

## Utilização da microespectroscopia infravermelha (FT-IR) para teste de algoritmos estatísticos na diferenciação dos micro-organismos *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis*

Patrícia Marcondes dos Santos\*, Maria Angélica Gargione Cardoso, Sônia Khouri, Alderico Rodrigues de Paula Júnior, Mituo Uehara†, Kumiko Koibuchi Sakane

**Resumo** As espécies do gênero *Candida* são causadoras de diversas infecções fúngicas e, nos últimos anos, tem sido desenvolvidas novas tecnologias para auxiliar nos diagnósticos microbiológicos. Dentre as técnicas está a espectroscopia infravermelha junto com a análise estatística multivariada. O objetivo deste trabalho é comparar dois métodos: estatístico (análise multivariada) e não-estatístico (ajuste de curva), utilizando os espectros infravermelhos de *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis* para testar o potencial do uso de Análise Estatística Multivariada para discriminação de espectros de micro-organismos. Para isso foram obtidos, utilizando o Spectrum Spotlight 400 da PerkinElmer, 54 espectros infravermelhos, sendo 18 de cada espécie, na faixa de 4000 a 1000  $\text{cm}^{-1}$ , com resolução de 4  $\text{cm}^{-1}$ , no modo de transmissão, a 20 °C. A análise dos espectros foi realizada através de três métodos: (1) inspeção visual direta dos espectros; (2) análise estatística multivariada; (3) ajuste de curva para a determinação de estruturas secundárias de proteínas. Na região de 1200 a 1000  $\text{cm}^{-1}$ , os espectros apresentam diferenças que podem ser percebidas numa inspeção visual direta. Uma banda próxima de 1070  $\text{cm}^{-1}$  e outra próxima de 1045  $\text{cm}^{-1}$  apresentam intensidades relativas diferentes para os três espectros. Por outro lado, as bandas da amida I, na região de 1710 a 1590  $\text{cm}^{-1}$ , apresentam aspectos visuais semelhantes com máximo em 1651  $\text{cm}^{-1}$  para os espectros dos três micro-organismos. Esse fato torna possível submeter a análise estatística multivariada a um teste de sua capacidade de diferenciar três espectros de *Candida*. A análise estatística multivariada foi aplicada aos 54 espectros para investigar as regiões de 4000 a 1000  $\text{cm}^{-1}$  com exceção da região de 2600 a 2300  $\text{cm}^{-1}$  e de 1710 a 1590  $\text{cm}^{-1}$  que corresponde a das bandas da amida I. A técnica selecionada foi a análise por componentes principais (PCA, *Principal Componente Analysis*), utilizando os primeiros quatro componentes principais, em conjunto com a técnica hierárquica de análise de agrupamento (HCA, *Hierarchical Clustering Analysis*) segundo o método de Ward. Foi utilizado para esta análise o software MINITAB 15 e o resultado mostra uma clara discriminação dos espectros dos três micro-organismos nas duas regiões consideradas. Adicionalmente foi obtido o espectro médio de cada micro-organismo nas bandas da amida I na região de 1710 a 1590  $\text{cm}^{-1}$ . Os três espectros médios assim obtidos foram analisados pelo método de ajuste de curva que não é estatístico para determinar as estruturas secundárias de proteínas. Para esta análise o software ORIGIN 7.5 foi utilizado e os resultados obtidos mostram estruturas conformacionais diferentes nos três micro-organismos. Esses resultados confirmam a discriminação obtida através da análise estatística multivariada e visual. Pode-se concluir que as análises estatísticas multivariadas baseadas em análise por componentes principais e análise de agrupamento com uso do algoritmo Ward é potencialmente útil para discriminar micro-organismos através de seus espectros infravermelhos. Além disso, as análises mostram que as bandas da amida I dos espectros infravermelhos de *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis* fornecem um conjunto de dados cuja estrutura de agrupamento é conhecida e que pode ser útil para testar e validar algoritmos estatísticos de análise de agrupamento.

**Palavras-chave** *Candida albicans*, *Candida dubliniensis*, *Candida parapsilosis*,  
Análise estatística multivariada, Microespectroscopia infravermelha,  
Estruturas secundárias de proteínas.

## ***Using FT-IR microspectroscopy for testing statistical algorithms for differentiation of *Candida albicans*, *Candida dubliniensis* and *Candida parapsilosis****

**Abstract** *Films of *Candida albicans*, *Candida dubliniensis* and *Candida parapsilosis* were prepared and the infrared spectra of these films were obtained in the region 4000 to 1000  $\text{cm}^{-1}$ , with resolution of 4  $\text{cm}^{-1}$ , in the transmission mode, at 20 °C. Fifty four spectra were obtained, 18 of each microorganism, with the PerkinElmer Spotlight 400 FT-IR, which has a microscope attached to a FT-IR spectrophotometer. The spectra were analyzed through three methods: (1) mere visual inspection; (2) multivariate statistical analysis; (3) curve-fitting for determining secondary structures of proteins. In the region 1200 to 1000  $\text{cm}^{-1}$ , the spectral bands show differences that can be seen by a mere visual inspection. On the other hand, the amide I bands, in the region 1710 to 1590  $\text{cm}^{-1}$ , have the same visual aspect for the three microorganisms. Multivariate statistical analysis was applied to analyze these amide I bands of all the 54 spectra. Principal component analysis (PCA) and techniques of hierarchical cluster analysis (HCA, Hierarchical Clustering Analysis) according to Ward's method were applied using the software MINITAB 15. The results show a clear discrimination of the three microorganisms. The average spectrum of each microorganism was obtained in the amide I band. Each average spectrum was analyzed by curve-fitting for the determination of secondary structures of proteins. The software used was the ORIGIN 7.5 and the results confirm the discrimination obtained through multivariate statistical analysis. This result shows that multivariate statistical analysis can be useful to discriminate infrared spectra of different microorganisms. Furthermore, this work shows that the amide I bands of the infrared spectra of *Candida albicans*, *Candida dubliniensis*, and *Candida parapsilosis* provide a set of data of known group structure that can be useful to test statistical algorithms of cluster analysis.*

**Keywords** *Candida albicans, Candida dubliniensis, Candida parapsilosis, Multivariate statistical analysis, Infrared microespectroscopy, Secondary structures of proteins.*

## Introdução

As espécies do gênero *Candida* são importantes agentes infecciosos em pacientes hospitalizados. A maioria dos sistemas de identificação disponível, em hospitais, baseia-se nas características fenotípicas dos micro-organismos. Entretanto, esses sistemas são geralmente demorados, apresentando acurácia limitada devido às dificuldades na leitura e interpretação dos resultados (Toubas *et al.*, 2007).

A *Candida albicans* pode ser considerada a mais importante espécie envolvida em infecções fúngicas, embora as espécies não *albicans*, como *Candida parapsilosis*, *Candida dubliniensis* vem aumentando em importância nos últimos anos (Celebi *et al.*, 2008).

Em estudo realizado por Mähns *et al.* (2005), foram comparados diferentes métodos fenotípicos para identificação de *Candida* spp, e concluíram que não existe um único método capaz de diferenciar todas as espécies, principalmente quando elas possuem características semelhantes, como é o caso de *Candida albicans* e *Candida dubliniensis*.

