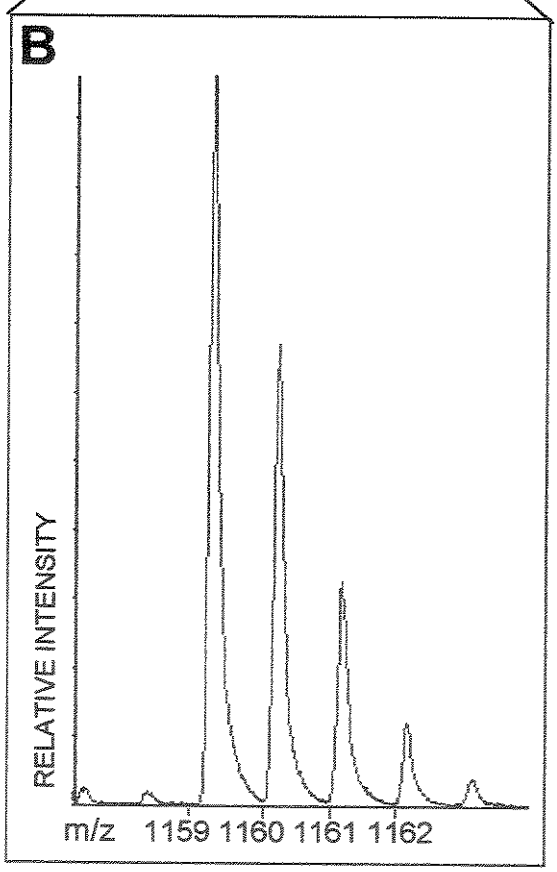

A
B

FIGURE 4: Silver stained 2DE of total protein extract from unfertilized (A) and fertilized (B) *Xenopus* egg. Arrows indicate some of the differentially expressed proteins.





C

AVSKVYARSVYDSRGNPTVEVELTTEKG

VFRSIVPSGASTGVHEALEMRDEDKSK

WMGKGVMNAVNNVNNVIAAAFVKANLD

VKDQKAVDDFLLSLDGTANKSKLGANAI

LGVSMAAARAAAAEKNVPLYQHLADLSK

SKTSPYVLPVPFLNVLNGGSHAGGALAL

QEFMIAPTGAKTFAEAMRIGSEVYHNLK

SLTKKRYGASAGNVGDEGGVAPNIQTA

EEALDLIVDAIKAAGHDGKVKIGLDCASS

EFFKDGKYDLDFKNPESDKSKWLTGVE

LADMYHSLMKRYPIVSIEDPFAEDDWEA

WSHFFKTAGIQIVADDLTVTNPARIATAIE

KKAADALLLKVNQIGTLSESIKAAQDSFA

ANWGVMVSHRSGETEDTFIADLVVGLR

TGQIKTGAPARSERLAKLNQLLRIEEELG

DKAVYAGENFHHGDKL

FIGURE 5: Peptide mass map of yeast enolase 2. Total yeast extract was separated by 2DE and silver stained. A selected spot was trypsin digested and the peptide masses analyzed in a MALDI-TOF mass spectrometer. Database search using the observed peptide masses indicated that protein was enolase 2. **(A)** Mass spectrum of tryptic peptides. The numbers indicate the corresponding peptide in Figure 5C. **(B)** Amplification of the region between 1158 and 1164Da showing the isotope cluster of an enolase 2 peptide with monoisotopic mass of 1159.66Da. **(C)** Amino acid sequence of yeast enolase 2. The number indicate the corresponding peaks in the mass spectrum. Note that trypsin cleaves on the C-terminal of lysine (L) or arginine (R).

Produção Científica

1. Artigos publicados

Smolka, M. B., Martins, D., Winck, F. V., Santoro, C. E., Castellari, R. R., Ferrari, F., Brum, I. J., Galembeck, E., Coletta-Filho, H. D., Machado, M. A., Marangoni, S. e Novello, J. C.. Proteome Analysis of the Plant Pathogen *Xylella fastidiosa* Reveals Major Cellular and Extracellular Proteins and a Peculiar Codon Bias Distribution. *Proteomics (Aceito para publicação)* 2002.

