YEAST METABOLIC STATE IDENTIFICATION BY FIBER OPTICS SPECTROSCOPY

C. C. Castro[1], J. S. Silva[1], V. V. Lopes[2], R. C. Martins[3] *
[1] *IBB - Institute for Biotechnology and BioEngineering, Universidade do Minho*
Campus de Gualtar, 4710-057 Braga, Portugal
[2] *INETI - Instituto Nacional de Engenharia Tecnologia e Inovação*
Estrada do Paço do Lumiar, 22, 1649-038 Lisboa
[3] *BioInformatics - Molecular and Environmental Biology Research Center, Universidade do Minho*
Campus de Gualtar, 4710-057 Braga, Portugal

Keywords: *Saccharomyces cerevisiae*, Morphology, LWUV-VIS-SWNIR reflectance spectroscopy, Singular value decomposition, Classification.

Abstract: In this manuscript we explore the feasibility of using LWUV-VIS-SWNIR (340 - 1100 nm) spectroscopy to classify *Saccharomyces cerevisiae* colony structures in YP agar and YPD agar, under different growth conditions, such as: i) no alcohol; ii) 1 % (v/v) Ethanol; iii) 1 % (v/v) 1-Propanol; iv) 1 % (v/v) 1- butanol; v) 1 % (v/v) Isopropanol; vi) 1 % (v/v) $(\pm)$-1-Phenylethanol; vii) 1 % (v/v) Isoamyl alcohol; viii) 1 % (v/v) tert-Amyl alcohol (2-Methyl-2-butanol); and ix) 1 % (v/v) Amyl alcohol. Results show that LWUV-VIS-SWNIR spectroscopy has the potential for yeasts metabolic state identification once the spectral signatures of colonies differs from each others, being possible to acheive 100% of classification in UV-VIS and VIS-SWNIR. The UV-VIS region present high discriminant information (350-450 nm), and different responses to UV excitation were obtained. Therefore, high precision is obtained because UV-VIS and VIS-NIR exhibit different kinds of information. In the future, high precision analytical chemistry techniques such as mass spectroscopy and molecular biology transcriptomic studies should be performed in order to understand the detailed cell metabolism and genomic phenomena that characterize the yeast colony state.

1 INTRODUCTION

Recent studies show that *S. cerevisiae* can form complex colony structures with an apparent cell specialization. Colonies of wild yeasts can contain all the varieties of cells, from which the mostly known are the diploid, haploid, hyphae form (diploid or haploid) and ascus (spore); opposing to the most well known yeast cell cycle - the budding yeast. Furthermore, it is known that *S. cerevisiae* can undergo changes in their replicative patterns and morphologies, according to environmental conditions (i.e., deleterious), to produce elongated cells joined-together in filaments (Dickinson, 2008) and colonies can signal each other (Palkova and Vachova, 2003).

The yeast-form and filamentous-form cell cycles are similar but, according to (Kron and Gow, 1995), in yeast-form growth, daughter cells are smaller than its mother and must undergo a period of further growth (in phase G1) before starting a new cell cycle (asymmetric cell division). On the other hand, filament-form cells have a symmetric cell division, once after mitosis and cell division, both mother and daughter cells are equal-sized and bud emerge starts in both cells. Furthermore mitochondrial mass and chitin deposition increases in filament form. The filament form cellular walls have a greater strength and rigidity than most of the other yeast forms, which has been suggested as a mean of penetration on solid media because of the yeast lack of natural mobility. The transcription of all of genes also decreases in filament forming yeast, and therefore has been proven difficult to find a direct transcriptomic relationship (Dickinson, 2008). Filament formation can be induced by nitrogen starvation or limitation (Rua et al., 2001), growth on a poor nitrogen source (Dickinson, 1994) or growth in the presence of low concentrations of fusel alcohols (end-products of cells catabolism). In the case of nitrogen starvation or limitation, filamentation can be explained as a foraging response because yeast is non-motile and cannot move to search

*Corresponding author: rui.martins@bio.uminho.pt

C. Castro C., S. Silva J., V. Lopes V. and C. Martins R. (2009).
YEAST METABOLIC STATE IDENTIFICATION BY FIBER OPTICS SPECTROSCOPY.
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pages 169-178
DOI: 10.5220/0001551201690178
Copyright © SciTePress