Nos últimos anos, novas tecnologias têm sido desenvolvidas para auxiliar o diagnóstico microbiológico de determinadas doenças infecciosas emergentes, principalmente as fúngicas, como as análises imunológicas, bioquímicas e moleculares e mais recentemente as análises baseadas em técnicas por espectroscopia infravermelha em conjunto com a análise estatística multivariada (Beekes *et al.*, 2007; Dziuba *et al.*, 2007; Essendoubi *et al.*, 2005; Maquelin *et al.*, 2002; Naumann, 2000; Sandt *et al.*, 2006; Timmins *et al.*, 1998; Toubas *et al.*, 2007).

A espectroscopia de absorção no infravermelho é uma das técnicas que fornece informações sobre os modos de vibração molecular, que por sua vez dependem da estrutura da molécula.

As vibrações moleculares envolvem energias que correspondem a energias de fótons infravermelhos. Na espectroscopia molecular, tais fótons podem ser absorvidos pela molécula que está sendo objeto de estudo, excitando modos de vibração da molécula, resultando, dessa interação entre fótons e molécula, espectros de absorção na região do infravermelho. Assim, pode-se investigar a geometria e as forças de interação entre os átomos que constituem a molécula, pois seus espectros dependem desses fatores.

O espectro de absorção no infravermelho é um processo que ocorre quando uma radiação eletromagnética incidente tem uma componente com valor de frequência igual à frequência correspondente a uma transição entre dois níveis vibracionais. Somente os modos de vibração que levam a variação no

momento dipolar elétrico são observados nos espectros infravermelhos (Colthup, 1997; Sala, 1995).

Há diversos métodos para se realizar a análise estatística multivariada dentre os quais está a análise por componentes principais (*PCA, Principal Componente Analysis*) e análise por agrupamento (*HCA, Hierarchical Clustering Analysis*) com o uso do algoritmo *Ward*. Este algoritmo busca formar grupos minimizando o desvio padrão entre os dados de cada grupo (Everitt, 1994; Ward, 1963). A aplicação de diferentes algoritmos estatísticos em uma análise pode levar a resultados discrepantes. Portanto, é importante comparar os resultados obtidos com a utilização do algoritmo estatístico com os resultados obtidos por métodos não-estatísticos (Manly, 2005). A técnica de ajuste de curvas para a determinação de estruturas secundárias de proteínas é um método não-estatístico que tem sido aplicado para diferenciar espectros infravermelhos de *Candida albicans* nas formas de levedura e filamentososa (Adt *et al.*, 2006).

O objetivo deste trabalho é comparar os dois métodos: estatístico e não-estatístico, utilizando os espectros infravermelhos de *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis* para testar o potencial do uso de Análise Estatística Multivariada para discriminação de espectros de micro-organismos.

## Materiais e Métodos

### Preparação das amostras para espectroscopia no infravermelho

Foram utilizadas cepas padrão (*ATCC – American Type Culture Collection*) de *Candida albicans* (*ATCC 10231*), *Candida dubliniensis* (*ATCC 778157*) e *Candida parapsilosis* (*ATCC 22019*) mantidas no Laboratório de Microbiologia da Faculdade de Ciências da Saúde – UNIVAP, respeitando as normas de Biossegurança.

As cepas padrão foram cultivadas em meios específicos para isolamento e identificação (Agar Sabouraud Dextrose – *DIFCO*) e incubadas por 24 a 48 horas a 35 °C, conforme protocolo de rotina. Em seguida, foram preparados inóculos de acordo com escala 0,5 de *Mc Farland*, que corresponde a  $1,5 \times 10^8$  células/mL. Uma alíquota (300 µL), dos inóculos previamente preparados, foi semeada em superfície com alça de *Drigalski* (Koneman *et al.*, 2001), em meio de cultura Agar Sabouraud Dextrose (*DIFCO*) e incubada por 12 horas a 37 °C. Decorrido o tempo de incubação, cerca de duas alças, de 1 µL, de colônias de leveduras foram ressuspensas em 40 µL de solução fisiológica estéril (0,9 %). A suspensão foi homogeneizada em Vortex por 1 minuto. Em

seguida, uma alíquota de 10  $\mu\text{L}$  foi transferida para a janela de fluoreto de cálcio ( $\text{CaF}_2$ ) e mantidas em estufa a 50  $^\circ\text{C}$  por uma hora, para obtenção de um filme transparente.

A solubilidade de fluoreto de cálcio em água é pequena, 0,0015 g/100 mL (18  $^\circ\text{C}$ ) e 0,0016 g/mL (20  $^\circ\text{C}$ ) (Budavari, 1996). Alguns autores tais como Barbosa (2007), Stuart (1997) e Toyran *et al.* (2004) consideram o fluoreto de cálcio insolúvel em água. No nosso experimento, foram utilizadas as janelas de  $\text{CaF}_2$  pelo tempo de contato de poucos minutos contando a partir da colocação da solução na janela até ser colocada na estufa. Além disso, os espectros obtidos de *C. albicans* conferem com os publicados na literatura (Adt *et al.*, 2006; Toubas *et al.*, 2007).

### Espectroscopia no infravermelho e pré-processamento dos dados

Os espectros foram obtidos na faixa de 4000 a 1000  $\text{cm}^{-1}$ , com resolução de 4  $\text{cm}^{-1}$ , no modo de transmissão com 32 varreduras, com intervalo de 1  $\text{cm}^{-1}$ , a 20  $^\circ\text{C}$ . Foi utilizado Spectrum Spotlight 400 FT-IR da PerkinElmer, que tem um microscópio acoplado a um espectrofotômetro infravermelho com transformada de Fourier.

Os dezoito espectros de cada espécie de *Candida* foram pré-processados com software Spectrum 5.3 (PerkinElmer), onde foram feitas as correções de linha de base, suavização espectral utilizando o algoritmo *Savitzky Golay* (9 pontos) e normalização. Os espectros médios de cada espécie são mostrados na Figura 1.

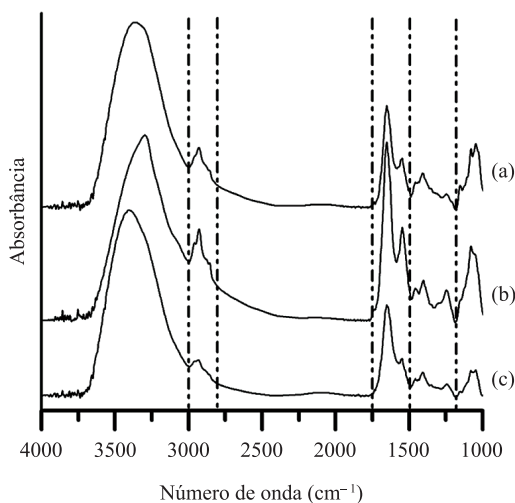


Figura 1. Espectros médios de (a) *Candida albicans*, (b) *Candida dubliniensis* e (c) *Candida parapsilosis*.

Figure 1. Average spectra of (a) *C. albicans*, (b) *C. dubliniensis* and (c) *C. parapsilosis*.

### Análise dos dados

Neste trabalho os espectros infravermelhos de *C. albicans*, *C. dubliniensis* e *C. parapsilosis* foram analisados através de três métodos: (1) inspeção visual direta dos espectros; (2) análise estatística multivariada; (3) ajuste de curva para o estudo das estruturas secundárias de proteínas.

