Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 787 (2003) 101–128
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

Analysis of the *Candida albicans* proteome I. Strategies and applications

Aida Pitarch[a], Miguel Sánchez[b], César Nombela[a], Concha Gil[a,*]

[a]*Departamento de Microbiología II, Facultad de Farmacia, Universidad Complutense de Madrid, Madrid, Spain*
[b]*Departamento de Microbiología y Genética IMB-CSCIC, Universidad de Salamanca, Salamanca, Spain*

Abstract

The alarming incidence of invasive candidiasis, predominantly among the recent expanding immunocompromised population, the appearance of antifungal-drug resistance, and the lack of specific diagnostic tests for it have demanded more impactful research into *Candida albicans* pathogenicity. Proteomic approaches can provide accurate clues about its biological complexity. Indeed, initial *C. albicans* proteome analyses have focused on the understanding of dimorphism, host responses, the cell wall, virulence factors and drug resistance, among others. This review aims to briefly outline the technology available for proteomics-based studies, surveying the main proteomic approaches applied to *C. albicans* research. Prefractionation techniques, two-dimensional gel electrophoresis and mass spectrometry continue to be the backbone of proteomic projects. Emerging strategies for protein separation, quantification and identification may, however, challenge the pivotal position of 2D-PAGE. Regardless of this, since we are now approaching the completion and annotation of *C. albicans* genome sequencing, systematic characterization of the proteome of this fungal pathogen, although still in its early stages, heralds an exciting expansion of our knowledge in years to come.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Review; *Candida albicans*; Proteomics

Contents

1. Introduction 102
2. Towards a multidimensional approach for the analysis of the ***C. albicans*** proteome 103
 2.1. 2D-PAGE, a useful tool for the simultaneous analysis of complex protein mixtures 104
  2.1.1. Proteomic studies to be addressed for the study of important C. albicans issues 104
 2.2. Prefractionation techniques, a powerful third dimension to 2D-PAGE 110
  2.2.1. Subcellular fractionation 110
  2.2.2. Chromatographic techniques 113
  2.2.3. Recycling IEF: a liquid-based IEF technique 114
  2.2.4. Multiple gel fractionation using narrow-range IPGs 114
3. Gateway linking the ***C. albicans*** genome to its proteome 115
 3.1. Classical protein identification methods, a relatively low-throughput approach 115

*Corresponding author. Tel.: +39-91-394-1748; fax: +39-91-394-1745.
E-mail address: conchagil@farm.ucm.es (C. Gil).

1570-0232/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
doi:10.1016/S1570-0232(02)00761-4

3.1.1. Immunoblotting 115
3.1.2. Edman degradation: N-terminal and internal peptide sequencing 117
3.2. Mass spectrometry, a key component of the rapid expansion in proteome research 118
3.2.1. Peptide mass fingerprinting: a high-throughput tool for identifying C. albicans proteins present in any of the public domain databases 119
3.2.2. Peptide sequence tags: a solution for proteins not annotated in the databases or ambiguous protein identifications 119
4. News corner 122
4.1. *C. albicans* proteomics: a boundless platform? 123
Acknowledgements 125
References 125

1. Introduction

The incidence of opportunistic fungal infections has increased dramatically during the last 2 decades as a result of the important expanding immunocompromised population, including victims of the acquired immunodeficiency syndrome (AIDS), cancer patients, and other individuals treated with aggressive chemotherapy or cytotoxic drugs, as well as increasing medical awareness directed towards prolonging patient survival (e.g. organ transplants, indwelling catheter utilization, longer hospitalization or broad-spectrum antibiotic use) [1–3]. The opportunistic fungal pathogen *Candida albicans*, a normal component of the oral cavity, gastrointestinal tract or vaginal microflora, remains one of the most frequent causative agent of this deep mycosis in humans, being the fourth most common cause of bloodstream infections [4]. The prognosis of *C. albicans* infections is mainly hampered by the reduced efficacy of currently available drugs, the development of antifungal resistance, and the lack of rapid and specific diagnostic tests [5–7]. Thus, candidiasis represents a major social and clinical problem. The pressing need to search for new antifungal targets and novel diagnostic strategies has made *C. albicans* an attractive organism to study.

The understanding of putative virulence factors that contribute to the pathogenicity of *C. albicans* [8–10] also supports its biological and medical interest. In addition, since *C. albicans* is a polymorphic fungus able to grow in the budding and filamentous forms during its proliferating in host tissues [11,12], it is also currently being used as a model in basic research studying morphogenesis and cellular differentiation. All the above reasons therefore suggest that *C. albicans* is a relevant eukaryotic human pathogen model amenable to research.

The application of molecular biology techniques to this yeast, such as transformation systems, gene disruption strategies, gene reporter assays, and expression-controlled systems [13] has allowed the genetic description of different biological and pathological features of *C. albicans*. However, in order to obtain global insight into these aspects and answers for key issues about *C. albicans* morphogenesis, pathogenicity and virulence, as well as other aspects of clinical interest, not only molecular genetic approaches but also protein–biochemical approaches are needed. The latter could provide accurate clues as to the biological complexity of this organism, since proteins are functional biomolecules that are directly involved in physiological and pathological processes.

The current postgenome era heralds a new phase for the biochemical understanding of final gene expression products through large-scale proteomic and functional studies. Proteomics offers a global and integrated view of the entire protein complement expressed by a genome (of a simple organism), a cell or a tissue (in multicellular organisms); namely the proteome [14,15]. Accordingly, unlike its genome, the proteome or the ‘‘genome operating system’’ of any cell system is dynamic and complex, because any given genome can yield a large number of proteomes [16,17]. Proteomics-based analyses involve the simultaneous separation, visualization and quantification of thousands of proteins by two-dimensional gel electrophoresis (2D-PAGE) and their subsequent identification and characterization by mass spectrometry (MS). Concomitantly, subproteomics, a novel tendency based on breaking

down a total proteome into several subproteomes [18], is playing a decisive role in the study of the protein complement of the genome. It is clear, however, that all these biochemical procedures generate a rapid accumulation of data. This evidently requires the enlargement and/or the development of new protein databases. Well-annotated protein and 2-DE databases are currently emerging in diverse biological and biomedical areas. In view of this, proteomics therefore implies the integration of 2D-PAGE, mass spectrometry and bioinformatics-based technologies.

Nevertheless, systematic characterization of the *C. albicans* proteome is still in its infancy, although mandatory expansion will be certainly undertaken in the coming years. Preliminary proteomic analyses applied to *C. albicans* have contributed towards a better understanding of process involved in morphogenesis, virulence and host response as well as the development of potential antifungal targets and diagnostic approaches.

The aim of this review is to highlight the different biochemical strategies available for proteome analysis, placing special emphasis on 2D-PAGE and MS-associated technology since these procedures have become paramount in large-scale proteomic projects [19]. In the next section, we discuss the relevance of two-dimensional gel electrophoresis as the main method for resolving complex protein mixtures and its applications to *C. albicans* research. An outline of the different prefractionation techniques accomplished prior to the separation of *C. albicans* proteins by 2D-PAGE is also given. After this, in the following section, we evaluate the methodology used for the identification of proteins from this human fungal pathogen. Finally, we survey recent proteomic trends, or alternative tools under development, in protein separation and characterization from other microorganisms, which could possibly be scaled up to the pathogen *C. albicans* in the near future.

2. Towards a multidimensional approach for the analysis of the *C. albicans* proteome

Protein separation methods have been developed on the basis of differences in the physical and chemical properties of proteins, such as size, charge, hydrophobicity, polarity or affinity for other molecules. Many techniques for the separation of cell proteins have been described extensively in the scientific literature, including centrifugation, membrane dialysis, immunoprecipitation, electrophoresis and chromatography (e.g. ion-exchange, size-exclusion, affinity and reversed-phase). Nonetheless, since thousands of proteins are present in any given cell system, and due to the overlapping physicochemical characteristics of some proteins, no single technology can be used to isolate all the proteins from a complex sample, multidimensional approaches (two- or three-dimensional methods) therefore being required, on-line or off-line, for proteomics-based analyses [20].

A combination of several biochemical procedures, which depend on different parameters (multidimensional arrays), was initially used for protein purification, such as (i) sequential application of different solvents in paper chromatography, (ii) differential centrifugation according to sedimentation rates and subsequent separation by density gradients, (iii) the combined use of chromatography and SDS–PAGE electrophoresis, or more recently, (iv) the association of capillary electrophoresis and liquid chromatography, (v) the combination of isoelectric focusing and SDS–PAGE electrophoresis (2D-PAGE), and (vi) the on-line use of different chromatographic techniques (e.g. size-exclusion and reversed-phase liquid chromatography [21]). Very few strategies, however, allow researchers to do proteomics or, more accurately, to isolate the complex protein mixtures—characteristic of the cell system proteomes—into their individual components simultaneously. At present, 2D-PAGE undoubtedly continues to be the preferred method for separating thousands of proteins at the highest resolution in a single experiment. 2D-PAGE is therefore the starting point of proteomics-based research projects, playing a crucial role in this [22,23].

In this section our focus is specifically restricted to 2D-PAGE technology as the main protein separation procedure for large-scale screenings. We critically discuss its impact on *C. albicans* inquiries, emphasizing the broad variety of its biological and biomedical applications. Finally, we also review some used approaches for *C. albicans* proteome

analyses in an attempt to circumvent certain 2-DE limitations.

2.1. 2D-PAGE, a useful tool for the simultaneous analysis of complex protein mixtures

Two-dimensional gel electrophoresis combines two different and sequential separation mechanisms. In the first dimension, the proteins are resolved according to their intrinsic charge or isoelectric point (p*I*) by isoelectric focusing (IEF). Each protein migrates through an immobilized pH gradient until it reaches the pH at which it has no net charge (p*I*). In the second dimension, the isoelectrofocused proteins are separated according to their size or apparent molecular mass by electrophoresis on polyacrylamide gels (SDS–PAGE). Thus, it is possible to estimate the p*I* and Mr of the resolved proteins. The high-resolution power of this technique lies in its two separation parameters since they are basically orthogonal and independent.

Since its development by O'Farrell [24] and Klose [25] in 1975 to the present day, this procedure has undergone significant improvements in its resolution, particularly in the first-dimension separation, reinforcing its crucial impact on the analysis of biological systems and/or medical problems [26]. The key to its success is based on the overwhelming replacement of first-dimension carrier ampholyte-generated pH gradients with immobilized pH gradients (IPGs) [27,28]. This enhances the resolution of patterns [29], interlaboratory reproducibility [30] and the protein load capacity—permitting the detection of low-abundance proteins and spot identification by means of micropreparative 2-DE [31]. The introduction of new reagents for protein solubilization in conjunction with the use of organic solvents has also provided a noteworthy increase in the total number of proteins isolated on 2-D gels and a better separation of hydrophobic proteins [32,33]. In addition, the development of very narrow-range IPGs has facilitated the resolution of proteins at extreme pH values or with low copy-numbers [18,29]. Finally, 2-D protein spots can be visualized with high sensitivity using silver stain, fluorescent dyes, radiolabels or immunoblotting in combination with enhanced chemiluminescence (ECL) [34].