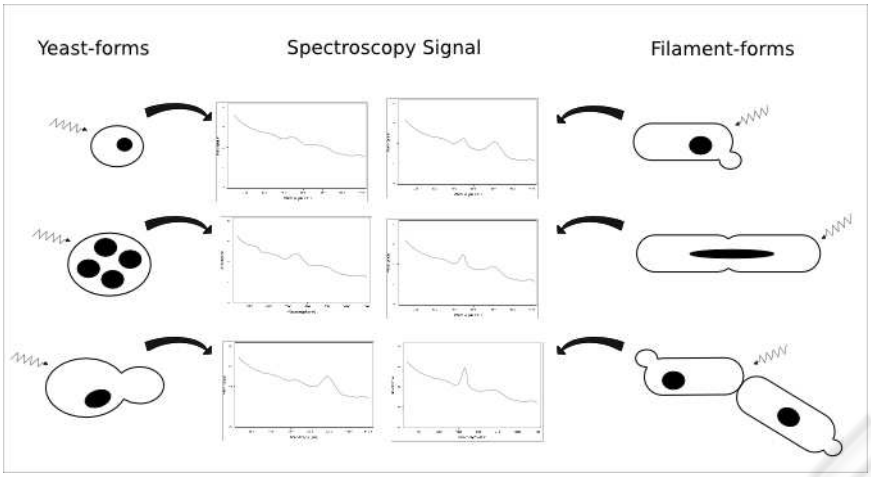

Figure 1: Spectroscopy signalling of different yeast structures.

for a richer supply of nutrients, it can only grow to explore its surroundings (Gimeno et al., 1992). Filamentation has also been induced by superior alcohols and AMPc, and has been argued that these may act as communication molecules between different yeast, allowing for the colony to synchronize its development, a phenomena known as quorum-sensing. Therefore, the yeast colony state and its dynamics is not yet explained. In order to understand this phenomena, an non-targeted, holistic and high-output approach is needed in order to gather the maximum information as possible to understand colony dynamics and yeast communication. In this sense, the use of spectroscopy, in conjunction with other techniques, allows to implement real-time and non-destructive methodologies that can explain the transcriptomics and metabolomics processes happening in yeast cellular communities.

Microorganisms traditional identification methods are supported by morphological and growth capacity in selective media (Gerard et al., 2006). The use of high-output methodologies to increase analysis capacity, such as mass spectroscopy, PCR and spectroscopy are becoming popular, not only because of the time needed for an effective identification, but more importantly because these methods are multivariate, which allows to obtain vast amounts of information in one measurement (Rosah et al., 2005).

Spectroscopy is a simple, precise, rapid, multivariate and non-destructive technique. Spectra is proportional to the chemical composition of the analite, acting as a non-destructive methodology capable of both fingerprint and quantifications. In this case, spectroscopy is may be able to classify the invariable structure of yeast, the metabolism and cell communication. Cells morphology is a visible expression of microorganisms physiology and metabolism (Treskatis et al., 1997). Different morphologies characterize different proteomic composition and different metabolism states that can be differentiated by UV-VIS-SWNIR spectroscopy.

Many spectroscopy techniques have been used for microorganisms identification, where NIR and Raman spectroscopy are the most popular (Stuart, 2004; Dziuba et al., 2007; Bhatta et al., 2005). Recently, a previous study revealed that UV-VIS-SWNIR is also a highly accurate spectroscopy method for microorganisms identification (Silva et al., 2008). UV spectroscopy records electronic transitions between electron energy levels from molecular levels in the UV-VIS region depend upon the energy involved. For any molecular bound (sharing a pair of electrons), orbitals are a mixture of two contributing orbitals σ and π, with corresponding anti-bounding orbitals σ^* and π^*, respectively. Some chemical bounds present characteristic orbital conditions, ordered by higher to lower order energy transitions: i) alkanes ($\sigma \rightarrow \sigma^*$; 150nm); ii) carbonyls ($\sigma \rightarrow \pi^*$; 170nm); iii) unsaturated compounds ($\pi \rightarrow \pi^*$; 180nm); iv) molecular bounds to O, N, S and halogens ($n \rightarrow \sigma^*$; 190nm); and v) carbonyls ($n \rightarrow \pi^*$; 300nm). As most UV-VIS spectrometers yield a minimum wavelength of 200nm, this technique has been considered to provide lower information in terms of functional groups when compared to IR, because spectral differences mostly attributed to conjugated $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions (Levine, 1975; Denney and Sinclair, 1987; Perkauparus et al., 1994).