Smolka, M., Zhou, H. e Aebersold, R. Quantitative Protein Profiling Using Two-dimensional Gel Electrophoresis, Isotope-coded Affinity Tag Labeling, and Mass Spectrometry. *Mol Cell Proteomics* 2002, 1, 19-29

Smolka, M. B., Zhou, H., Purkayastha, S. e Aebersold, R. Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal Biochem* 2001, 297, 25-31.

Não relacionados ao projeto:

Freire, M. G., Machado, O. L., Smolka, M. B., Marangoni, S., Novello, J. C. e Macedo, M. L. Isolation and characterization of isolectins from Talisia esculenta seeds. *J Protein Chem* 2001,20, 495-500.

Smolka, M. B., Zoppi, C. C., Alves, A. A., Silveira, L. R., Marangoni, S., Pereira-Da-Silva, L., Novello, J. C. e Macedo, D. V. HSP72 as a complementary protection against oxidative stress induced by exercise in the soleus muscle of rats. *Am J Physiol Regul Integr Comp Physiol.* 2000, 279, R1539-1545.

Smolka, M. B, Marangoni, S., Oliveira, B. e Novello J. C. Purification and partial characterization of a thrombin-like enzyme, balterobin, from the venom of Bothrops alternatus. *Toxicon* 1998, 36, 1059-1063.

2. Apresentações orais

"Proteome Analysis of the Plant Pathogen *Xylela fastidiosa* Strain 9a5c: Secreted and Adhesion Proteins", XXXI Reunião Anual da Sociedade Brasileira de Bioquímica - SBBq. Caxambu, MG, 21 de maio de 2002.

"O Papel Atual de Eletroforese de Duas Dimensões na Análise Proteômica", Ciclo de Palestra da Amersham Biosciences, XXXI Reunião Anual da Sociedade Brasileira de Bioquímica - SBBq. Caxambu, MG, 21 de maio de 2002.

"Projeto Proteoma da *Xylella* fastidiosa", Reunião de Soluções para Proteoma, Applied Biosystems. São Paulo, SP, 6 de dezembro de 2001.

"Projeto Proteoma da *Xylella* fastidiosa", I Simpósio Genoma Funcional da *Xylella fastidiosa* - FAPESP. Serra Negra, SP, 11 de dezembro de 2001.

"Proteoma de *Xyllela fastidiosa*", I WORKSHOP SOBRE PROTEÔMICA PARASITÁRIA. RECIFE, PE, 23 de Novembro de 2001.

"Quantitative Proteomics: using mass spectrometry and Icat reagents for large scale quantification and identification of proteins", Ciclo de Palestras da Applied Biosystems, XXX Reunião Anual da Sociedade Brasileira de Bioquímica - SBBq. Caxambu, MG, 21 de maio de 2001.

"Análise de Proteomas", Programa de Seminários do Dept. de Bioquímica - IB - UNICAMP. Campinas, SP, 8 de outubro de 1999.

3. Cursos ministrados

Mini-curso: "Biotecnologia", V Congresso Aberto aos Estudantes de Biologia -CAEB. UNICAMP, Campinas, SP, 8 a 12 de outubro de 2001.

Mini-curso: "Análise de Proteoma na Fisiologia Vegetal", VIII Congresso Brasileiro de Fisiologia Vegetal. Ilhéus, BA, 2 a 7 de setembro de 2001.

"I Curso Teórico-Prático de Estudos Pós-Genômicos em Parasitas". Recife, PE, 19 a 24 de novembro de 2001.

4. Cursos realizados

"Proteoma e Eletroforese de Duas Dimensões em Gel de Poliacrilamida". Genebra, Suíça, junho de 1998.

"Análise de Imagens de Géis 2DE". Freiburg, Alemanha, julho de 1998.

"Computação Aplicada à Biologia". Petrópolis, RJ, fevereiro de 1999.