2.1.1. Proteomic studies to be addressed for the study of important C. albicans issues

Regardless of current improvements in the resolution, reproducibility and sensitivity of 2D-PAGE, *C. albicans* proteomic studies are still in their infancy. This may be due to the lack of substantial automation and the high costs associated with this technology [35] and to the difficulty involved in analyzing low-abundance or very hydrophobic proteins (see Section 2.2) with potential roles in pathogenesis or virulence. However, some 2-DE gels have been developed in the last years in an effort to investigate different *C. albicans* issues of biological and clinical interest, such as dimorphism, host responses (humoral immune response), cell wall, proteasomes, virulence factors, drug resistance, etc. Table 1 summarizes the 2-DE applications undertaken for the proteomic study of this fungal pathogen.

The aim of this methodology applied to *C. albicans* has not only been to separate complex protein mixtures and to achieve a global and integrated view of gene expression at protein level, but also (i) to compare protein expression profiles simultaneously under specified sets of conditions (expression proteomics), (ii) to establish reference protein maps from entire (yeast/hyphal) cells, specific subcellular organelles, or protein complexes (structural proteomics or cell map proteomics), (iii) to detect and characterize proteins of interest, and (iv) to identify post-translational modifications by means of differences in charge and/or mass in the protein (Fig. 1). Functional proteomics is hitherto unreported in the *C. albicans* literature, but could provide further key information about interactions between protein targets and antifungal agents and/or about pathogenicity mechanisms. Although a brief survey of proteomic applications to *C. albicans* research is outlined below, the reader is also referred to a recent review on this topic [36].

(i) Dimorphism. One of the main goals in proteomics is to determine which proteins are up- or downregulated when challenged by a particular environmental cue. The morphogenetic transition of *C. albicans* can be regulated by three external variables: temperature, pH and nutrients (such as serum, *N*-acetylglucosamine or proline) [12,37]. Pioneering *C. albicans* studies in the 2-DE field

Table 1
Summary of proteomic applications. Main separation and identification methods to the analysis of *Candida* proteome

Application	General aims	Results	Separation methods	Identification methods	Year	References
Dimorphism	Comparison of different yeast and hyphal protein expression profiles	Detection of 33 yeast and 10 hyphal specific proteins during temperature-regulated dimorphism. Immunoabsorption and nonfilamentous mutant analyses revealed that hypha-specific proteins were modifications of proteins present in yeasts. Prolonged labeling. Autoradiography	Antibody crossabsorption + 2D-PAGE		1980	[38]
	(1) Cytoplasmic fraction	Detection of five yeast specific proteins and two hyphal downregulated proteins during temperature-regulated dimorphism. Labeling at four intervals. Autoradiography.	2D-PAGE		1981	[39]
		Detection of only one yeast- and one hyphal-specific protein during pH-regulated dimorphism. This was concluded from nonfilamentous mutant analyses. Labeling at three intervals Autoradiography.	2D-PAGE		1985	[40]
	(2) S100 fraction	2-DE comparison of pH-dependent germ tube induction by glucose and galactose with pH-independent induction by GlcNAc. Minor changes associated only with growth and carbon source metabolism. Silver stain	Fractionation + 2D-PAGE		1996	[41]
	(3) Cell wall fraction	Qualitative and quantitative changes in different yeast and hyphal cell wall protein fractions. Important increase in hyphal proteins linked to chitin. Identification of several up and downregulated proteins. Silver stain	Subcellular fractionation + 2D-PAGE	Immunoblot analysis + MALDI-TOF +MALDI-TOF/TOF MS	2002	[42]
Host response (humoral immune response)	Protein detection and characterization	2-DE comparison of yeast and hyphal cytoplasmic antigens Immunoabsorption of [^{35}S]-labeled yeast and mycelial cytoplasmic proteins with anti-yeast and anti-mycelial polyclonal antibodies	Antibody crossabsorption + 2D-PAGE		1980	[45]
		Immunodetection of cytoplasmic antigens by a supernatant of hybridoma cells containing monoclonal antibodies against *C. albicans*	2D-PAGE+ Immunoblot		1990	[46]
		Immunodetection of different *C. albicans* immunogenic isoforms of 41000 and p*I* 6.1–6.9 by a monoclonal antibody (RJ5)	2D-PAGE+ Immunoblot		1991	[47]
		Five major cytoplasmic antigens present in four clinical *C. albicans* isolates recovered from neutropenic cancer patients	2D-PAGE		1994	[152]
		Immunodetection of several *C. albicans* antigens from cytoplasmic extracts, protoplast lysates and proteins secreted by protoplasts in sera from systemic candidiasis patients. Identification of two antigens (Eno1p and Tdh3p). Eno1p and a 34 000 secreted protein were the most reactive	Subcellular fractionation+ 2D-PAGE+ Immunoblot	Immunoblot analysis	1999	[48]
		Immunodetection of several *C. albicans* cytoplasmic antigens in murine and human sera. Identification of Eno1p as antigen	2D-PAGE+ Immunoblot	Immunoblot analysis	1999	[49]
		Identification of four *C. albicans* antigens (Met6p, Aco1p, Pyk1p and Gpm1p) by matching of tandem mass spectrometry data to *S. cerevisiae* proteins.	2D-PAGE	MALDI-TOF+ nESI-MS–MS	2000	[51]
		Identification of several antigens of *C. albicans* in human sera by mass spectrometry.	2D-PAGE	MALDI-TOF +nESI-IT–MS		[52]

Table 1. Continued

Application	General aims	Results	Separation methods	Identification methods	Year	References
	Comparison of differentially expressed proteins	Different serologic response in BALB/c and CBA/H mice Strong increase in Eno1p in resistant strain (BALB/c) and of Met6p, Ssb1p and Pgkp in susceptible strain (CBA/H) Identification of Pgkp, Adh1p, Met6p, Imhp, Pgkp, Fbap and Tpip	2D-PAGE + Immunoblot	Immunoblot analysis + *N*-terminal sequencing	2001	[50]
Cell wall	Sequencing	Separation of hydrophobic *C. albicans* cell wall proteins by preparative isoelectric focusing and continuous elution preparative electrophoresis, maintaining the proteins in soluble form to minimize loss during their separation. This provides sufficient amounts for sequencing analysis	Fractionation+ liquid IEF+ electrophoresis	ESI-MS–MS	1998	[57]
	Collection of reference	Reference 2-DE map of *C. albicans* proteins secreted by protoplasts. Identification of Bgl2p, Eno1p and Tdh3p	Subcellular fractionation+ 2D-PAGE	Immunoblot analysis	1999	[48]
	2-DE protein maps	2-DE separation of *C. albicans* β-ME extracts and subsequent detection by a combined silver–Coomassie staining method to improve the visualization of several acidic and basic protein spots	Subcellular fractionation+ 2D-PAGE		2000	[58]
		Reference 2-DE maps of different *C. albicans* cell wall protein fractions. Improvement in the enrichment and solubilization of wall proteins by chemical and enzymatic treatments. Identification of several cell surface-associated proteins	Subcellular fractionation+ 2D-PAGE	Immunoblot analysis+ MALDI-TOF+ MALDI-TOF /TOF MS	2002	[42]
Proteasome	Collection of reference 2-DE protein maps	20S proteasome. Reference map with at least 14 polypeptides. Identification of α6/Pre5, α3/Y13, α5/Pup2 and α7/Prs1. Determination of a phosphorylated protein (α7/Prs1) by acid phosphatase treatment	Fractionation+ chromatography+ 2D-PAGE+ immunoblot	Immunoblot analysis+ N-terminal sequencing	2000	[60]
Virulence factors	Protein detection and characterization of	Immunodetection of three ubiquitinated cell surface proteins of 37 000, 40 000 and 58 000 (*C. albicans* cell surface receptors) in β-ME extracts	2D-PAGE + Immunoblot	Immunoblot analysis	1996	[61]
	(1) Adhesins	Identification of *C. albicans* Csh1p. Protein purification on hydrophobic interaction chromatography column	Chromatography+ 2D-PAGE+ Immunoblot	Immunoblot analysis+ LC–IT-MS	2001	[62]
	(2) Lytic enzymes	Identification of *C. albicans* Sap2p. Evidence of its mucinolytic activity by A zymogram from a native IEF gel	2D-PAGE+ Immunoblot	Immunoblot analysis+ N-terminal sequencing	1996	[63]
Drug resistance	Comparison of differentially expressed proteins	2-DE comparison of fluconazole-sensitive and resistant *C. glabrata* isolates[a]. The latter displayed 25 upregulated and 76 downregulated proteins	2D-PAGE		1997	[64]
Others	Comparison of differentially expressed proteins	Overexpression of Ef-1αp, Ssa2p and Rps5p in a Cgt1-deficient mutant. CGT1 encodes mRNA GTasa.	2D-PAGE	N-Terminal sequencing+ MALDI + nESI-Q-TOF	2000	[65]
	Sequencing	Identification of Eno1p, Adh1p, Pgkp, Tdh3p, Pmmp, Ssa1p, Ssb1p by mass fingerprinting of SYPRO-Red-stained 2-DE gels.	2D-PAGE	MALDI-TOF	2000	[114]

[a] 2D-PAGE analysis in non-*albicans Candida* spp.

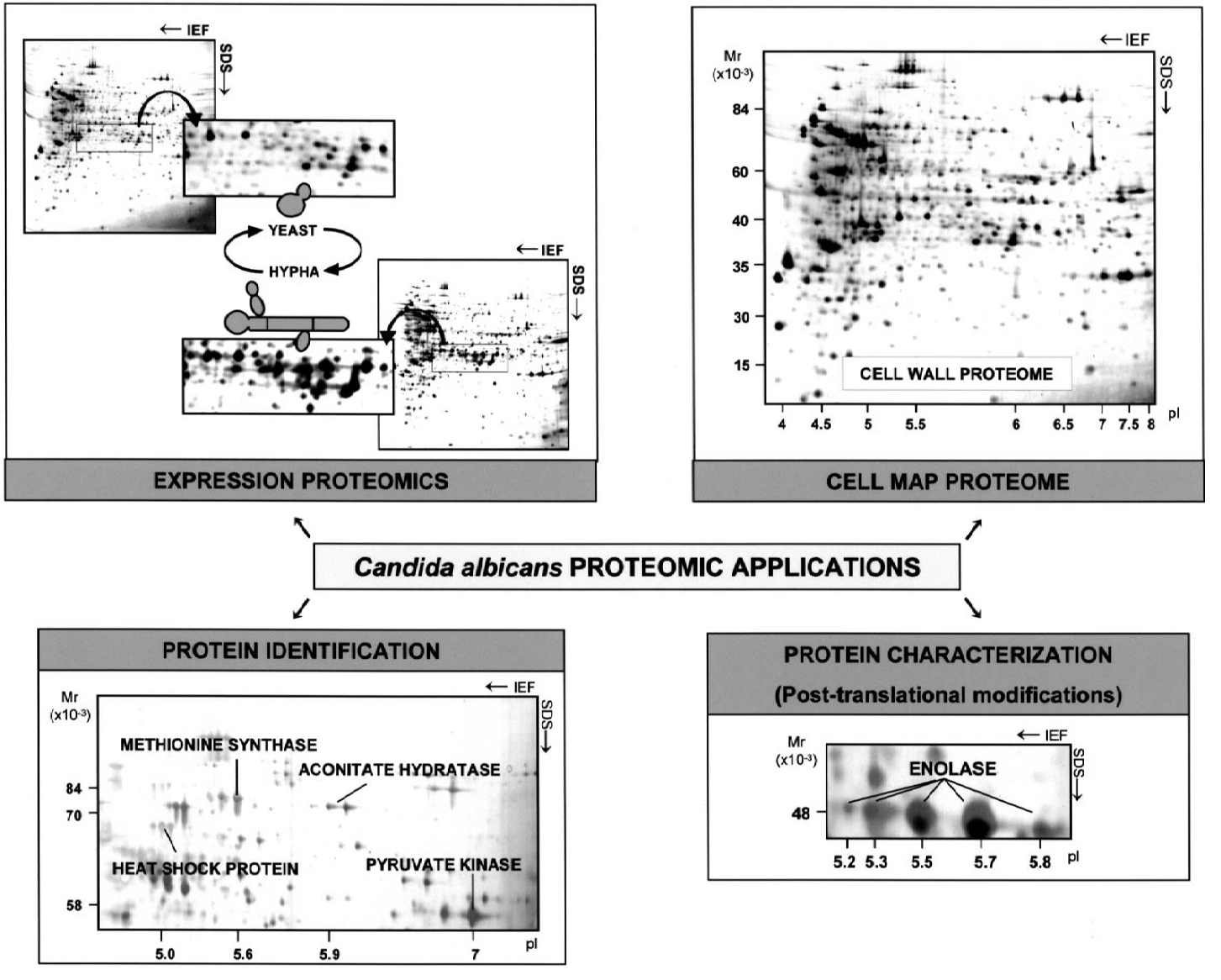

Fig. 1. Main current aims of the application of proteomics-based analysis to *C. albicans*.