Many organic molecules present conjugated unsaturated and carbonyls bounds, such as aminoacids, phospholipids, free fatty acids, phenols and flavonoids, peroxides, peptides and proteins,

sugars and their polymers absorb in these bands. UV-VIS not only records the effect of electron excitation, but also the effect of return to lower orbitals, which result in vibrational and rotational modes, increasing the characteristic spectra of biological materials. This effect enhances photochemical reactions and fluorescence which are important features for microbiological identification (Levine, 1975; Coyle, 1989; Klessinger and Michl, 1995) and help to identify metabolic states of yeast. Many biological molecules also present chromophore groups, which increase the absorption in the UV-VIS region, such as: nitro, nitroso, azo, azo-amino, azoxy, carbonyl and thiocarbonyl (Coyle, 1989; Klessinger and Michl, 1995). Moreover the sensitivity of today's spectrometers has highly increased, being possible to obtain low noise to signal ratios which expands the detection limits (Optics, 2006). LWUV-VIS-SWNIR has as main advantages the minimization of liquid water absorbance and effect of temperature. Furthemore, as state of the art spectrometers also include high frequency vibrational infrared (SWNIR), it is also possible to obtain important information on water, fats and proteins (Burns and Ciurczak, 2001; Devices, 2005).

The main objective of this exploratory work is to determine if UV-VIS-SWNIR is a suitable methodology that may be used to recognize the state of *S. cerevisiae* colonies by spectral signal processing to obtain discrimination among different induced metabolism, cellular communication, morphology and growth media.

2 MATERIALS AND METHODS

2.1 Sample Preparation

Saccharomyces cerevisiae wild type was obtained from the microbiological collection of the IBB - Institute for Biotechnology and BioEngineering at the University of Minho.

The incubation was performed in YPD broth medium (Sigma Aldrich - ref. Y1375) during 12 hours at 25 ^oC under constant agitation (250 rpm). Wild type yeast (20 μl) was inoculated on the surface of YP and YPD agar mediums using different growth conditions, such as: without alcohol and with 1 % (v/v) of an alcohol and was incubated at 25 C during 144 h. Studied alcohols were: Ethanol (Riedel-de Han - ref. 32221), 1-Propanol (Sigma Aldrich - ref. 538000), 1- Butanol (Sigma Aldrich - ref. BT105), Isopropanol (Sigma Aldrich - ref. 190764), (±)-1-Phenylethanol (Fluka - ref 09449), Isoamyl alcohol (SAFC - ref. W205710), tert-Amyl alcohol (2-Methyl-2-butanol) (Sigma Aldrich - ref. 152463) and Amyl alcohol (SAFC - ref. 205605) (Sigma-Aldrich Quimica, 2008).

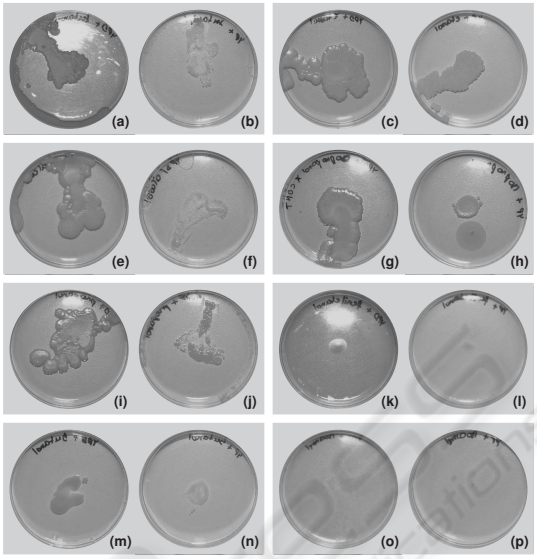

Figure 2: (a) YPD tert-amylOH; (b) YP tert-Amyl; (c) YPD Ethanol; (d) YP Ethanol; (e) YPD without alcohol; (f) YP without alcohol; (g) YPD Isopropanol; (h) YP Isopropanol; (i) YPD Propanol; (j) YP Propanol; (k) YPD Phenylethanol; (l) YP Phenylethanol; (m) YPD Butanol; (n) YP Butanol; (o) YPD Isoamyl; (p) YP Isoamyl.