Para a análise visual foi calculado o espectro médio de cada tipo de cepa utilizando 18 espectros pré-processados. Os espectros médios de cada tipo de cepa foram impressos na forma de gráfico em uma escala e faixa apropriada para ressaltar as diferenças entre eles.

Os algoritmos utilizados para a análise estatística multivariada foram a Análise de Componentes Principais (PCA) juntamente com a Análise de Agrupamento Hierárquico (HCA). O PCA consiste em projetar os parâmetros espectrais originais em novos eixos de forma que o primeiro eixo fica na direção de maior variabilidade dos parâmetros. Os demais eixos são perpendiculares aos eixos anteriores sempre na direção de maior variabilidade remanescente. As novas variáveis são denominadas de componentes principais e são os autovetores da matriz de covariância dos parâmetros originais. Os autovalores desta matriz representam variâncias das novas variáveis. Para aplicações em espectroscopia, em geral, os primeiros principais componentes (três a cinco) captam mais de 95% da variabilidade dos parâmetros originais e os demais podem ser considerados como ruído e desprezados. Isto permite uma grande redução da dimensionalidade de parâmetros o que facilita a fase de classificação (Johnson e Wichern, 1992). Após a redução da dimensionalidade, algoritmos de classificação podem ser utilizados. Um método que permite visualizar os dados é o método HCA que procura agrupamento de dados semelhantes utilizando alguma medida de distância entre os dados. Neste trabalho foi selecionado o algoritmo de *Ward*. Para visualizar os resultados da análise de agrupamento pode-se utilizar o dendrograma também disponível em diversos softwares, dentre eles, o MINITAB 15, utilizado neste trabalho.

Aplicamos o método de ajuste de curvas para determinar as estruturas secundárias de proteínas, na região de 1710 a 1590  $\text{cm}^{-1}$ , correspondente à banda da amida I. Nesse método supõe-se que a banda da amida I seja a soma de várias componentes lorentzianas. A função lorentziana,  $L(x)$ , é definida por:

$$L(x) = \frac{2A}{\pi} \frac{w}{4(x - x_0)^2 + w^2} \quad (1)$$

onde  $x_0$  dá a posição do pico da função,  $w$  é um parâmetro que especifica a largura e  $A$  é a área total sob a curva a partir da linha de base. O valor de pico,  $L(x_0)$ , é dado por:

$$L(x_0) = \frac{2}{\pi w} \quad (2)$$

de modo que quanto menor for a largura  $w$ , maior será o valor de pico.

A derivada segunda da banda da amida I fornece os picos das componentes lorentzianas (Byler e Susy, 1986). As larguras das componentes lorentzianas são ajustadas, pelo método dos mínimos quadrados, para que a soma das componentes lorentzianas reproduza, o melhor possível, a banda original. Para o cálculo da derivada segunda e o ajuste pelo método dos mínimos quadrados foram utilizados os softwares Spectrum 5.3 e ORIGIN 7.5 respectivamente. Alguns autores utilizam gaussianas em lugar de lorentzianas para o ajuste da banda da amida I (Toyran *et al.*, 2004).

As áreas de componentes das bandas de amida I são relacionadas às quantidades de estruturas conformacionais. As porcentagens das conformações  $\alpha$ -hélice,  $\beta$ -folha, ao acaso e dobras/voltas são estimadas calculando a razão entre a área de cada componente e a área total que é a área de todos os componentes (Stuart, 1997; Surewicz e Mantsch, 1988).

A Tabela 1 mostra os parâmetros do espectrofotômetro e softwares utilizados no trabalho.

## Resultados e Discussões

A Figura 1 apresenta o espectro infravermelho médio dos três micro-organismos analisados.

O espectro de *Candida albicans* está de acordo com o publicado na literatura. (Adt *et al.*, 2006; Toubas *et al.*, 2007). Essendoubi *et al.* (2005) utilizaram a segunda derivada dos espectros infravermelhos de

**Tabela 1.** Tabela dos parâmetros do espectrofotômetro e softwares utilizados.

**Table 1.** Table of parameters of the spectrophotometer and softwares.

Obtenção dos espectros infravermelhos	
Espectrofotômetro	Spectrum Spotlight 400 FT-IR (PerkinElmer)
Resolução	4 cm <sup>-1</sup>
Modo	transmissão
Número de varreduras	32
Intervalo	1 cm <sup>-1</sup>
Temperatura de operação	20 °C
Análise de dados	
Softwares utilizados	Spectrum 5.3; ORIGIN 7.5; MINITAB 15

espécies *Candida*, entre eles os de *Candida albicans* e *Candida parapsilosis* para rápida identificação utilizando análise de cluster hierárquica. Tintelnot *et al.* (2000) compararam duas espécies fortemente relacionadas: *Candida albicans* e *Candida dubliniensis* utilizando espectroscopia FT-IR combinada com análise de cluster hierárquico. Timmins *et al.* (1998) notaram uma diferença muito pequena nos espectros infravermelhos de micro-organismos estudados, entre eles: *Candida albicans* e *Candida dubliniensis*. Suas bandas são muito largas com contornos muito complexos, o que tornam sua discriminação difícil por inspeção visual rápida e sugerem como o método para discriminação, a análise estatística multivariada.

A Tabela 2 apresenta as regiões espectrais em que se encontram bandas características de biomoléculas e as interpretações aproximadas dessas bandas (Mantsch e Chapman, 1996; Maquelin *et al.*, 2002; Twardowski e Anzenbacher, 1994).

A Figura 1 mostra, conforme as interpretações dadas na Tabela 2, que os espectros das três espécies de *Candida* apresentam predominância das bandas de proteínas e de carboidratos.

### Análise dos espectros infravermelhos

São apresentadas a seguir as análises dos espectros obtidos a partir de três métodos distintos: (1) inspeção visual direta; (2) análise estatística multivariada (*MVA*, *Multivariate Statistical Analysis*); (3) ajuste de curva para analisar a banda da amida I, na região de 1710 a 1590 cm<sup>-1</sup>.

**Tabela 2.** Bandas características no infravermelho de biomoléculas.  
**Table 2.** Characteristic infrared bands of biological specimens.

Número de onda (cm <sup>-1</sup> )	Descrições aproximadas
4000-3000	estiramento da ligação O-H do grupo hidroxila; estiramento da ligação N-H de amida A de proteínas.
3000-2800	estiramento dos grupos -CH <sub>3</sub> e >CH <sub>2</sub> dos ácidos graxos.
1750-1500	estiramento da ligação C=O do grupo ester e de ácidos nucléicos; amidas I e II de proteínas.
1500-1200	deformação angular dos grupos >CH <sub>2</sub> e -CH <sub>3</sub> dos ácidos graxos; estiramento simétrico de C=O de COO <sup>-</sup> ; amida III de proteínas; estiramento assimétrico da ligação P=O dos compostos com grupo fosfato.
1200-1000	estiramento das ligações C-O, C-C; deformações angulares de C-O-H, C-O-C de carboidratos presentes na parede celular; estiramento simétrico da ligação P=O do grupo fosfato.