focused on the analysis of differential protein expression from yeast and mycelial cytoplasmic extracts at global scale and in a nondiscriminatory way [38]. Several subsequent works were also performed in this mode but including detailed temporal assays [39,40]. Likewise, Niimi et al. [41] evaluated the initiation of *C. albicans* morphogenesis using soluble proteins of cell-free S-100 fractions. However, the results of all these initial investigations revealed few differences in the detection of alternative morphologies using 2D-PAGE. It is likely that a large number of regulatory proteins remain undetected due to limitations of the method (e.g. the temporal and/or spatial appearance or the low abundance of key proteins in the cell [40]). In contrast, certain qualitative and quantitative changes in gene product expression have been visualized in different enriched fractions of cell wall proteins when yeast and hyphal forms were compared [42]. These results support cell wall remodeling during morphogenetic conversion, since this latter process involves alterations in wall composition and organization [43]. It has been postulated that the ability of *C. albicans* to switch between the budding and hyphal forms could represent a virulence factor, although its genetic basis remains obscure [44]. The combination of proteomics with functional studies could therefore contribute to elucidating this issue in future research.

(ii) Host response. Early analyses of *C. albicans* proteome also concentrated on understanding the humoral response [45]. 2D-PAGE is a powerful tool for simultaneously separating the complex antigenic composition of *C. albicans*. The detection of *C. albicans* antigens has been accomplished by combining of 2D-PAGE with immunoabsorption [45] or with immunoblotting using monoclonal or polyclonal antibodies [46–50]. Different serologic responses to systemic candidiasis in a murine model along the course of infection were reported between a resistant- and a susceptible-strain of mice when sublethal and lethal doses, respectively, were used [50]. Alternatively, several *C. albicans* immunogenic proteins, with different patterns of protein expression and subcellular localization, were also detected and characterized in sera from patients with confirmed systemic candidiasis [48,49,51,52]. The goal of this strategy is to obtain an overall and integrated view of all immunogenic proteins that are being produced in a patient. It could be helpful for the identification of potential markers of the infection and protective antigens, monitoring of the evolution of candidiasis and the development of novel diagnostic methods and future vaccines (Fig. 2).

(iii) Cell wall. High-resolution 2-DE has also been used to study the *C. albicans* cell wall. Owing to its privileged location within the cell—the outermost cellular structure—the cell wall is the initial point of contact between the fungus and its environment. This host–fungus interaction is directly related to pathogenesis, virulence, adhesion to host tissue, and immunomodulation of the host immune response [53,54]. Furthermore, the cell wall also determines the shape of the fungus, playing a key role in morphogenesis. Thus, proteomics might become a useful tool to gain insight into cell wall implications and to search for new antifungal targets, since this structure is absent in mammalian hosts [55]. However, cell wall proteins are difficult to analyze due to their high heterogeneity, interconnections with cell wall polysaccharides (mannan, glucan and chitin), low abundance, low solubility and hydrophobic nature [36,56]. In the quest to solve these problems, Masuoka et al. [57] have suggested a preparative electrophoretic technique for hydrophobic wall protein isolation that minimizes the loss of protein during their separation. Also, Pitarch et al. [48] have reported another novel proteome approach for studying *C. albicans* cell wall proteins based on the analysis of proteins secreted into medium during the regeneration of protoplast cell walls. A modified staining method that combines silver and Coomassie blue stain in order to improve the 2-D detection of *C. albicans* cell surface proteins extracted under reducing agents has recently been reported [58]. By sequential extraction of cell wall proteins according to the type of linkages they establish with other wall components, it has been possible to enhance the enrichment and solubilisation of these low-abundance proteins [42]. The following section examines these proteomic approaches in more detail (see 2.2.1 and 2.2.3).

(iv) Proteasome. Proteomics can also be used for the analysis of different subsets of a proteome, such as subcellular compartments, organelles, macromolecular structures and multiprotein complexes, in an effort to establish subcellular proteomes [59].

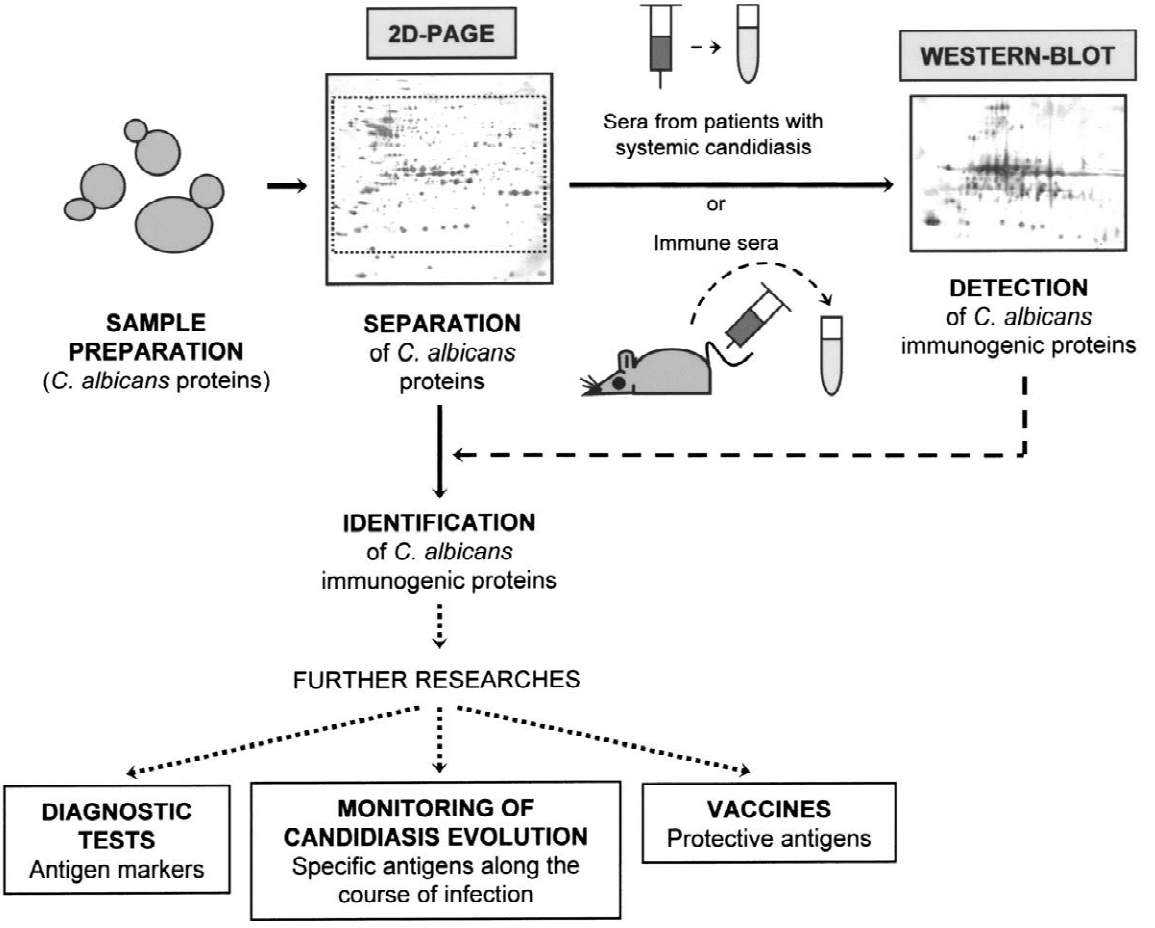

Fig. 2. Proteomic approach for the identification of immunogenic proteins from *C. albicans*. The combination of 2D-PAGE with immunoblotting using sera from patients with systemic candidiasis or immune sera is a powerful tool for the detection of antigenic proteins. The potential goals of this proteomics-based strategy are also shown.

Recently, the 20S proteasome from *C. albicans* has been elucidated by combining successive chromatographic steps with 2D-PAGE [60]. The 20S proteasome, a hollow cylindrical particle, is the central proteolytic component of the ubiquitin–proteasome pathway, which is responsible for the elimination of specific proteins and can also play a role in antigen presentation, stress response, signal transduction and cell differentiation, among others. 2-D immunoblotting of acid phosphatase-treated 20S proteasome afforded a multiple-phosphorylated subunit. This assay is an attractive approach for detecting post-translational modifications of proteins. In any case, current 2-DE analyses of *C. albicans* macromolecular structures (i.e. 20S proteasome) and subcellular compartments (i.e. cell wall) are only preliminary steps towards an attainable characterization of the complete proteome of this fungal pathogen.

(v) Virulence factors. The application of proteomics to the study of putative *C. albicans* virulence factors has also been exploited in the last few years. Adherence to host tissues, response to environmental changes (such as phenotypic switching and dimorphism) and the secretion of proteases and hydrolases are the main putative virulence attributes of this fungal pathogen [8,10]. Therefore, a better knowledge of the precise pathogenicity- and virulence-involved mechanisms could reveal poten-

tial antifungal targets that may be useful in the search for better therapies. In this sense, a helpful *C. albicans* proteomic approach based on the combination of 2D-PAGE or IEF and immunoblotting with specific antibodies has enabled the detection and characterization of (i) three ubiquitinated cell surface proteins —the 37 000 laminin receptor, the 58 000 fibrinogen-binding mannoprotein, and the candidal C3d receptor, suggesting that ubiquitin-like epitopes would be associated with cell surface receptors and thus with adhesion to host structures [61], (ii) a hydrophobic cell surface protein (Csh1p) that mediates binding to host target proteins [62], and (iii) a known secretory aspartyl proteinase (Sap2p) that is involved in mucin proteolysis, facilitating penetration of the gastrointestinal mucosa barrier by this pathogen [63].