"Espectrometria de Massa em Aparelho MALDI-TOF". Framingham, EUA, setembro de 2001.

5. Participação em disciplinas

Aula: "Projetos Proteoma", Disciplina de Genética Molecular para Ciências Exatas, UNICAMP. Campinas, SP, 16 de maio de 2001. Responsável: Prof. Dr. Marcelo Menossi.

Aula: "Eletroforese de Proteínas", Disciplina de Bioquímica de Macromoléculas (BB-582), Curso de Ciências Biológicas, UNICAMP. Campinas, SP, 4 de abril de 2000. Responsável: Prof. Dr. Hiroshi Aoyama.

Aula: "Eletroforese e Proteoma", Disciplina de Bioquímica de Macromoléculas (BB-582), Curso de Ciências Biológicas, UNICAMP. Campinas, SP, 31 de março de 1999. Responsável: Prof. Dr. Hiroshi Aoyama.

Aula: "Adaptações Musculares: um Visão Bioquímica", Disciplina de Estudo dos Movimentos Corporais, Faculdade de Educação Física, UNICAMP. Campinas, 23 de abril de 1999. Responsável: Profa. Dra. Antonia Bankoff.

Aula: "Eletroforese de Proteínas", Disciplina de Bioquímica Médica (BB-680), Curso de Ciências Biológicas, UNICAMP. Campinas, SP, 27 de setembro de 1998. Responsável: Prof. Dr. José Camillo Novello.

6. Comunicações em congressos

SMOLKA, M. B., WINCK, F. V., MARANGONI, S., SANTORO, C. E., MARTINS, D., COLETTA FILHO, H. D., MACHADO, M. A., LEMOS, E., BRUM, I. J., GALEMBECK, E., TOYAMA, M., NOVELLO, J. C. Proteome analysis of the plant pathogen Xylella fastidiosa strain 9A5C: secreted and adhesion proteins In: XXXI Reunião Anual - SBBq, 2002, Caxambu.

MARTINS, D., WINCK, F. V., SANTORO, C. E., MARANGONI, S., NOVELLO, J. C. **SMOLKA, M. B.** Xylella fastidiosa is a peculiar organism in terms of gene codon bias In: XXXI Reunião Anual - SBBq, 2002, Caxambu.

BRUM, I. J., **SMOLKA, M. B.**, NOVELLO, J. C., GALEMBECK, E. Theoretical avaliation of a new protein identification method based on multiple database constrains In: XXXI Reunião Anual - SBBq, 2002.

SANTORO, C. E., **SMOLKA, M. B.**, WINCK, F. V., MARTINS, D., CASTELLARI, R. R., FERRARI, F., MARANGONI, S., NOVELLO, J. C. Identification of basic proteins from Xylella fastidiosa In: XXXI Reunião Anual - SBBq, 2002, Caxambu.

WINCK, F. V., **SMOLKA, M. B.**, MARTINS, D., SANTORO, C. E., CASTELLARI, R. R., FERRARI, F., MARANGONI, S., NOVELLO, J. C. Fractionation of the Xylella fastidiosa Proteome using liquid chromatography and 2-DE analysis In: XXXI Reunião Anual - SBBq, 2002, Caxambu.

SANTORO, C. E., **SMOLKA, M. B.**, WINCK, F. V., MARANGONI, S., MARTINS, D., NOVELLO, J. C. A 2D reference map of basic proteins from Xylella fastidiosa In: XXX Reunião Anual - SBBq, 2001.

WINCK, F. V., **SMOLKA, M. B.**, MARTINS, D., SANTORO, C. E., MARANGONI, S., NOVELLO, J. C. Analysis of Xylella fastidiosa proteome using narrow pi range 2D maps In: XXX Reunião Anual - SBBq, 2001

BRUM, I. J., **SMOLKA, M. B.**, NOVELLO, J. C., GALEMBECK, E. Calculation of pI, MW, and theoretical 2D map: a computational tool for proteome analysis In: XXX Reunião Anual - SBBq, 2001, Caxambu.