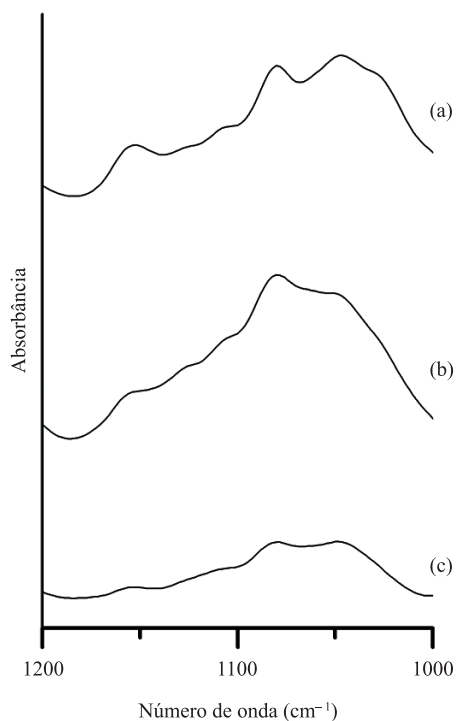


### Inspeção visual direta

A região de 1200 a 1000  $\text{cm}^{-1}$  corresponde a bandas de absorção de carboidratos presentes na parede celular dos três micro-organismos.

A parede celular é uma estrutura rígida que mantém a integridade da célula e interage com o meio que se encontra. É localizada na parte externa da membrana celular e é composta principalmente de glucanas e mananas e de pequenas quantidades de quitinas e lipídios. As glucanas e mananas se combinam formando as glicoproteínas, manoproteínas e glicomanoproteínas (Chaffin, 2008; Gompertz *et al.*, 2008). Manoproteínas se encontram na superfície externa da célula e sua quantidade varia exponencialmente entre as fases estacionárias e de crescimento das células e também conforme as condições do meio, tais como nutrientes e disponibilidade de oxigênio (Chaffin, 2008).

Na região de 1200 a 1000  $\text{cm}^{-1}$  observam-se duas bandas nos três espectros, sendo uma próxima de 1070  $\text{cm}^{-1}$  e outra próxima de 1045  $\text{cm}^{-1}$ . Adt *et al.* (2006) atribuíram essas bandas a vibrações das ligações C-O-C do anel de polissacarídeos. A primeira corresponde a do glicogênio e dos ácidos nucleicos e a segunda de manoproteínas. As intensidades relativas dessas duas bandas são diferentes para os três espectros, mostrados na Figura 2.



**Figura 2.** Espectros médios de (a) *Candida albicans*, (b) *Candida dubliniensis* e (c) *Candida parapsilosis* na região de 1200-1000  $\text{cm}^{-1}$ .  
**Figure 2.** Average spectra of (a) *C. albicans*, (b) *C. dubliniensis* and (c) *C. parapsilosis* in the region 1200-1000  $\text{cm}^{-1}$ .

A inspeção visual direta permite distinguir os espectros infravermelhos das três espécies de *Candida*. Esse fato torna possível submeter a análise estatística multivariada a um teste da sua capacidade de diferenciar espectros de micro-organismos.

### Análise estatística multivariada

A análise estatística multivariada tem sido utilizada para o estudo de espectros de micro-organismos, com vistas a uma rápida diferenciação desses micro-organismos (Beekes *et al.*, 2007; Dziuba *et al.*, 2007; Essendoubi *et al.*, 2005; Maquelin *et al.*, 2002; Naumann, 2000; Sandt *et al.*, 2006; Timmins *et al.*, 1998; Toubas *et al.*, 2007).

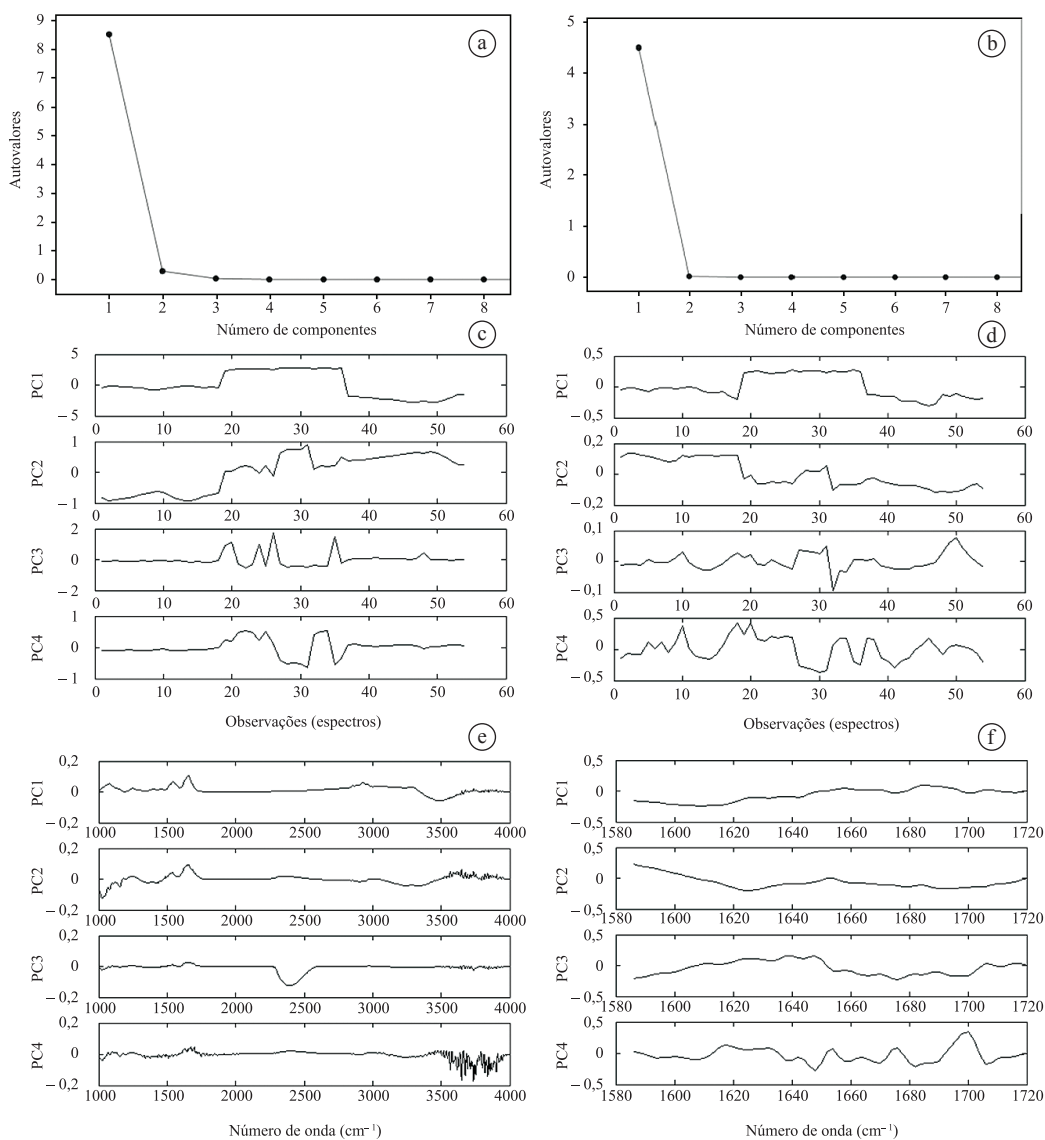
Análises estatísticas multivariadas baseadas em análise por componentes principais (PCA, *Principal Component Analysis*) e análise de agrupamento (*HCA*, *Hierarchical Clustering Analysis*) foram utilizadas para analisar os espectros dos três micro-organismos. Para as análises de agrupamento foram utilizados os quatro primeiros componentes principais, por conter aproximadamente 99% das informações iniciais, conforme mostra pelos *Scree plot*, Figuras 3a, b. Esse alto grau de representação em um pequeno número de componentes demonstra uma grande correlação entre os dados analisados. Pelos gráficos dos *Score*, Figuras 3c, d, pode-se confirmar a grande contribuição do primeiro componente em separar os espectros analisados. O *Loading plot* da Figuras 3e, f, projetado a partir dos componentes principais, ilustra a contribuição das variáveis espectrais originais para a formação dos quatro primeiros componentes principais. As principais variações podem ser observadas nas regiões entre 1200 a 1000  $\text{cm}^{-1}$  onde é possível fazer a diferenciação dos espectros por inspeção direta visual.