(vi) Drug resistance. 2D-PAGE technology can also play a relevant role in the understanding of antifungal resistance mechanisms through comparisons of protein expression profiles of drug-sensitive versus -resistant strains. Marichal et al. [64] have demonstrated that a putative origin of azole-resistance in *Candida glabrata* isolates is the duplication of the chromosome that contains the CYP51 gene, which encodes the azole-target enzyme (i.e. 14-α-sterol demethylase). When the 2-DE protein patterns of fluconazole-susceptible and revertant isolates were compared with resistant ones, significant differences were seen. This suggested that the chromosome duplication resulted in pleiotropic effects on other genes in addition to the CYP51 gene. Further characterizations of these up- or downregulated proteins might therefore facilitate the elucidation of more azole-resistance mechanisms.

(vi) Others. Other 2-DE studies directed towards the comparison of differentially expressed proteins have also been carried out. De Backer et al. [65] have recently reported that at least three proteins are upregulated in a disrupted strain of the CGT1 gene. CGT1 encodes mRNA GTase, which is a potential antifungal target. 2-DE assays revealed the influence in the inhibition of CGT1 activity in *C. albicans*, pointing to the overexpression of a translation elongation factor and a heat shock protein. This confirms the pleiotropic effect of CGT1 depletion (i.e. an increase of resistance to hygromycin B—an antibiotic that affects translational elongation—and to heat stress).

To sum up, 2-DE applications to *C. albicans* surveys are wide-ranging. Although proteome research in *C. albicans* is still in its early stages, it is fast expanding. During the 1980s, only five studies were performed using this technology in *C. albicans*. However, the frequency of publications concerning this issue increased up to two-fold over the decade from 1990 to 1999. This trend currently continues, with seven proteomics works about *C. albicans* in the first 2 years of the present century (Table 1). However, although there are diverse limitations to this methodology, which are being circumvented as described below, proteomics promises an exciting expansion in the *C. albicans* field in years to come.

2.2. Prefractionation techniques, a powerful third dimension to 2D-PAGE

As a result of the difficulties involved in isolating low-abundance and/or hydrophobic proteins (e.g. cell surface, membrane, other cell organelles, macromolecular structure-associated or DNA-binding proteins, among others) further separation steps before 2D-PAGE are required for resolving them. Some approaches for bypassing these drawbacks, such as subcellular fractionation, chromatographic techniques, liquid-based preparative IEF or gel fractionation by narrow-range IPGs, have been reported in the *C. albicans* literature (Table 1). These biochemical procedures mostly aim at (i) reducing sample complexity and thus enhancing the number of low copy-number proteins by removing the major components, and/or (ii) improving the solubilization of certain proteins. Indeed, these prefractionation techniques could provide a powerful third dimension to 2D-PAGE and therefore constitute a multidimensional approach in proteome analyses. In view of this, the future trend in proteomics-based research will probably be the analysis of subproteomes instead of whole proteomes [18]. Below we review these strategies.

2.2.1. Subcellular fractionation

The goal of this method is not only to isolate proteins from different cellular structures or compartments in order to define their intracellular location but also to enrich samples for specific proteins of known cellular location that could be masked on 2-DE gels by the most abundant soluble gene

products. Sequential extraction can also help to increase the solubility of some proteins (such as hydrophobic proteins from the cell wall or associated with membranes), using specific extraction reagents for each subcellular fraction [42,66]. In addition, cell-fractionation prior to proteome analysis can offer a useful approach for (i) the understanding of protein–protein interactions or signal transduction pathways, and/or the discover of new regulatory networks that are exclusive to one organelle or compartment, (ii) the mapping and identification of the proteins that make up a particular subset, generating 2-D reference protein maps for future research, and (iii) the monitoring of abnormal protein expression localized to a given organelle [59]. The combination of this biochemical tool with 2D-PAGE has recently been described in *C. albicans* [41,42,48,57,58,60,61].

The initial step in the subcellular fractionation of *C. albicans* yeast or hyphal cells is to rupture the cell wall and plasma membrane. Although many methods are able to achieve cellular disintegration, mechanical disruption using glass beads is undoubtedly one of the most rapid and effective procedures. Subsequently, a fractionation of the homogenate obtained is performed. To this end, numerous biochemical techniques can be used to isolate the different populations of cellular compartments or structures, such as sequential extraction, differential centrifugation, sucrose density gradients and/or free-flow-electrophoresis (on the basis of differences in their relative solubility in several buffers, in their mass, in their density and/or in their charge, respectively) as well as immunoprecipitation (using organelle-specific antibodies). After cell disruption, a differential centrifugation at relatively low speed is commonly carried out in order to separate walls from other cell components (crude extract). A schematic diagram of *C. albicans* cell-fractionation that employs differential centrifugation, sequential extraction and density gradients is shown in Fig. 3.

(i) Cell walls. Since various cell wall proteins of *C. albicans* are covalently linked to the wall structural polysaccharides (e.g. mannans, β-1,3- or β-1,6-glucans and chitin) [54,67], the degree of complexity, insolubility and hydrophobicity of the sample is extremely high, hampering their isolation and their ensuing resolution by 2D-PAGE. Different strategies for cell wall fractionation have been developed for the purpose of solubilizing and extracting *C. albicans* proteins from this compartment (Fig. 3).

In the first instance, after mechanical disruption and differential centrifugation as described above, proteins from isolated walls of *C. albicans* may be solubilized by a wide variety of reagents (e.g. reducing agents, detergents, alkalis, hydrolytic enzymes, etc.) that can be used either alone or in combination [42,68–72]. Firstly, reducing agents, such as dithiothreitol (DTT) and β-mercaptoethanol (βME) mainly release proteins that are loosely associated, either by disulfide bridges or noncovalent bonds, with other wall components, in addition to increasing cell wall porosity [73], thus facilitating the subsequent action of wall degrading enzymes. Secondly, detergents, such as sodium dodecylsulfate (SDS) or *n*-octylglucoside, also solubilize mannoproteins loosely associated with the cell wall framework through noncovalent interactions [68,74]. Thirdly, the β-elimination process (using 30 m*M* NaOH) extracts proteins linked to β-1,3-glucan via an alkali-labile linkage (e.g. through their *O*-glycosyl side chains) [71,75]. Finally, different hydrolytic enzymes, such as β-1,3-glucanases (quantazyme, zymolyase, lyticase or laminarinase), β-1,6-glucanases or chitinases, solubilize proteins covalently anchored to the glucan or chitin networks, respectively [72]. We have achieved a sequential extraction of different mannoproteins that make up *C. albicans* cell wall architecture in combination with 2D-PAGE [42]. To this end, yeast and hyphal proteins from isolated walls were released by hot SDS and DTT treatment, and then either by mild alkali or by enzymatic treatment with glucanases and chitinases (Fig. 3), obtaining eight different cell wall subfractions and hence eight reference 2-DE maps. Consequently, this approach improves the enrichment in and solubility of proteins from this compartment, and also provides a global and integrated view of the cell wall proteome of *C. albicans*.

Alternatively, wall proteins may also be isolated from intact cells under mild conditions, either by reducing agents or by limited enzymatic digestion, in order to avoid cell lysis and thus contamination with intracellular material [57,58,61,69,76] (Fig. 3). Preliminary 2-DE separations of *C. albicans* cell surface proteins extracted under βME-treatment have been undertaken [58,61].

Interestingly, an exciting strategy for the study of

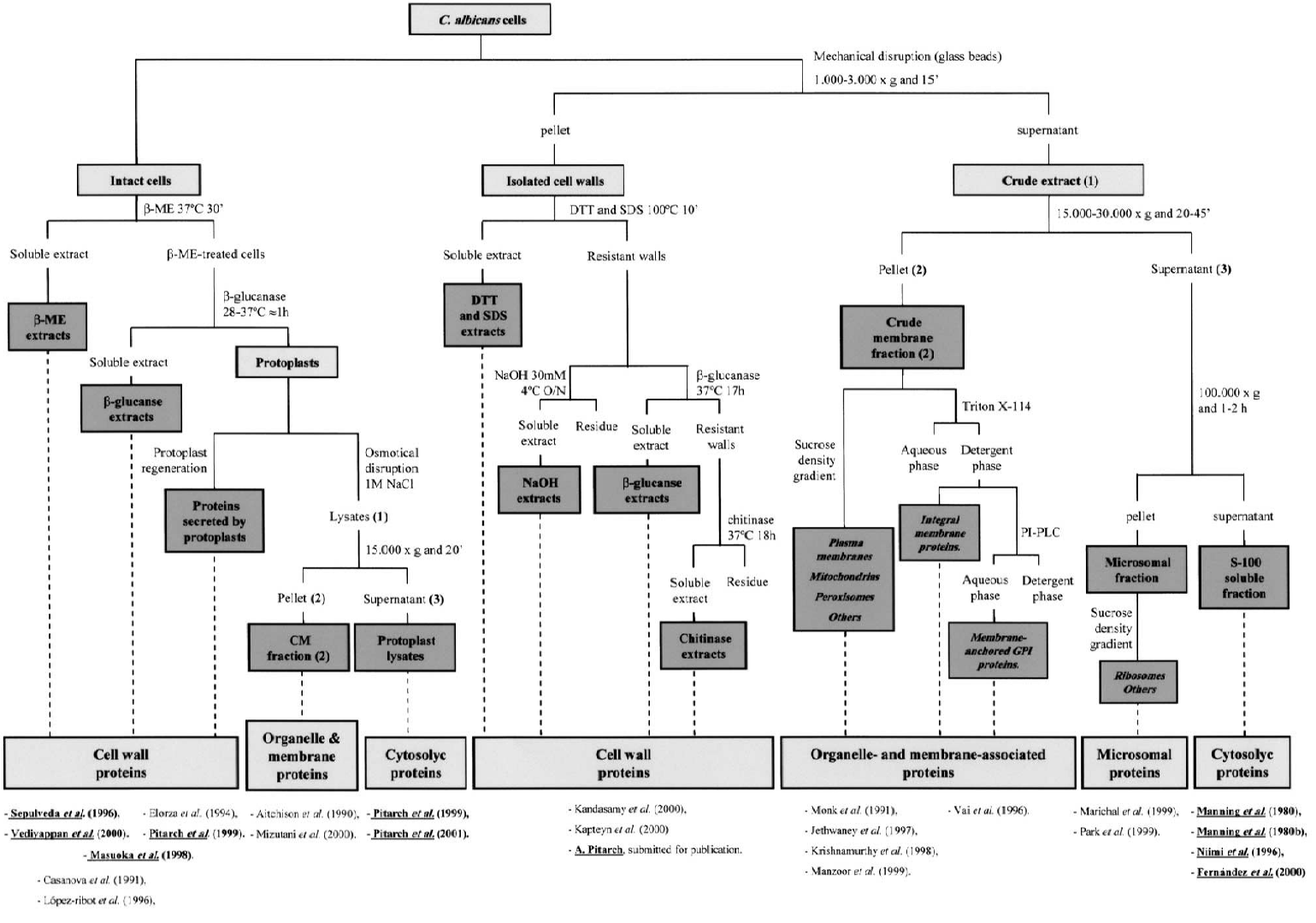

Fig. 3. Flow chart for the collection of specific subsets of the *C. albicans* proteome is proposed. Differential centrifugation, sequential extraction and density gradients are mostly used in subcellular fractionation. Numbers in brackets indicate that the workflow for the fractionation of protoplast lysates is equivalent to that shown in crude extracts. At bottom, underlined references correspond to journal articles related to 2D-PAGE analysis.