MARTINS, D., **SMOLKA, M. B.**, SANTORO, C. E., BRUM, I. J., GALEMBECK, E., MARANGONI, S., NOVELLO, J. C. Database constrain tool for protein identification based on pI and MW prediction from genome sequence In: XXX Reunião Anual - SBBq, 2001.

SMOLKA, M. B., WINCK, F. V., SANTORO, C. E., MARTINS, D., GREENE, L., FAÇA, V., COLETTA FILHO, H. D., MACHADO, M. A., MARANGONI, S., NOVELLO, J. C. Proteome analysis of Xylella fastidiosa: first subset of identified proteins In: XXX Reunião Anual - SBBq, 2001, Caxambu.

Não relacionados ao projeto:

MARTINS, L. T. D., FREIRE, M. G., **SMOLKA, M. B.**, TOYAMA, M., DAMICO, D. C. S., MARANGONI, S., NOVELLO, J. C., MACEDO, M. L. Characterization of a lectin from Mamea americana seeds In: XXXI Reunião Anual - SBBq, 2002, Caxambu.

CABRINI, E. C., **SMOLKA, M. B.**, FREIRE, M. G., DAMICO, D. C. S., NOVELLO, J. C., MARANGONI, S., MACEDO, M. L. Biochemical characterization of a lectin from Pouteria Caimito seeds In: XXXI Reunião Anual - SBBq, 2002, Caxambu.

OLIVEIRA, B., SILVA, J. A., **SMOLKA, M. B.**, MATTIOLI, M. A. P., NOVELLO, J. C., MARANGONI, S. 2D Electrophoresis and Mass Spectrometry reveal: trypsin inhibitor from Copaifera langsdorffii seeds probably belongs to kunitz family! In: XXXI Reunião. Anual - SBBq, 2002, Caxambu.

FREIRE, M. G., MACHADO, O. L., **SMOLKA, M. B.**, NOVELLO, J. C., MARANGONI, S., MACEDO, M. L. Isolation and characterization of isolectins from Talisia esculenta seeds (TEL) In: XXXI Reunião Anual - SBBq, 2002

MATTIOLI, M. A. P., SILVA, **SMOLKA, M. B.**, J. A., BALDASSO, P., NOVELLO, J. C., MARANGONI, S. Molecular characterisation of BvcL, a lectin from Bauhinia variegata candida, usind electrophoresis, circular dichroism and mass spectrometry techniquesIn: XXXI Reunião Anual - SBBq, 2002.

REGO, E., CARVALHO, D., **SMOLKA, M. B.**, MARANGONI, S., OLIVEIRA, B., NOVELLO, J. C. Biochemical Studies of Lectin from Leguminosae seeds: Crotalaria pallida In: XXIX Reunião Anual - SBBq, 2000, Caxambu.

KASSAB, B., CARVALHO, D., **SMOLKA, M. B.**, NOVELLO, J. C. Snake venom lectins: a comparative study with BMool (Bothrops moojeni Lectin) and BJcul (Bothrops jararacussu Lectin) In: XXIX Reunião Anual - SBBq, 2000, Caxambu.

7. Co-orientações

Flávia Vischi Winck, Iniciação Científica - PIBIC, Fracionamento de proteomas por cromatografia líquida.

Daniel Martins, Iniciação científica - FAPESP, Análise do codon bias das proteínas mais expressas da *Xylella fastidiosa*.

Carlos Santoro, Iniciação Científica - FAPESP, análise de proteínas básicas da *Xylella fastidiosa.*

Itaraju Brum, Iniciação Científica - FAPESP, ferramentas computacionais para identificação de proteínas

Paula Oblessuc, Iniciação Científica.

Diego, Iniciação Científica, Análise proteômica do mecanismo de resistência a antibioticos na bactéria Helicobacter pylori.

Rafael, Capacitação técnica - FAPESP, Análise proteômica de *X. fastidiosa* por ICAT-MS.

Fernanda Ferrari, Capacitação técnica - FAPESP