Growth medium present the following constitutions: YPD broth medium (Sigma-Aldrich ref. Y1357): 10 $g.l^{-1}$ Yeast extract , 20 $g.l^{-1}$ Peptone and 20 $g.l^{-1}$ Glucose (Sigma-Aldrich Quimica, 2008); YP agar: 10 $g.l^{-1}$ Yeast extract (Fluka - ref. 70161), 20 $g.l^{-1}$ Peptone ($Bacto^{TM}$ - ref. 211677) and 15 $g.l^{-1}$ Agar (Fluka ref. 05039) and YPD agar medium (Sigma Aldrich - ref. Y1500): 10 $g.l^{-1}$ Yeast extract , 20 $g.l^{-1}$ Peptone, 15 $g.l^{-1}$ Agar and 20 $g.l^{-1}$ Glucose (Sigma-Aldrich Quimica, 2008).

Both agar medium were prepared according to the indications of the manufacturer: i) suspension of the dehydrated media in purified water (amounts defined by the manufacturer); ii) heating of the media, with frequent agitation, until complete dilution; iii) autoclave of the mixture at 121^oC for 15 minutes; and iv) shed in petri plate (Sigma-Aldrich Quimica, 2008).

2.2 Spectroscopy

Saccharomyces cerevisiae UV-VIS-SWNIR spectroscopy was performed with: i) Avantes multichannel fiber optic spectrometer AvaSpec-2048-4-DT (200 to 1100 nm; 2048 pixel) (Avantes, 2007); ii) reflection UV-VIS and VIS-SWNIR probes,

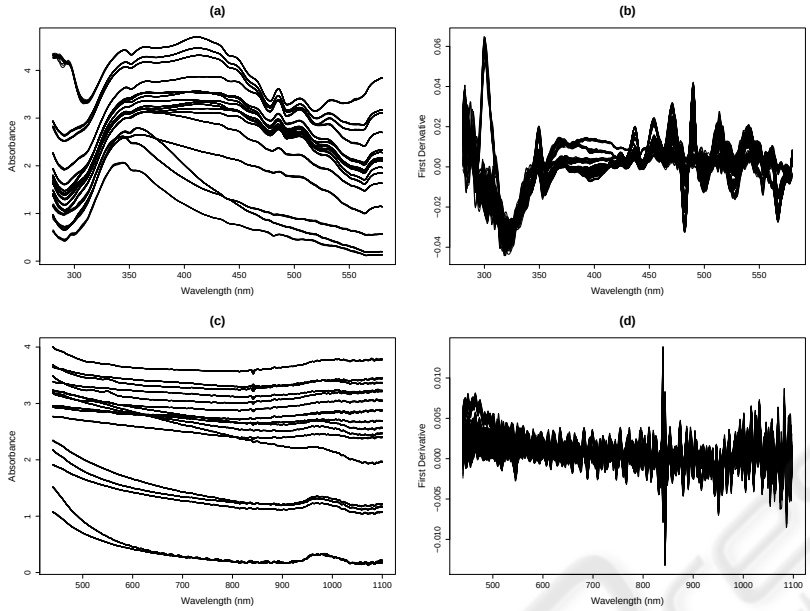

Figure 3: Microorganisms spectra: (a) Absorbance spectrum LWUV-VIS; (b) First derivative spectrum LWUV-VIS; (c) Absorbance spectrum VIS-SWNIR; (d) First derivative spectrum VIS-SWNIR.

models FCR-7UV200-2ME and FCR-7IR200-2-ME (Avantes, 2007); and iii) a balanced deuterium-tungsten halogen light source, model DH-2000-BAL (Micropack, 2008). AvaSoft 6.0 was used to control the spectrometer and data acquisition (Avantes, 2007).