C. albicans wall proteins is to analyze the proteins secreted when protoplasts are regenerating their cell walls [48,77]. During the early stages of the regeneration process, significant amounts of proteins are secreted into medium but their covalent incorporation into the nascent cell wall is significantly delayed [77]. Accordingly, this approach may yield further information about the composition of cell wall precursors before they become retained in the wall and also about their putative interactions [78]. The main advantage of this method that does not destroy covalent linkages is that it bypasses the underlying problems with chemical and enzymatic extractions (e.g. potential modifications of proteins or the presence of proteins bearing glucan and chitin side-chain residues [79]). Recently, a reference 2-DE map of *C. albicans* proteins secreted by regenerating protoplasts during the two first hours of incubation has been reported [48], affording an overall view of the gene expression products involved in de novo cell wall biosynthesis.

(ii) Crude extracts. After removing *C. albicans* cell walls, the method most frequently applied to obtain specific cytoplasmic and membrane components is to isolate them by differential centrifugation at increasingly higher speeds and longer times. For instance, at fairly high speeds, a pellet of membranes, mitochondria and other organelles is sedimented, whereas at higher speeds and with longer centrifugation times, microsomes (such as ribosomes and small vesicles, among others) are deposited [80–82] (Fig. 1). Then, each pelleted fraction is resuspended and can be separated by a sucrose density gradient in order to isolate its specific organelles or structures through the migration of each of these until their appropriate equilibrium density has been reached [83–85].

This cell-fractionation may also be accomplished from protoplast lysates in the same way [86,87]. Instead of mechanical disruption and removal of the walls, these are completely eliminated by enzymatic digestion in an isotonic medium in order to attain the formation of protoplasts. Their osmotic rupture allows the cellular homogenate to be collected. The different subsets can then be separated by the above procedure.

An effective method for solubilizing membrane proteins is through biphasic separation in a nonionic detergent (e.g. Triton X-114), affording integral membrane proteins in the detergent fraction. In turn, this fraction can also be subjected to phosphatidylinositol-specific phospholipase or PI–PLC treatment to extract proteins attached to the membrane through GPI (glycosylphosphatidylinositol) in the aqueous phase [88] (Fig. 3). These GPI–proteins are found in all eukaryotic cells, and in yeasts they may be anchored to the plasma membrane and/or linked to β-1,6-glucan of the cell wall [89].

To date, 2-DE analyses of different *C. albicans* cellular subfractions have been concentrated exclusively on the cell wall [42,48,58,61], the 20S proteasome [60] and soluble cytoplasmic proteins [38,41,45,48,50]. Proteins isolated from organelles, membrane systems, microsomes or other macromolecular cellular structures of *C. albicans* have been resolved by SDS–PAGE and other separation techniques. Obviously, subcellular fractionation of these organelles combined with 2D-PAGE will doubtless be required in future research directed towards the establishment of the complete *C. albicans* proteome.

2.2.2. Chromatographic techniques

Another strategy aimed at improving the detection of low copy-number proteins of functional interest and/or the resolution efficacy of hydrophobic proteins on 2-DE gels involves chromatographic techniques. Depending on to the type of column matrix chosen, proteins can be fractionated on the basis of differences in their size (size-exclusion chromatography or SEC), their charge (ion-exchange chromatography or IEC), their hydrophobicity (hydrophobic interaction chromatography or HIC) or their ability to bind to other molecules or ligands (affinity chromatography).

The combination of both methodologies has allowed the enrichment of specific proteins from *C. albicans*. Although few examples have been reported for this fungal pathogen [60,62], this multidimensional approach promises significant advantages for its future proteomic analyses.

In this sense, successive chromatographic columns (on DEAE-Sephacel, Q-Sepharose, Sepharose CL-4B, phenyl-Sepharose and hydroxyapatite) have been used in combination with 2D-PAGE in an attempt to purify *C. albicans* 20S proteasomal subunits, which

are proteins with low expression levels [60]. This allowed the resolution of at least 14 polypeptides on 2-DE gels.

Hydrophobic interaction chromatographic columns before 2D-PAGE have been employed in order to isolate hydrophobic cell surface proteins of *C. albicans* [62]. Cell wall proteins with exposed hydrophobic regions (e.g. mannosyl groups [56]) are retained in these columns and then eluted with an ammonium sulfate gradient. This type of chromatography permits an enrichment of hydrophobic proteins and thus of low-abundance hydrophobic proteins, such as cell surface proteins.

Heparin–agarose affinity chromatography has been applied to S100 cell-free extracts with a view to separating potential DNA-binding polypeptides associated with the early phases of *C. albicans* germ tube formation [41]. Due to polyanionic nature of heparin, these columns can be used as a first stage in the purification of these DNA-binding proteins. A small group of soluble proteins isolated by this procedure was differentially synthesized during the process of morphogenesis. This suggested that regulatory polypeptides involved in the initial events of the dimorphic transition might be present at low expression levels. However, this sample prefractionation technique was only combined with SDS–PAGE. Perhaps, further 2-DE separations will be necessary to increase the protein resolution of these putative regulatory proteins.

2.2.3. Recycling IEF: a liquid-based IEF technique

As above outlined, low-abundance hydrophobic proteins, such as those of the cell surface or those associated with membranes, are difficult to isolate and analyze. This is mainly due to (i) their loss during the purification procedure, through their adsorption onto apparatus surfaces; (ii) the need for a large number of cultures to purify sufficient amounts of protein for subsequent biochemical and sequence analyses, and (iii) their poor solubilization during sample preparation or the first dimension separation. All these factors restrict biochemical study of their putative biological roles, such as adherence to host tissue and virulence, in which *C. albicans* cell wall proteins seem to be involved [90].

In an effort to minimize these drawbacks, Masuoka et al. [57] have reported a two-dimensional system that combines preparative isoelectric focusing with continuous elution preparative electrophoresis. This strategy employs a liquid IEF system (the Rotofor™ from Bio-Rad) that maintains protein solubility since the proteins are in solution instead of in gel matrices. This affords a reduction in protein–surface interaction and hence protein loss due to adsorption. After IEF, the fractions containing the protein of interest are pooled and isolated by preparative electrophoresis. Following this, the separated fractions are directly placed in tubes or collected in a dot–blot apparatus for subsequent sequencing. This enables significant intermediate steps to be circumvented, thereby limiting protein loss.

Accordingly, liquid-phase IEF combined with preparative electrophoresis provides sufficient amounts of material for sequencing since it minimizes protein loss by reducing (i) purification steps, (ii) the time of protein exposure to apparatus surfaces and (iii) the actual number of surfaces to which proteins are exposed.

2.2.4. Multiple gel fractionation using narrow-range IPGs

The use of very narrow-range immobilized pH gradients (nrIPGs) during the first-dimensional separation of 2D-PAGE allows researchers to increase the loading capacity, the resolving power, and the solubility of protein samples within a given pH range because of the reduction in sample complexity and the introduction of solubilizing reagents for more specific groups of proteins [18]. This prefractionation technique based on charge differences is therefore a powerful tool for (i) separating proteins with low expression levels, (ii) resolving overlapping protein spots through the use of wide-range IPGs, (iii) improving the resolution of proteins with extreme p*I* values, and (iv) constructing complete proteome maps through series of 2-DE patterns with overlapping ranges for IEF (proteomic contigs).

Contrariwise, *C. albicans* 2-DE analyses have focused on proteome screens using wide-range standard pH 3.0–10.0 gradients with a view to obtaining an overall view of the most abundant proteins in a nondiscriminatory way. Although some *C. albicans* acidic proteins (i.e. some secretory aspartyl proteinases and 20S proteasome subunits) have been isolated by means of a pH 4.0–7.0 linear gradient for

the first-dimension separation [60,63], the isoelectric focusing of proteins on single pH unit IPGs has not yet been undertaken. Hence, this field requires further development in order to gain a finer degree of 2-DE separation of the proteins from this fungus.

3. Gateway linking the *C. albicans* genome to its proteome

The ultimate goal of protein separation by 2D-PAGE is to identify and characterize the final gene expression products. In turn, protein identification pursues ambitious objectives, such as (i) the creation of reference maps able to define all the proteins expressed in a particular biological system (e.g. a cell, an organelle, a macromolecular structure, or a tissue) under specified physiological conditions, thus linking genomes to proteomes [91]; (ii) the gateway to protein characterization, including protein function, post-translational modifications, protein–protein interactions, complex formation, regulation of protein activity, metabolic and signaling pathways and structural analyses, among others [19,92,93] and (iii) verification of the information generated in genome sequencing projects (for instance, predictions of the ORFs that have been discovered [91,94]).

Separated proteins can be identified in many ways (e.g. immunoprecipitation, immunoblotting, co-electrophoresis, co-migration analyses, comparison of 2-DE gel positions, Edman sequencing, amino acid composition and mass spectrometry). Notwithstanding this wide variety of techniques, early studies on the characterization of *C. albicans* proteins resolved by 2D-PAGE were carried out using their isoelectric points, their size and/or their relative abundance. However, in the mid-1990s *C. albicans* proteins from 2-DE gels were subjected to different identification methods, such as antibody staining, Edman degradation, and/or MS (see Table 1).

After 2-DE separation, protein spots are (i) blotted onto the surface of a chemically inert membrane, such as nitrocellulose or PVDF, and then visualized using specific antibodies (by immunoblotting) or excised from the membrane (for subsequent identification by *N*-terminal sequencing or MS), or alternatively, (ii) excised directly from the gel (for MS-based characterization).

A flow chart of the general strategy for the identification of *C. albicans* proteins isolated by 2D-PAGE is outlined in Fig. 4. In this section we survey these procedures—from the relatively low-throughput methods of immunoblotting and Edman sequencing to the method of choice in the large-scale proteomic projects (i.e. mass spectrometry), underscoring their applications and their drawbacks.

3.1. Classical protein identification methods, a relatively low-throughput approach

The conventional protein identification methods are generally laborious, slow and/or expensive procedures to be used in screening large numbers of proteins, limiting their application to low-throughput approaches [95]. As above stated, two traditional techniques for characterizing proteins of proteomes resolved on 2-DE gels (antibody staining or immunoblotting, and *N*-terminal sequencing or Edman degradation) have been achieved in the gene-product analyses of this human fungal pathogen.

3.1.1. Immunoblotting

The combination of both biochemical tools—2D-PAGE and Western blotting—in the detection of specific proteins immobilized onto an inert support and conjugated directly or indirectly with labeled antibodies, integrates the high resolving power of 2-DE gels, the specificity of antibodies and the extreme sensitivity of radioactivity or enzyme assays. This powerful procedure allows researchers not only to identify a particular protein in a complex mixture, but also to detect post-translational modifications, to analyze immune responses, and/or to visualize low-abundance proteins [96].