Os componentes principais, utilizadas nesta análise (Figuras 3a, c, e, foram calculadas a partir da matriz de covariância formada por dados espectrais em quase toda sua extensão, região de 4000 a 1000  $\text{cm}^{-1}$ , com exceção da região de  $\text{CO}_2$ , entre 2600 e 2300  $\text{cm}^{-1}$ .

Para o agrupamento hierárquico foi utilizado o algoritmo Ward, considerado o mais eficiente método de agrupamento hierárquico (Everitt, 1994; Ward, 1963). O resultado, mostrado no dendrograma da Figura 4 apresenta uma clara discriminação entre os três espectros. No dendrograma são visualizadas as distâncias entre os grupos (*cluster*) de forma decrescente, e quanto menor a distância, maior a semelhança entre as observações. Essa distância é mensurada a partir da matriz das distâncias euclidianas utilizando-se a soma dos desvios quadráticos dos pontos para o centroide, calculo feito pelo algoritmo Ward. Para a criação da matriz das distâncias euclidianas

são utilizadas as componentes principais (PC's) originadas da matriz de covariância dos valores de absorvância dos espectros (medida adimensional). Logo, a distância no dendrograma representa a distância geométrica entre pontos no espaço multidimensional, sendo esta, um parâmetro adimensional (Cormack, 1971; Romesburg, 1990).

As bandas da amida I, na região de 1710 a 1590  $\text{cm}^{-1}$ , apresentam o mesmo aspecto visual nos espectros dos três micro-organismos. Aplicamos a análise estatística multivariada a essas bandas da amida I, utilizando os mesmos parâmetros da análise anterior e, o resultado, apresentado no dendrograma da Figura 5, mostra que o algoritmo utilizado conseguiu

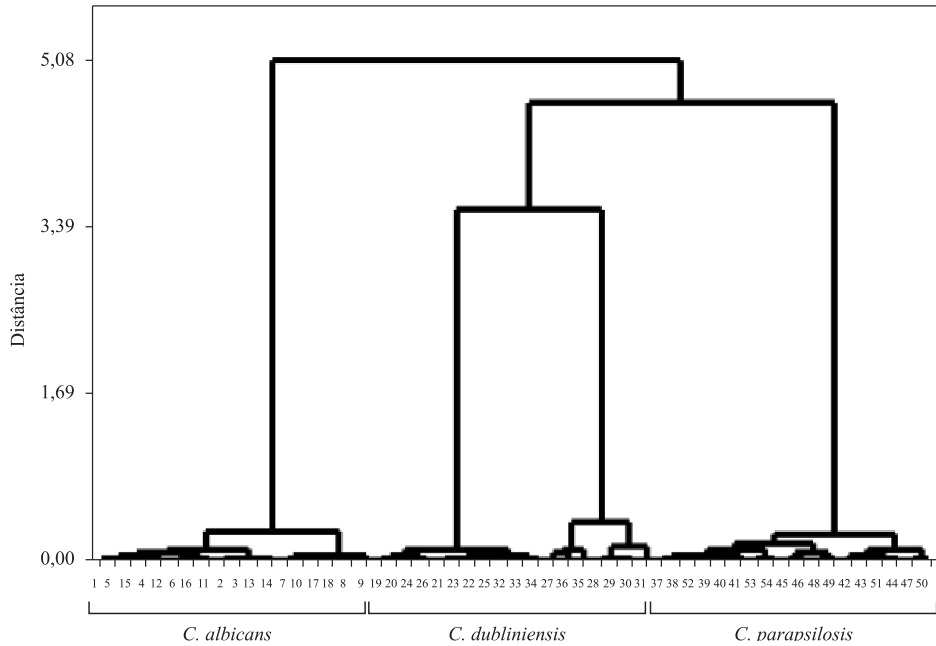


**Figura 3.** Gráficos dos componentes principais dos espectros das espécies de *Candida*: (a) e (b) Gráficos dos autovalores de cada componente principal – Scree plot; (c) e (d) Gráficos dos autovetores (scores) dos componentes principais; (e) e (f) Gráfico de loading plot dos componentes principais; (a), (c) e (e) são valores calculados a partir do intervalo espectral de 4000 a 2600 e 2300 a 1000  $\text{cm}^{-1}$ ; (b), (d) e (f) são valores calculados a partir do intervalo espectral da banda de Amida I (1710 a 1590  $\text{cm}^{-1}$ ).

**Figure 3.** Graphics of the four main components of the spectra of *Candida* species (a) and (b) Graphics of the eigenvalues of each main component - Scree plot; (c) and (d) Graphics of the eigenvectors (scores) of the principal components; (E) and (F) Graphic of "loading plot" of the principal components; (a), (c) and (e) are calculated values from the spectral range 4000-2600 and 2300-1000  $\text{cm}^{-1}$  and (b), (d) and (f) are calculated values from the spectral range of Amide I band (1710-1590  $\text{cm}^{-1}$ ).

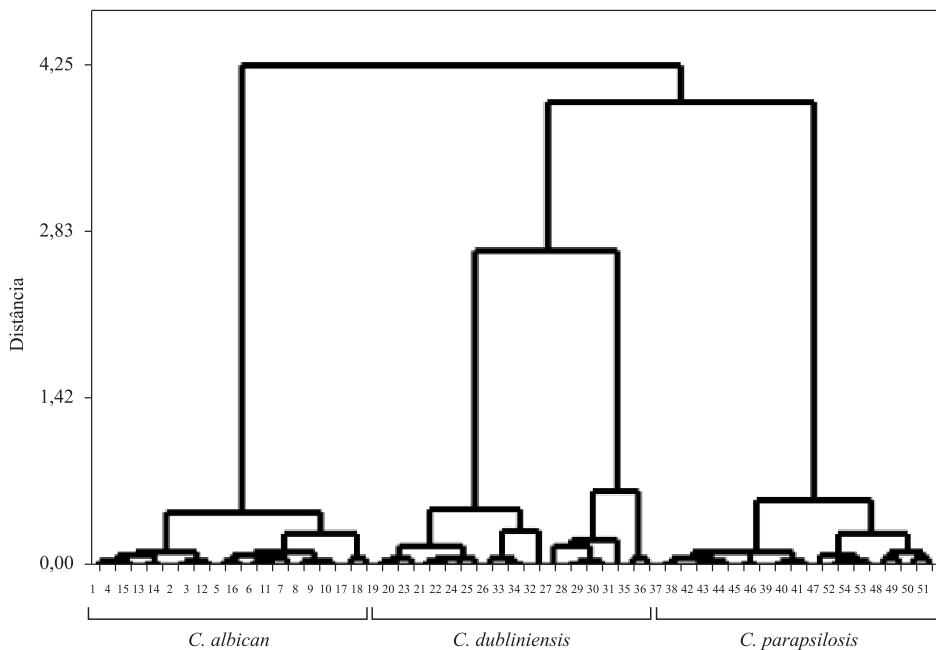
diferenciar os três espectros. Esse resultado demonstra que, mesmo nos casos em que os espectros de diferentes micro-organismos apresentem o mesmo aspecto visual, a análise estatística multivariada pode ser

potencialmente útil para diferenciar os espectros desses micro-organismos (Essendoubi *et al.*, 2005; Naumann, 2000; Timmins *et al.*, 1998; Tintelnot *et al.*, 2000; Toubas *et al.*, 2007).