Spectra were obtained at the room temperature of $18 \pm 2^{o}C$ and the light source in (a) UV-VIS: the deuterium lamp was let to stabilize during 20 min; and (b) VIS-NIR: the tungsten lamp lamp was let to stabilize during 15 min. The dark spectra was recorded and measurements were taken with linear and electric dark correction. Both light spectra were monitored by statistically assessing the reproducibility of the light source with measurements of light during the several days of the experiment. Twenty spectra replicates were recorded of UV-VIS and VIS-SWNIR measurement of both plate count agar and microorganisms colonies to study scattering effects. Furthermore, spectra were obtained inside a box designed to isolate the environmental light and maintain the probe horizontally.

2.3 Spectral Analysis

2.3.1 Robust Mean Scattering Correction

The collected spectrum were smoothed by using a Savisky-Golay filter (length = 4, Order= 2) (Savitzky and Golay, 1964) and afterwards, was pre-processed using a modified robust multiplicative scatter correction algorithm (RMSC) (Gallager et al., 2005; Martens and Stark, 1991; Martens et al., 2003): $\mathbf{x}_{corr} = \mathbf{x}\mathbf{b} + \mathbf{a}$. The $\mathbf{a}$ and $\mathbf{b}$ are computed by minimizing the following error: $\mathbf{e}_j = \mathbf{b}\mathbf{x}_j + \mathbf{a} - \mathbf{x}_{ref}$; where the $\mathbf{x}_j$ is the j sample spectrum and $\mathbf{x}_{ref}$ is the growth media spectrum.

The RMSC algorithm is based on the application of the robust least squares method to determine the $\mathbf{a}$ and $\mathbf{b}$ matrices, ensuring that spectral areas that do not correspond to scattering artifacts are not taken into consideration. The robust least squares algorithm is implemented by the re-weighted least squares with the weights computed using the Huber function. The algorithm high breakdown point (50%) means that existent outliers will not distort the model fitting (eq. **??**) and thus, the $\mathbf{a}$ and $\mathbf{b}$ scatter correction parameters are determined using only the consistent spectral areas. The iterative algorithm can be described, briefly as follow: 1) set the reference spectrum ($\mathbf{x}_{ref}$) equal to the sample spectrum closest to the median spectrum; 2) correct the remaining sample spectrum by applying the above described robust least squares procedure; and 3) recompute the median spectrum and iterate until convergence.

2.3.2 Singular Value Decomposition

Singular value decomposition (SVD) is a blind signal decomposition technique widely used in spectroscopy data, where the corrected spectrum ($\mathbf{x}_{corr}$) is decom-