This strategy has been used widely for the analysis and characterization of the *C. albicans* proteome, with the same goals as above. First, fourteen *C. albicans* proteins were isolated by 2-DE and identified through specific antibodies [42,48–50,60–63]. Second, post-translational modifications of some *C. albicans* gene products were detected by immunological cross-reactivity of different protein isoforms [47,48,50,60]. Phosphorylation of the proteasome α7/Prs1p subunit was elucidated by 20S proteasome incubation with acid phosphatase prior to its immunodetection on 2-D gel blots [60]. Nevertheless,

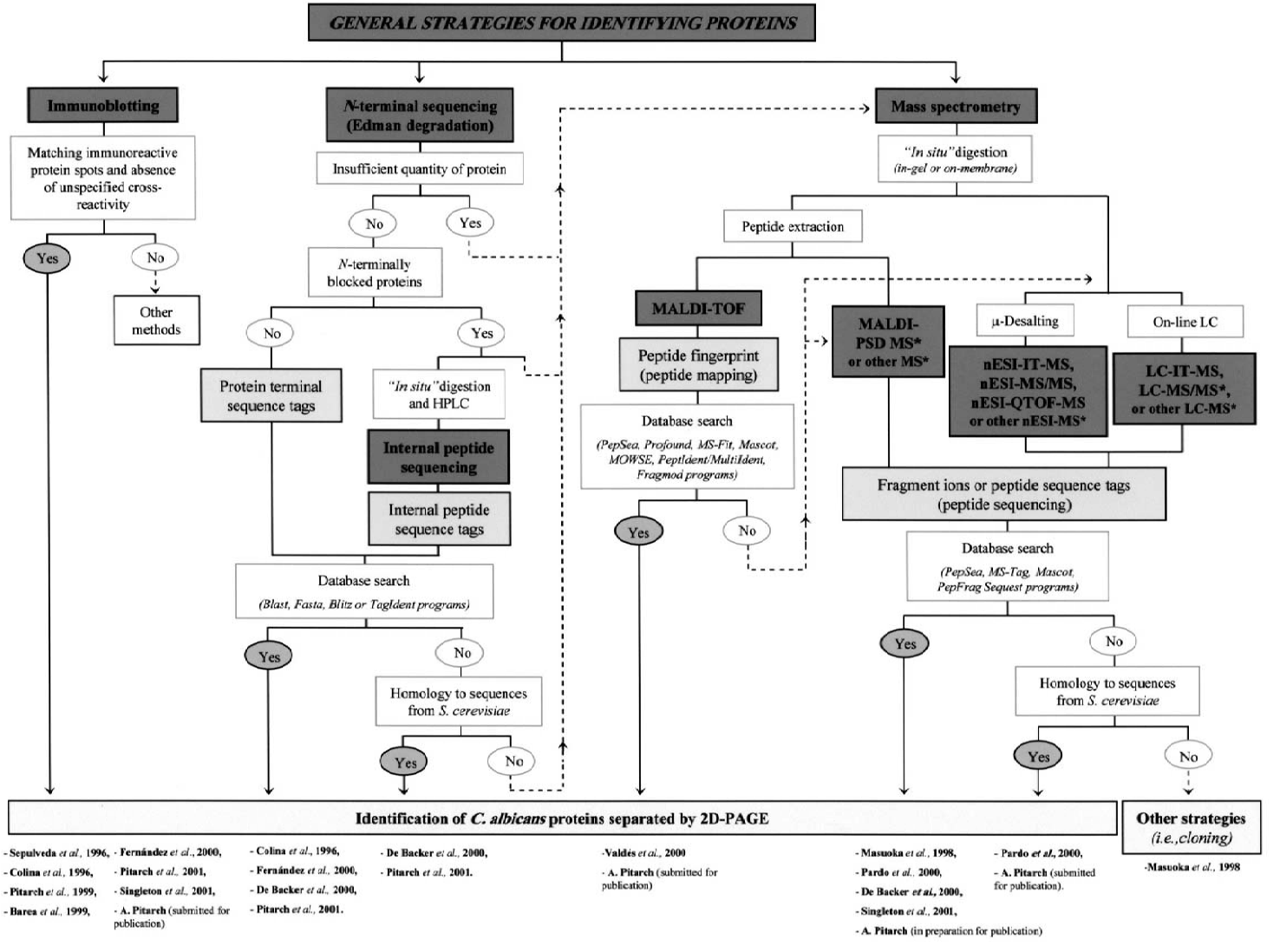

Fig. 4. Flow chart of the general strategy for identifying proteins of *C. albicans* proteomes separated by 2D-PAGE is given. Asterisks denote mass spectrometers that have hitherto been unreported in *C. albicans* scientific literature.

this comprehensive approach only allows one to detect the presence of the post-translational modification but not to identify either the phosphorylated amino acids or the phosphorylation sites. For the latter goals, MS-based methods are required. Third, 2D-PAGE followed by Western blotting with sera from patients with systemic candidiasis [48,49,52] and infected mice [46,49,50] allowed antigenic proteins from this human fungal pathogen to be immunorecognized. Finally, *C. albicans* proteins with low expression levels can be visualized as a result of the amplification of their detection by the enhanced chemilluminescence Western blotting detection reagent (ECL™) [48,50,60].

Despite these advantages, this tool is evidently a low-throughput approach since it only enables the analysis of one or a few proteins per experiment. In addition, procedure is subject to certain limitations, such as (i) the availability of specific antibodies for immunoaffinity identification, (ii) unspecified cross-reactivity with other proteins with similar epitopes and (iii) the difficulty involved in matching certain immunoreactive protein spots against silver-stained protein spots (see Fig. 4). However, various commercially available software packages for 2-D gel analysis (such as MELANIE, PDQUEST and IMAGEMASTER [97,98], among others) enable the chemiluminescent film over the silver-stained gel to be overlayed, as well as the subsequent automatic detection of the protein spot of interest.

3.1.2. Edman degradation: N-terminal and internal peptide sequencing

(i) Micro-sequence analysis. Since the middle of the last century, Edman degradation has been the traditional method for determining the amino acid sequence of a given protein. In this procedure, the N-terminal amino acid is cleaved from the polypeptide and identified by reversed-phase HPLC. This process is repeated on the residual shortened polypeptide—with a new free N-terminus—until at least five amino acid residues have been characterized.

Microsequence analyses of *C. albicans* protein spots from preparative 2-DE gels or 2-DE gel blots have recently been undertaken in two different ways: either by N-terminal sequencing or by internal peptide sequencing (Fig. 4). The first approach is performed on the whole protein [50,63]. Consequently, a short sequence of 5–15 amino acids, named protein terminal sequence tag, is obtained from the N-terminus of the intact protein. Owing to the specificity of these tags, it is possible to identify proteins from organisms with small genomes [91]. Alternatively, automated Edman degradation can be carried out on the internal peptide fragments generated by in situ (in-gel or on-membrane) tryptic digestion [60,65]. These fragments are previously separated by reversed-phase HPLC in order to sequence individual peptide fragments. Unlike the former method, this procedure is less sensitive but it does allow researchers to circumvent N-terminal protein blockage (see below) and to obtain multiple internal sequences [99].

Nevertheless, Edman sequencing does have some shortcomings. First, this technique is ineffective for high-throughput approaches, since it is very expensive and slow (30 min per amino acid or approximately 5 h per protein [100]). Second, this technique is unsuccessful for the analysis of subpicomole quantities of protein in individual spots, preparative 2-DE gels being required. Nonetheless a combination of Edman degradation with accelerator MS currently enables protein sequencing at attomole level [101]. Finally, the most common failure of this procedure derives from N-terminally blocked proteins, this being a frequent phenomenon in eukaryotic cells. It is due to the presence of formyl, acetyl and pyroglutamyl groups on the N-terminal residues [102]

(ii) Database searching. After microsequence analysis, proteins are identified using sequence database searching algorithms (e.g. BLAST [103], FASTA [104] or BLITZ [105]) that are accessible over the world wide web. The BLITZ program is generally applied for full-length protein sequences, whereas the BLAST and FASTA heuristic algorithms are mainly used for sequence fragments. These alignment search tools permit researchers to match the amino acid sequence generated against the protein databases and/or the DNA databanks dynamically translated in all six reading frames (by the TBLASTN or TFASTX/Y programs). For very short sequences, the ExPASy TagIdent tool allows users to identify a protein from the data about the N- and C-terminal sequence tag of

five amino acids or fewer, its p*I* and molecular mass, and the taxonomic category or species of interest [106].

Although a *C. albicans* genome-sequencing project is currently nearing completion at the Stanford DNA Sequencing and Technology Center, until recently not many *C. albicans* proteins were annotated in the databases [107]. However, using N-terminal sequence tag information, two different approaches for the identification of *C. albicans* proteins not present in databases were proposed. These are based on homology with sequences either from other proteins whose function has already been reported [65] or from other related species with their genomes fully sequenced (such as *Saccharomyces cerevisiae* [108]) [50,63], since the protein *N*-terminus is generally well conserved across species boundaries. At present, a *C. albicans* genomics database (namely CandidaDB) is now freely available through the Internet at http://genolist.pasteur.fr/CandidaDB, and this could be of help in both these sequencing studies and further mass spectrometry-based analyses (see below).

3.2. Mass spectrometry, a key component of the rapid expansion in proteome research

Proteomics-based analyses have undergone a renaissance in recent years. This is a result of recent improvements in mass spectrometric technologies for the systematic study of gene products and is also due to the rapid growth of completed and annotated DNA databases of many model organisms. The introduction of (i) soft ionization methods for converting large, polar and nonvolatile biomolecules to gas-phase ions (ii) very low-flow-rate electrospray source (nanoESI) and microscale capillary reversed-phase HPLC coupled to ESI MS (LC–MS), which permit high sensitivity with low sample consumption, (iii) tandem mass spectrometry for the characterization of unknown proteins, (iv) new peptide extraction methods, and (v) innovative software algorithms able to correlate mass spectrometric information with sequence databases, have revolutionized the area of medical and biological MS for protein and peptide analysis [109]. All these advances have helped towards the successful use of mass spectrometry as the method of choice for dynamic proteome identification and also for the determination of post-translational modifications in proteins.

A mass spectrometer consists of at least an ion source, a mass analyzer and a detector, and measures the mass-to-charge ratio (m/z) of gas-phase ions. Thus, this robust technique, preceded by site-specific proteolysis (usually with trypsin), involves (i) the conversion of peptides to gas-phase ions by soft ionization techniques, such as matrix-assisted laser desorption ionization (MALDI) from the solid state or electrospray ionization (ESI) from the liquid state, (ii) separation of the created ions according to their m/z ratio in a mass analyzer (e.g. time-of-flight (TOF), quadrupole, ion trap, magnetic/electrostatic sector or Fourier-transform ion cyclotron resonance (FT-ICR) analyzers), (iii) the optional fragmentation of selected ion species either by metastable decay (in the postsource decay or PSD technique) or by collision-induced dissociation (CID) performed in tandem mass spectrometry, combining two different analyzers or using the same one twice (see Section 3.2.2), and (iv) measurement of their masses in the detector, obtaining a mass spectrum that displays ion abundance versus m/z values [110,111]. Although several combinations of ionization sources and mass analyzers have been developed, MALDI is typically coupled to a TOF analyzer whereas electrospray ionization is usually combined with triple quadrupole, ion trap or hybrid quadrupole time-of-flight (Q-TOF) MS. However, recent breakthroughs in MS instrumentation have also enabled the development of a MALDI source coupled to a Q-TOF analyzer or to a tandem TOF flight tube [112,113].

Two strategies for protein identification using mass spectrometers have been developed: (i) peptide mapping or peptide mass fingerprinting (PMF) by MALDI-TOF MS (high-throughput) and (ii) peptide fragmentation tagging or partial amino acid sequencing using PSD or CID methods (low-throughput). Both methodologies have recently been used in *C. albicans* proteomics-based research [42,51,52,57,62,65,114]. A general workflow is to analyze the protein sample first by its PMF and, if the protein cannot be identified confidently with this means, further peptide fragmentation analyses are then carried out in order to generate sequence tags. A

schematic representation of this approach for protein characterization by MS is shown in Fig. 5.