**Figura 4.** Dendrograma de *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis* para região espectral de 4000-2600  $\text{cm}^{-1}$  e 2300-1000  $\text{cm}^{-1}$ .

**Figure 4.** Dendrogram of *Candida albicans*, *Candida dubliniensis* and *Candida parapsilosis* using spectra in 4000-2600  $\text{cm}^{-1}$  and 2300-1000  $\text{cm}^{-1}$  spectral zones.



**Figura 5.** Dendrograma de *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis* para região espectral de 1710-1590  $\text{cm}^{-1}$ .

**Figure 5.** Dendrogram of *Candida albicans*, *Candida dubliniensis* and *Candida parapsilosis* using spectra in 1710-1590  $\text{cm}^{-1}$  spectral zone.



Um bom teste para qualquer algoritmo estatístico é tomar um conjunto de dados com uma estrutura de agrupamento conhecida e verificar se o algoritmo é capaz de reproduzir essa estrutura. Com relação aos espectros de *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis*, a banda de amida I foi analisada através de um método não-estatístico, com base em conhecimento a respeito de estruturas secundárias de proteínas.

#### Ajuste de curva

As estruturas secundárias de proteínas se refletem nas bandas de amida I e amida II. No entanto, na de amida II, surgem muitas vezes misturadas as bandas de aminoácidos residuais. Assim, a amida I que corresponde às vibrações de estiramento de C=O do grupo peptídico é mais utilizada para análise de estruturas secundárias de proteínas (Barth, 2007; Byler e Susi, 1986; Stuart, 1997; Surewicz e Mantsch, 1988; Twardowski e Anzenbacher, 1994).

Os espectros no infravermelho entre 1590 a 1710  $\text{cm}^{-1}$  dos três micro-organismos são visualmente semelhantes apresentando um contorno com máximo

em 1651  $\text{cm}^{-1}$  acompanhado de um pequeno ombro (vide Figura 1).

Muitos autores (Byler e Susi, 1986; Grdadolnik, 2002; Stuart, 1997) indicam o método de ajuste de curva da banda de amida I como um dos melhores métodos para estimar a estrutura secundária. Os parâmetros necessários para o ajuste tais como número e posição de bandas, são obtidos pela segunda derivada dos espectros (Stuart, 1997). As posições destas bandas servem de parâmetros de entrada para ajuste de curva para a banda de amida I original.

As Figuras 6-9 apresentam as áreas dos componentes da amida I e a decomposição da banda da amida I em seus componentes lorentzianos, para os três micro-organismos.

As frações das áreas dos componentes de amida I são mostradas na Tabela 3.

Observamos que as posições de números de ondas dos componentes de amida I entre os micro-organismos são muito próximas, mas suas áreas assumem valores diferentes. Como a área se relaciona com a população, o resultado de cálculos indica quantidades diferentes de

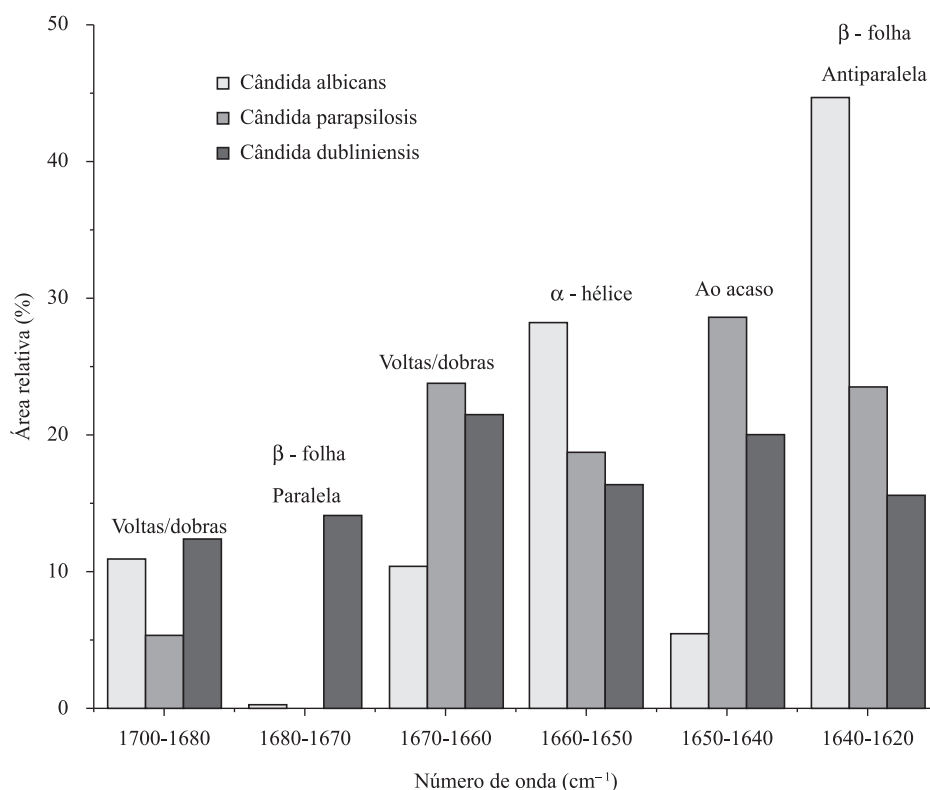
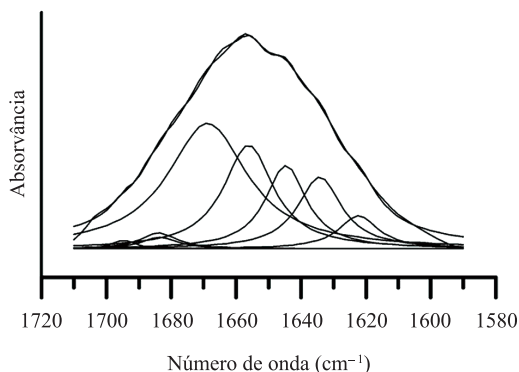
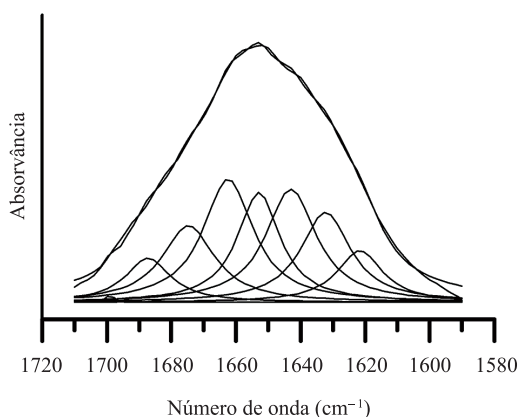


Figura 6. Distribuição das estruturas secundárias de proteínas entre *C. albicans*, *C. dubliniensis* e *C. parapsilosis*.

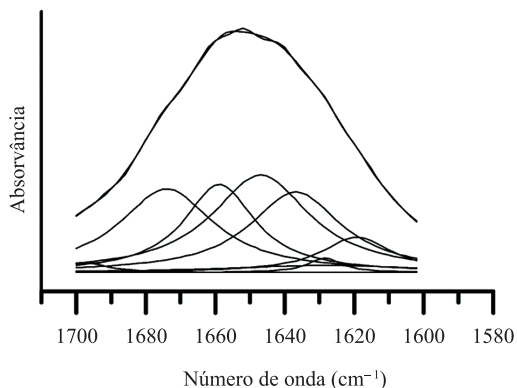
Figure 6. Distribution of the secondary structure of proteins to *C. albicans*, *C. dubliniensis* and *C. parapsilosis*.



**Figura 7.** Ajuste de curva da banda de amida I da *Candida albicans*.  
**Figure 7.** Curve fitting amida I band of the *Candida albicans*.



**Figura 8.** Ajuste de curva da banda de amida I da *Candida dubliniensis*.  
**Figure 8.** Curve fitting amida I band of the *Candida dubliniensis*.



**Figura 9.** Ajuste de curva da banda de amida I da *Candida parapsilosis*.  
**Figure 9.** Curve fitting amida I band of the *Candida parapsilosis*.

estruturas secundárias nos três micro-organismos. Por exemplo, a banda de amida I de *Candida parapsilosis* referente à conformação *parallel β-sheet* ( $\beta$  – folha paralela) não é observada e a de *Candida albicans* é muito reduzida. Na *Candida dubliniensis*, todas as conformações estão presentes e a maior área corresponde a *turns & bends* (voltas/dobras), ao passo

**Tabela 3.** Análise da banda de amida I de *C. albicans*, *C. dubliniensis* e *C. parapsilosis*.

**Table 3.** Analysis of amide I band of *C. albicans*, *C. dubliniensis* e *C. parapsilosis*.

<i>Candida albicans</i>		
Número de ondas (cm <sup>-1</sup> )	Área relativa (%)	
1695 e 1687	10,9	<i>Turns &amp; bends</i>
1679	0,3	<i>Parallel β-sheet</i>
1669	10,4	<i>Turns &amp; bends</i>
1656	28,2	$\alpha$ - <i>helix</i>
1646	5,5	<i>Random coil</i>
1635 e 1621	44,7	<i>Antiparallel β-sheet</i>
<i>Candida dubliniensis</i>		
Número de ondas (cm <sup>-1</sup> )	Área relativa (%)	
1695 e 1687	12,4	<i>Turns &amp; bends</i>
1678	14,1	<i>Parallel β-sheet</i>
1665	21,5	<i>Turns &amp; bends</i>
1656	16,4	$\alpha$ - <i>helix</i>
1645	20,0	<i>Random coil</i>
1634	15,6	<i>Antiparallel β-sheet</i>
<i>Candida parapsilosis</i>		
Número de ondas (cm <sup>-1</sup> )	Área relativa (%)	
1695 e 1686	5,4	<i>Turns &amp; bends</i>
-	-	
1669	23,8	<i>Turns &amp; bends</i>
1656	18,3	$\alpha$ - <i>helix</i>
1646	28,6	<i>Random coil</i>
1635 e 1623	23,5	<i>Antiparallel β-sheet</i>

que na *Candida albicans*, a maior área corresponde a *antiparallel β-sheet* ( $\beta$  – folha antiparalela) e para *Candida parapsilosis*, *random coil* (ao acaso).

Os resultados confirmam a discriminação dos espectros obtida através da análise estatística multivariada.

## Conclusão

Os espectros infravermelhos de *Candida albicans*, *Candida dubliniensis* e *Candida parapsilosis* foram analisados por inspeção visual direta, através da análise estatística multivariada e através do método de ajuste de curva. Os resultados e conclusões tiradas deste trabalho são:

- Os espectros apresentam, na região de 1200 a 1000 cm<sup>-1</sup>, diferenças que podem ser percebidas numa inspeção visual direta. As bandas da amida I na região de 1710 a 1590 cm<sup>-1</sup> apresentam o aspecto visual semelhante nos espectros dos três micro-organismos. É necessário, portanto, submeter a análise

estatística multivariada a um teste de sua capacidade para diferenciar espectros de micro-organismos;

- Análises estatísticas multivariadas baseadas em análise por componentes principais (PCA, *Principal Component Analysis*), com a utilização dos quatro primeiros componentes principais e análise de agrupamento (HCA, *Hierarchical Clustering Analysis*) com o uso do algoritmo Ward, foram aplicadas aos 54 espectros para investigar as regiões de 4000 a 1000  $\text{cm}^{-1}$  com exceção da região de 2600 a 2300  $\text{cm}^{-1}$  e de 1710 a 1590  $\text{cm}^{-1}$  que corresponde a das bandas da amida I. Os resultados mostram uma clara discriminação entre os três micro-organismos em ambas as análises; e
- O método do ajuste de curva que não é estatístico é aplicado a essas bandas de amida I para determinação de estruturas secundárias de proteínas. Os resultados obtidos indicam quantidades diferentes de estruturas secundárias nos três micro-organismos, confirmando a discriminação feita pela análise estatística multivariada. Esse estudo mostra o potencial desta técnica para diferenciação de micro-organismos através de seus espectros infravermelhos.

Além disso, as bandas da amida I dos espectros infravermelhos, dos três micro-organismos investigados, fornecem um conjunto de dados cuja estrutura de agrupamento é conhecida e que pode ser útil para testar algoritmos estatísticos de análise de agrupamento.

O método estatístico apresentado neste trabalho pode ser aplicado a outros algoritmos de análise de agrupamento que se pretenda testar e utilizar para a discriminação de micro-organismos através de seus espectros infravermelhos.

## Agradecimentos

Um dos autores (P. M. S.) agradece pela bolsa de estudos concedida pela CAPES.

## Referências

Adt I, Toubas D, Pinon JM, Manfait M, Sockalingum GD. FTIR spectroscopy as a potential tool to analyse structural modifications during morphogenesis of *Candida albicans*. *Archives of Microbiology*. 2006; 185:277-85. <http://dx.doi.org/10.1007/s00203-006-0094-8>.

Barbosa LCA. Espectroscopia no infravermelho na caracterização de compostos orgânicos. Viçosa: UFV; 2007.

Barth A. Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta*. 2007; 1767:1073-101. <http://dx.doi.org/10.1016/j.bbabi.2007.06.004>.

Beekes M, Lasch P, Naumann D. Analytical applications of Fourier Transform-Infrared (FT-IR) spectroscopy in microbiology and prion research. *Veterinary Microbiology*. 2007; 123:305-19. <http://dx.doi.org/10.1016/j.vetmic.2007.04.010>.

Budavari S. The Merck index: An encyclopedia of chemicals, drugs, and biologicals. 12th ed. Whitehouse Station: Merck; 1996.

Byler DM, Susi H. Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers*. 1986; 25:469-87. <http://dx.doi.org/10.1002/bip.360250307>.