3.2.1. Peptide mass fingerprinting: a high-throughput tool for identifying C. albicans proteins present in any of the public domain databases

(i) Mass analysis. Peptide mapping is routinely used for rapid protein identification, typically after SDS–PAGE or 2D-PAGE, and is usually accomplished on MALDI-TOF mass spectrometers. In this approach, the protein sample is in situ digested with a specific endoproteinase (such as trypsin, which cleaves proteins at their arginine or lysine amino acids, if they are not followed by a proline). The peptide mixture is then sequentially extracted from the gel, absorbed on a metal target, co-crystallized with an acidified matrix, and ionized by laser pulses. Subsequently, an electric field accelerates these ions toward the detector, the m/z value of each ion being determined by its time-of-flight to travel from the source to the detector [115].

(ii) Database searching. The peptide mass fingerprint (PMF) of a given protein is therefore the set of peptide masses generated from digesting a pure protein with a sequence-specific protease. Several computer algorithms (such as PEPSEA [116], PROFOUND [117], MASCOT [118], MS-FIT [119], PEPTIDENT/MULTIIDENT [120,121] or MOWSE [122] programs, among others) available on the Internet compare all the experimentally obtained peptide masses with those calculated theoretically from protein or translated genomic databases. Unambiguous protein identification can be achieved if there is a high score of peptide mass matches and protein sequence coverage with a protein in the database.

Unfortunately, a significant number of proteins cannot be conclusively identified or may afford ambiguous database research results by PMF alone, owing to (i) the presence of few tryptic peptides from enzymatic digestion, as a result of inefficient proteolysis or low quantities of protein sample, thus limiting the protein sequence coverage and increasing peptide mass redundancy, (ii) chemical protein modifications (e.g. phosphorylation or glycosylation), differing experimental and theoretical m/z values, (iii) protein mixtures, enhancing the complexity of PMF, and (iv) the nonavailability of full-length gene sequences or of annotated proteins in any of the public domain databases on the organism of interest [16,123]. Despite these limitations, it is possible to avoid the first three ones by using (i) preparative protein gels and/or fluorescent dyes, which do not modify proteins covalently [114], (ii) appropriate database search filters, or the FINDMOD algorithm, in order to characterize potential post-translational modifications, or (iii) the PROFOUND program, which allows the identification of simple protein mixtures. However, until not long ago the latter shortcoming—characteristic of organisms with nonfully sequenced genomes and/or unannotated DNA databases, such as *C. albicans*—could not be circumvented by means of database searches based on homology with peptide mass fingerprints from other cross-species, since tryptic peptide maps are poorly conserved across species boundaries [124] and hence sequence information (peptide sequence tags) is required for this purpose. Accordingly, PMF is the tool of choice for identifying proteins from organisms with completely sequenced genomes, such as *S. cerevisiae* [108].

Accordingly, a small number of *C. albicans* proteins (cytoplasmic and cell wall proteins) collected with 2-DE gels have been identified at a high level of confidence by peptide mapping [42,114]. Conversely, several peptide mass spectra yielded ambiguous results because these proteins were not be present in *C. albicans* databases and/or could display post-translational modifications that change the peptide mass accuracy, and thus the mass spectrum of the modified protein fails to match the theoretical mass spectrum of the unmodified one generated from the database. As described below, in cases where PMF is unsuccessful, sequence information by tandem mass spectrometry is needed.

3.2.2. Peptide sequence tags: a solution for proteins not annotated in the databases or ambiguous protein identifications

(i) Mass analysis. MS/MS yields fragmentation-based peptide sequence information. MS/MS can be undertaken with any ionization source (MALDI or ESI). Nevertheless, in ESI, unlike MALDI, peptide samples do require some purification method after in-gel digestion in order to remove MS-interfering substances (e.g. salts, detergents and/or buffers) [125]. Using either extensive on-line HPLC washing or an off-line procedure (such as minicolumns or

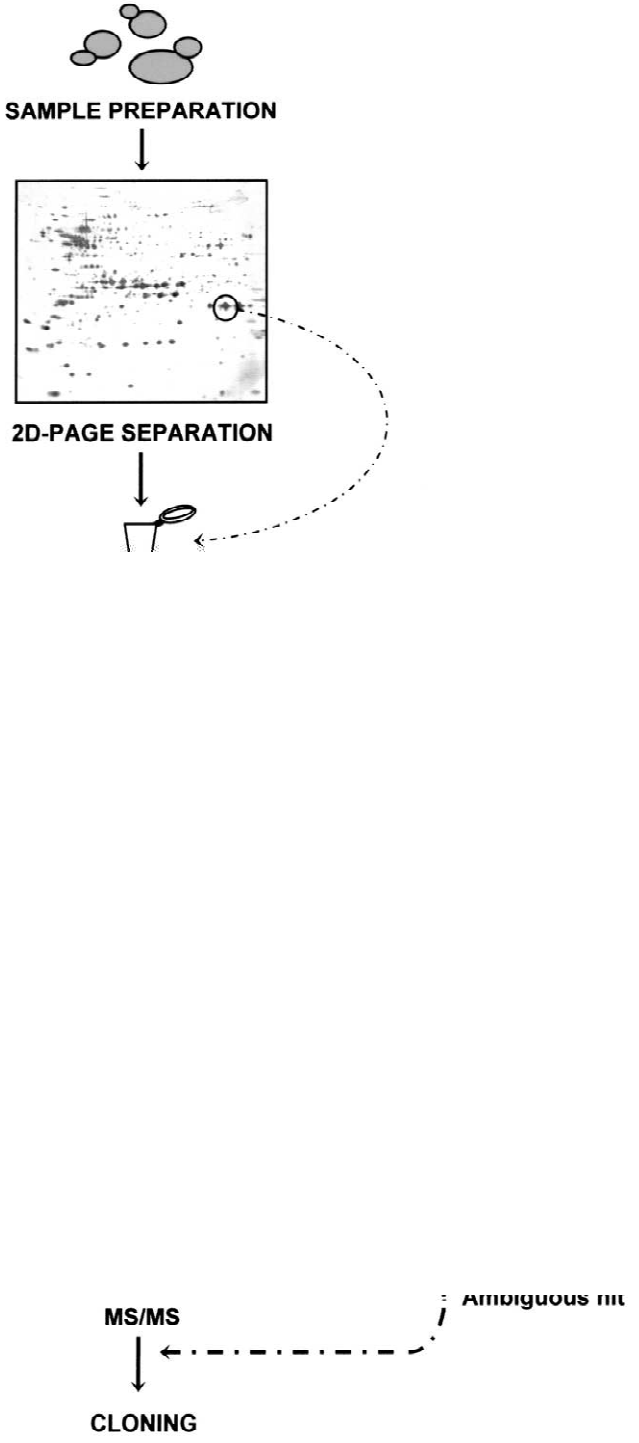

Fig. 5. Schematic representation of the typical approach for characterizing *C. albicans* proteins by mass spectrometry.

microtips filled with reversed-phase material) it is often possible to concentrate and desalt peptide mixtures and remove detergents from them. Alternatively, LC or CE (capillary electrophoresis) can also be coupled to an ESI interface with a view to separating complex peptide mixtures, allowing researchers to perform large-scale proteomic analyses and to circumvent the 2D-PAGE technology (see Section 4). On the other hand, the revolutionary nanoelectrospray ionization system—a miniaturized electrospray source consisting of a gold-coated borosilicate capillary with a internal tip diameter of 1 μm—also enables many sequencing experiments to be carried out on the same sample because nanoESI operates at very low flow-rates (5–10 nl/min) [126].

Regardless of the ionization source used, the peptide ions generated can be selected (generally according the peptide mass data obtained by PMF) and fragmented either by post-source decay (PSD) or by collisional excitation (CID). The masses of the fragment ions are then measured to supply partial sequence information. The PSD method is accomplished on MALDI reflector TOF mass spectrometers (MALDI-PSD MS/MS). However, the PSD spectra are usually difficult to interpret, since the peptide fragmentation patterns are fairly unpredictable. Alternatively, collision-induced dissociation (CID) or collisionally-activated dissociation (CAD) can be exploited in any ion source and mass analyzer. Although a brief outline of some of these mass spectrometers is given below, readers are referred to [112,125,127,128] for further information about this topic. (i) In triple quadrupole mass spectrometers (coupled to an ESI source), the first quadrupole analyzer selectively separates the ion of interest, this being dissociated by activation in the collision cell filled with an inert gas, such as argon. The third quadrupole measures the m/z values of dissociation products. (ii) Hybrid quadrupole-TOF mass spectrometers (Q-TOF MS) combine a selection quadrupole, a collision hexapole and a TOF fragment ion analyzer. (iii) In ion-trapping mass spectrometers, the undesired ions are ejected from the trap. Subsequently, the isolated ion is fragmented and the ion products are analyzed. This type of MS allows proteome researchers to carry out multiple fragmentation reactions (MS^n) and thus the de novo sequencing of proteins. Finally, (iv) in tandem TOF instruments (equipped with a MALDI source), the ions fragmented in the CID cell are reaccelerated and analyzed by the second TOF analyzer.

Most studies about the peptide fragmentation of *C. albicans* protein samples from 2-DE gels have been based on the use of mass spectrometers with an ESI interface. As summarized in Fig. 4, tryptic peptides are previously desalted and concentrated either in minicolumns (filled with a Poros R2 perfusion sorbent (PerSeptive Biosystems, Framingham, MA, USA) [51] or with a few reversed-phase C_{18} beads (Vydag and Millipore) [52,65]) or using on-line LC [62]. Afterwards, the peptide sequence analyses are implemented by means of (i) electrospray ionization triple–quadrupole tandem mass spectrometers (ESI-MS/MS) [57], (ii) nanoelectrospray ionization followed by low-energy collision-induced dissociation (CID) performed on triple quadrupole [51], hybrid quadrupole time-of-flight [65], or ion trap [52] mass analyzers (nanoESI-MS/MS, nanoESI-QTOF-MS or nanoESI-IT-MS, respectively), or (iii) liquid chromatography–ion trap mass spectrometers (LC–IT-MS) [62].

(ii) Database searching. MS/MS spectra can provide structural information since the difference in mass between two consecutive ion-fragments of the same type (i.e. y or b series ions, according to charge location), corresponding to the loss of an amino acid residue, affords the identity and position of that amino acid in the peptide. It is therefore possible to acquire the complete peptide sequence through the data from any good quality tandem mass spectrum. Typically, three different database searching strategies have been developed for interpreting the fragment mass spectra measured, and thus for identifying the protein of interest [115,129]. Several bioinformatic tools (such as PEPSEA [116], SEQUEST [130,131], PEPFRAG [132], MS-TAG [119] and MASCOT [118] algorithms) can be employed for this purpose.

In the first approach, similar to peptide mass fingerprinting database searching, the experimental tandem mass spectrum is compared with the theoretical fragmentation spectra derived from protein or translated genomic databases, using diverse search algorithms (see above). Evidently, unambiguous protein identification will be attained if a high number of peptides match the same entry in a database. One of the main advantages of using this

procedure is that, unlike the ones to be described below, it does not involve the interpretation of MS/MS spectra.