Celebi S, Hacimustafaoglu M, Ozdemir O. Nosocomial candidemia in children: results of a 9-year study. *Mycoses*. 2008; 51:248-57. <http://dx.doi.org/10.1111/j.1439-0507.2007.01464.x>.

Chaffin WLJ. *Candida albicans* cell wall proteins. *Microbiology and Molecular Biology Reviews*. 2008; 72:495-544. <http://dx.doi.org/10.1128/MMBR.00032-07>.

Colthup NB, Daly LH, Wiberley SE. Introduction to Infrared and Raman Spectroscopy. 3rd ed. New York: Academic Press; 1997.

Cormack RA. Review of classification. *Journal of the Royal Statistical Society (Series A)*. 1971; 134(3):321-67.

Dziuba B, Babuchowski A, Nalecz D, Niklewicz M. Identification of lactic acid bacteria using FTIR spectroscopy and cluster analysis. *International Dairy Journal*. 2007; 17:183-9. <http://dx.doi.org/10.1016/j.idairyj.2006.02.013>.

Essendoubi M, Toubas D, Bouzaggou M, Pinon JM, Manfait M, Sockalingum GD. Rapid identification of *Candida* species by FT-IR microspectroscopy. *Biochimica et Biophysica Acta*. 2005; 1724:239-47. <http://dx.doi.org/10.1016/j.bbagen.2005.04.019>.

Everitt BS. Statistical methods for medical investigations. 2nd ed. London: Edward Arnold; 1994.

Gompertz OF, Gambale W, Paula CR, Corrêa B. Características gerais dos fungos. In: Trabulsi LR, Alterthum F, editors. *Microbiologia*. 5th ed. São Paulo: Atheneu; 2008. p. 479-82.

Grdadolnik J. A FTIR investigation of protein conformation. *Bulletin of the Chemists and Technologists of Macedonia*. 2002; 21(1):23-34.

Johnson RA, Wichern DW. Applied multivariate statistical analysis. 3rd ed. New Jersey: Prentice Hall, Englewood Cliffs; 1992.

Koneman EW, Allen SD, Janda WM, Scherecken-Berger PC, Win Junior WC. Diagnóstico microbiológico - Texto e Atlas Colorido. 5. ed. Rio de Janeiro: MDSI; 2001.

Mähns B, Sther F, Schäfer W, Neuber K. Comparison of standard phenotypic assays a PCR method to

- discriminate *Candida albicans* and *C. dubliniensis*. *Mycoses*. 2005; 48:55-61. <http://dx.doi.org/10.1111/j.1439-0507.2004.01054.x>.
- Manly FJ. *Multivariate statistical methods – A Primer*. 3rd ed. Boca Raton: Chapman & Hall/CRC; 2005.
- Mantsch HH, Chapman D. *Infrared spectroscopy of biomolecules*. New York: Wiley-Liss, Inc.; 1996.
- Maquelin K, Kirschner C, Choo-Smith LP, Braak NVD, Endtz HP, Naumann D, Puppels GJ. Identification of medically relevant microorganisms by vibrational spectroscopy. *Journal of Microbiological Methods*. 2002; 51(3):255-71. PMID:12223286.
- Naumann D. *Infrared Spectroscopy in Microbiology*. In: Meyers RA, editor. *Encyclopedia of analytical chemistry*. Chichester: John Wiley & Sons, Ltd; 2000. p. 102-31.
- Peltroche-Llacsahuanga H, Schmidt S, Seibold M, Lutticken R, Haase G. Differentiation between *Candida dubliniensis* and *Candida albicans* by fatty acid methyl ester analysis using gas-liquid chromatography. *Journal of Clinical Microbiology*. 2000; 38(10):3696-704. PMID:11015386.
- Romesburg HC. *Cluster analysis for researchers*. Malabar, Florida: Robert E. Krieger Publishing Co.; 1990.
- Sala O. *Fundamentos da espectroscopia Raman e no Infravermelho*. São Paulo: Editora UNESP; 1995.
- Sandt C, Madoulet C, Kohler A, Allouch P, Champs C, Manfait M, Sockalingum GD. FT-IR microspectroscopy for early identification of some clinically relevant pathogens. *Journal of Applied Microbiology*. 2006; 101:785-97. <http://dx.doi.org/10.1111/j.1365-2672.2006.02969.x>.
- Stuart B. *Biological applications of infrared spectroscopy*. Chichester: John Wiley & Sons, Ltd.; 1997.
- Surewicz WK, Mantsch HH. New insight into protein secondary from resolution-enhanced infrared spectra. *Biochimica et Biophysica Acta*. 1988; 952:115-30. PMID:3276352
- Timmins ÉM, Howell SA, Alsberg BK, Noble WC, Goodacre R. Rapid differentiation of closely related *Candida* species and strains by Pyrolysis-Mass Spectrometry and Fourier Transform-Infrared Spectroscopy. *Journal of Clinical Microbiology*. 1998; 36(2):367-74. PMID:9466743.
- Tintelnot K, Haase G, Seibold M, Bergmann F, Staemmler M, Franz T, Naumann D. Evaluation of phenotypic markers for selection and identification of *Candida dubliniensis*. *Journal of Clinical Microbiology*. 2000; 38(4):1599-608. PMID:10747150.
- Toubas D, Essendoubi M, Adt I, Pinon JM, Manfait M, Sockalingum GD. FTIR spectroscopy in medical mycology: applications to the differentiation and typing of *Candida*. *Analytical Bioanalytical Chemistry*. 2007; 387:1729-37. <http://dx.doi.org/10.1007/s00216-006-1005-1>.
- Toyran N, Zorlu F, Dönmez G, Öge K, Severcan F. Chronic hypoperfusion alters the content and structure of proteins and lipids of rat brain homogenates: a Fourier Transform infrared spectroscopy study. *European Biophysics Journal*. 2004; 33: 549-54. <http://dx.doi.org/10.1007/s00249-004-0396-1>.
- Twardowski J, Anzenbache P. *Raman and IR spectroscopy in biology and biochemistry*. New York: Ellis Horwood; 1994.
- Ward JH. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 1963; 58:236-44. <http://dx.doi.org/10.2307/2282967>.

---

## Autores

### Patrícia Marcondes dos Santos\*

Instituto de Pesquisa de Desenvolvimento, Universidade do Vale do Paraíba – Univap, Av. Shishima Hifumi, 2911, Urbanova, CEP 12244-000, São José dos Campos, SP, Brasil.

### Maria Angélica Gargione Cardoso

Laboratório de Imunologia e Farmacognosia, Instituto de Pesquisa de Desenvolvimento, Universidade do Vale do Paraíba – Univap, São José dos Campos, SP, Brasil.

### Sônia Khouri

Laboratório de Microbiologia, Faculdade de Ciências da Saúde, Universidade do Vale do Paraíba – Univap, São José dos Campos, SP, Brasil.

### Alderico Rodrigues de Paula Júnior

Processamento de Sinais Biológicos, Instituto de Pesquisa de Desenvolvimento, Universidade do Vale do Paraíba – Univap, São José dos Campos, SP, Brasil.

### Mituo Uehara †, Kumiko Koibuchi Sakane

Laboratório de Espectroscopia no Infravermelho, Instituto de Pesquisa e Desenvolvimento, Universidade do Vale do Paraíba – Univap, São José dos Campos, SP, Brasil.