The second search procedure, called peptide mass tag searching, requires the partial interpretation of CID spectra. The information about a partial amino acid sequence, the peptide mass (parent mass), and the exact location of the sequence in the peptide (the masses of the fragments at the start and at the end of the interpreted region) constitutes the sequence tag and this is sufficient to find the protein from which it comes in databases [116]. Proteolytic cleavage specificity can also be included in the search. Sequence tags can be used to query the databases that contain incomplete gene information—such as expressed sequence tag (EST) databases—to identify the protein from which the peptides are derived [133].

Finally, the de novo sequencing method is the strategy of choice for organisms that have incomplete or unannotated DNA databases, such as—until a short time ago—*C. albicans*. The goal of this approach is to obtain several complete amino acid sequences deduced de novo from the MS/MS spectra (either with computer assistance or by manual interpretation), in an attempt to carry out sequence similarity searches against well-defined protein sequence databases [134,135]. This procedure therefore allows researchers to find known homologous proteins in other species using sequence-database search programs, such as BLAST or FASTA tools. In addition, de novo sequence analysis algorithms in homology mode can be used to identify highly homologous proteins in the databases.

The latter MS/MS data-searching strategy has been used for the identification of 2D-PAGE-separated proteins from *C. albicans* that were not present in the public domain databases [51,52,62,65]. Drawing on the high degree of similarity between *S. cerevisiae* and *C. albicans* sequences, unambiguous cross-species identifications of some of the proteins of this fungal pathogen were attained. The information from CID spectra was used to query the *S. cerevisiae* databases to identity the *S. cerevisiae* counterparts of the *C. albicans* proteins. Subsequently, the assigned homologous protein sequence was matched with the unannotated *C. albicans* genome database (http://www-sequence.stanford.edu/group/candida) using the TBLASTN or TFASTX/Y bioinformatic tools in order to confirm whether the corresponding contigs or DNA fragments—predicted from the *C. albicans* genome sequence information—encode proteins with these experimentally determined peptides.

Nevertheless, if the protein of interest cannot be unambiguously identified by the latter method or if it is assigned as an unknown protein, the sequence information can then be used as a basis for cloning the corresponding gene [57,136]. MS/MS spectra therefore provide information for sequencing and/or cloning a given protein.

4. News corner

Presently, emerging non-gel-based proteomic technologies, such as 2-D liquid chromatography or capillary isoelectric focusing coupled to MS [137–139], isotope-coded affinity tags [140] and protein arrays [141], herald new opportunities for rapid and high-throughput analyses of the final gene expression products. Although these innovative tendencies are hitherto unreported in the *C. albicans* literature, below we briefly outline the main contributions that might be exploited for this pathogen in the future.

In the quest to bypass some gel-associated limitations in protein separation and characterization, Link et al. and Washburn et al. [137,138] have described a new non-gel-based multidimensional approach using two-dimensional liquid chromatography (a strong cation-exchange and reversed-phase capillary column), in conjunction with on-line mass spectrometry (SCX-RP–ESI-MS/MS) for the analysis of large protein complexes (multidimensional protein identification technology or MudPIT). This attractive strategy involves protein digestion prior to sequential peptide separation: first by their electrostatic charge and then by their hydrophobicity. Likewise, O'Connor et al. [139,142] have developed a system that separates peptides in liquid phase according to their p*I* values by capillary isoelectric focusing. The peptides are subsequently eluted directly into an ESI-MS for detection and characterization (cIEF–ESI-MS). All these MS-based methods allow the resolution of thousands of proteins and their rapid identification in a single experiment.

New quantitative proteomic technologies are im-

proving measurements of differential protein expression among different samples (see Fig. 6). Gygi et al. [140] have reported an exciting method that employs isotope-coded affinity tags or ICATs to determine the relative amount of protein in two given samples. To this end, the cysteine side chains of proteins from each cell growth condition are tagged with either a light or a heavy ICAT reagent (a normal or deuterated label, respectively). Subsequently, both protein samples are mixed and digested. ICAT-labeled peptides are affinity-isolated and analyzed by liquid-chromatography coupled to MS/MS. The relative intensity of identical peptides from each sample reveals the relative abundance of the protein from which they derive. In addition, the MS/MS spectrum yields its identity (Fig. 6A). In an alternative approach, Tonge et al. [143] have recently reported a fluorescence 2-D differential gel electrophoresis procedure (2-D DIGE) aimed at comparing different samples on the same gel. In this approach, each protein sample is previously labeled with different fluorophores before mixing them and resolving them on the same 2-DE gel (Fig. 6B). Thus, both strategies minimize the time-consuming analysis and enhance the degree of precision in quantitative comparisons of protein expression profiles.

In the near future, the current development of the protein array-based technology may challenge the course of proteomic research, since these protein arrays can simultaneously analyze thousands of parameters in a single experiment. These microarrays may therefore be exploited in high-throughput screenings for the analysis of expression profiling, the study of molecular interactions (e.g. antibody–antigen, other protein–protein or protein–small molecule interactions), the discovery of biomarkers and/or functional annotations of gene products [144–147]. To this end, protein mixtures are separated on chromatographic-based (e.g. ion-exchange, reversed-phase and hydrophobic, among others) or biologically active (e.g. antibodies, receptors, DNA, and other ligands) surface arrays, such as the SELDI ProteinChip (surface-enhanced laser desorption/ionization) [145] or SPR-BIA chip (surface plasmon resonance–biomolecular interaction analysis) [148]. Then, the proteins retained or affinity-captured on the chip surface are detected directly and identified by MALDI-TOF-MS.

4.1. C. albicans proteomics: a boundless platform?

There is no doubt that the 21st century will witness noteworthy progress in the analysis of microbial proteomes, as a result of (i) the completion of genome sequencing for several microorganisms and the subsequent emergence of well-annotated proteomic databases, and (ii) improvements in the limitations of current proteomic technologies and/or the development of further high-throughput innovations. The main challenge of the postgenomic era will therefore be the characterization of proteomes of biological systems (including gene expression, protein identification, post-translational modifications, protein–protein interactions and cell phenotypes). Accordingly, proteomics provides new insights for biological and medical researchers due to its infinite biomedical opportunities, such as the identification of new therapeutic targets and diagnostic markers and/or the elucidation of the pathogenic mechanisms involved [149,150].

At present, proteomics-based studies applied to *C. albicans* are still in their infancy. The sequencing of the *S. cerevisiae* genome [108] has significantly supported these analyses. *S. cerevisiae* protein data have therefore been employed for the identification of certain *C. albicans* proteins owing to the high degree of homology between most sequences from both yeast organisms. However, a *C. albicans* genome sequencing project is currently almost completed at the Stanford Genome Technology Center (http://www-sequence.stanford.edu/group/candida). The availability of the complete *C. albicans* genome in the foreseeable future will undoubtedly facilitate the systematic characterization of its gene products and the creation of *C. albicans* proteome databases. Recently, our research team is developing an annotated *C. albicans* 2D-PAGE database dealing with subcellular proteomes, accessible via the world wide web at http://www.expasy.ch/ch2d/2d-index.html [107]. It may be useful for comparing 2-DE patterns of similar samples obtained at different laboratories with the reference one, available on the Internet, and for suggesting protein identifications.

Preliminary proteomic research performed on this human fungal pathogen has been directed towards the understanding of dimorphism, host responses, the cell wall, virulence factors, drug resistance and other

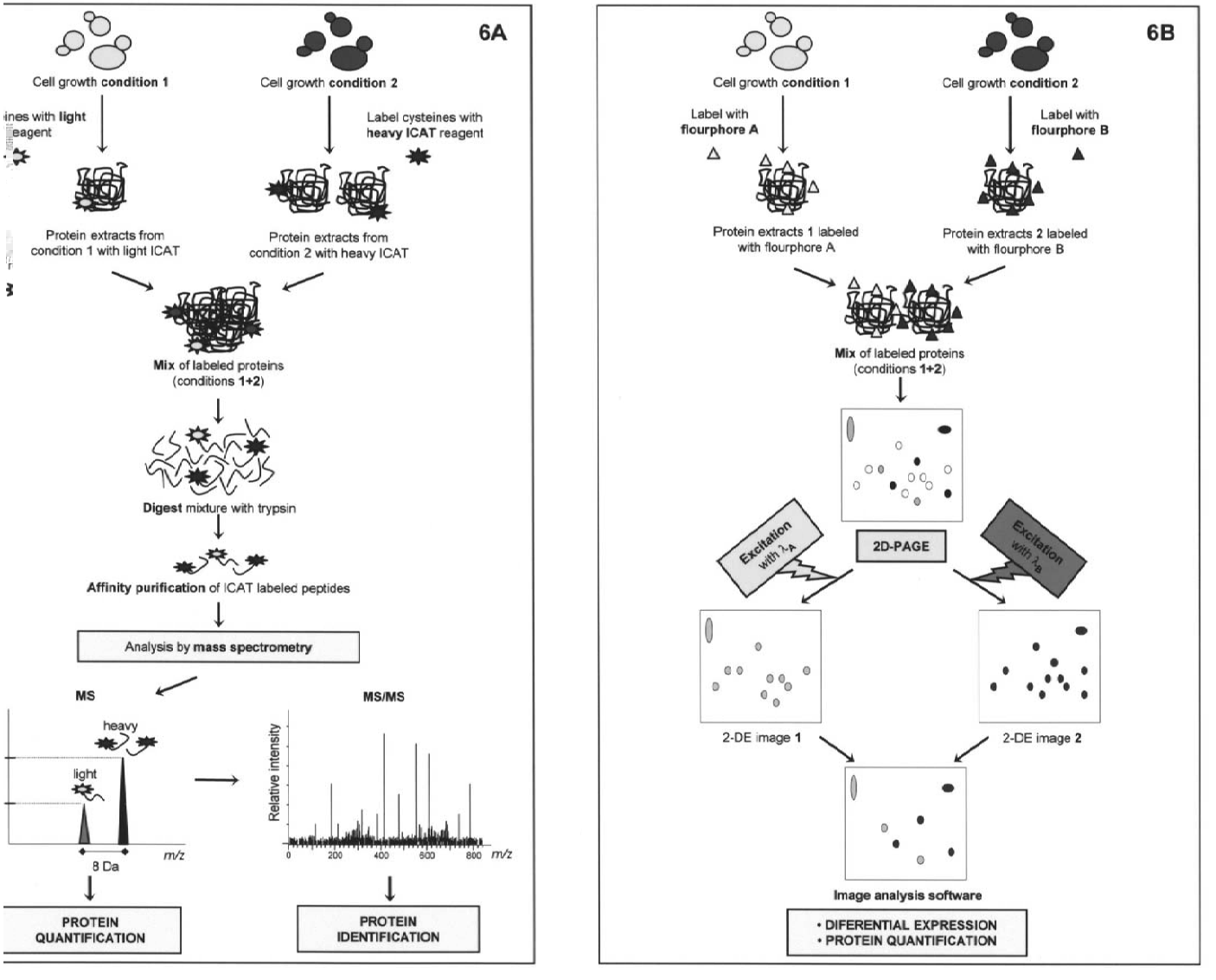

Fig. 6. Schematic illustration of two novel approaches for the detection and quantification of differential protein expression (quantitative proteome analysis). (A) ICAT strategy. To simplify the representation, a single protein is shown. In the example, this protein is expressed more strongly in cell growth condition 2 than in condition 1, the ratio being 2:1. (B) 2-D DIGE proteomic technology.