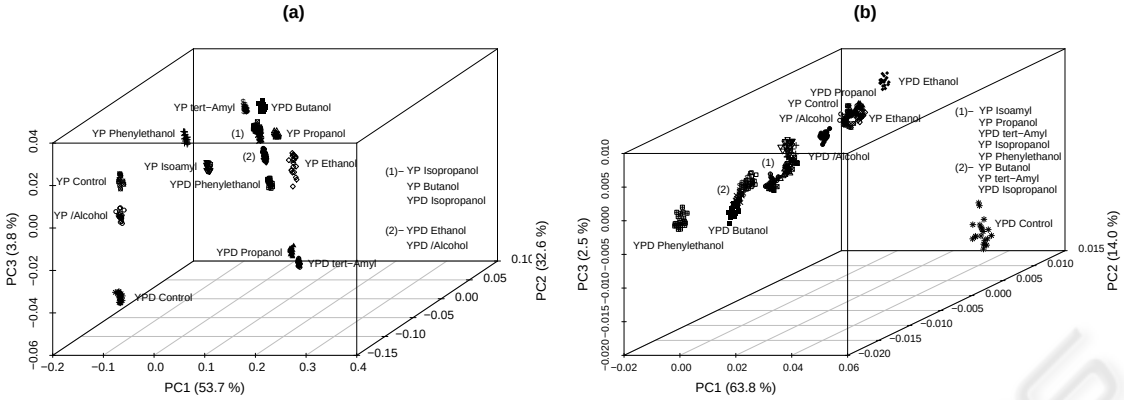

Figure 4: First derivative spectra PCA analysis: (a) LWUV-VIS Gabriel Plot (PC1 (78,40%), PC2 (8,03%); Symbols: i) YPD terta-Amyl ($\oplus$); ii) YP terta-Amyl ($); iii) YPD Ethanol ($\blacklozenge$); iv) YP Ethanol ($\lozenge$); v) YPD without alcohol ($\bullet$); vi) YP without alcohol ($\circ$); vii) YPD Isopropanol ($\times$); viii) YP Isopropanol ($\boxtimes$); ix) YPD Propanol ($\blacktriangle$); x) YP Propanol ($\triangle$); xi) YPD Phenylethanol ($\boxplus$); xii) Phenylethanol ($+$); xiii) YPD Butanol ($\blacksquare$); xiv) YP Butanol ($\square$); xv) YP Isoamyl ($\triangledown$); xvi) YPD control ($*$); xvii) YP control ($\otimes$).

posed in order of magnitude of variation directions in the variable space (wavelengths). Generally, most variability is captured in the first principal components (PC), where as, in good signal to noise spectral data, noise is captured in the last decompositions. Therefore a spectrum can be decomposed as: $\mathbf{x}_{corr} = \widehat{\mathbf{x}} + \varepsilon(\mathbf{x})$; where $\widehat{\mathbf{x}}$ is the signal and $\varepsilon(\mathbf{x})$ is the estimated noise of $\mathbf{x}$. This decomposition is possible to be performed by singular value decomposition (SVD):

$$\mathbf{X} = \mathbf{USV^T} \tag{1}$$

where $\mathbf{X}$ is a matrix that contains all the corrected spectra, $\mathbf{US}$ are the scores, $\mathbf{V^T}$ the loadings and the $\mathbf{S}$ singular values (Jolliffe, 1986; Krzanowski, 1998; Baig and Rehman, 2006).

To distinguish between the number of relevant decompositions, a randomization test is performed to the original matrix ($\mathbf{x}$) to determine the number of relevant singular values (Manly, 1998). In this research, 500 randomizations were performed by permutation the spectral scope value for the same wavelengths among the different samples, to do not violate the spectral continuity. By comparing the singular values of randomized spectrum with the original spectrum, the number of independent singular values and decompositions that discriminate the different microorganisms spectrum are obtained, so that:

$$\widehat{\mathbf{X}} = \mathbf{US}_{relv}\mathbf{V^T}_{relv} \tag{2}$$

where $\mathbf{US_{relv}}$ and $\mathbf{V^T_{relv}}$ are the statistically relevant scores and loading of $\mathbf{X}$. To further discriminate between the microorganisms spectrum, the relevant PC's scores($\mathbf{US}_{relv}$) were subjected to hierarchical clustering analysis using the eucledian distance. Further class identification was performed using soft independent class analogy (SIMCA) (Doytchinova and Flower, 2006).

Not all features in the spectrum fingerprint preserve the same quality after signal decomposition into relevant principal components. In these cases, their reconstruction is statistically impossible. In practical terms, features that are not compressed, cannot be analyzed in the score plot. Feature extraction quality can be assessed by the Q-statistic (square prediction error) of the relevant decomposition developed by (Jackson and Mudholkar, 1979):

$$Q_i = \mathbf{e}_i\mathbf{e}_i^T \tag{3}$$

where $\mathbf{e}_i = \mathbf{x}_i - \widehat{\mathbf{x}}_i$ is the reconstruction error for the i^{th} spectrum present in the data matrix (the i^{th} row of $\mathbf{X}$). An accepted way of computing the Q statistical confidence interval (Q_α) is defined as:

$$Q_\alpha = \theta_1 \cdot \left[1 + \frac{Z_{\alpha/2}}{\theta_1}\sqrt{2\theta_2 h_0^2} + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2}\right]^{\frac{1}{h_0}} \tag{4}$$

where Z_α is the inverse normal distribution value for the significance level ($\alpha/2$), $\theta_j = \sum(S_i)^j$ and $h_0 = 1 - \frac{2}{3}\theta_1\theta_3/\theta_2^2$ (Choi et al., 2005). The samples with a Q-statistic above Q_α are outside the PCA model reconstruction and contain non-common features(Conlin et al., 2000).

In these cases, the contribution plot is estimated to determine which variables are affecting the Q-statistics (Miller et al., 2003; Dunia et al., 1996), and

diagnostic why features are not captured. The reconstructed sample $\widehat{\mathbf{x}}_i$, the variable contribution for the reconstruction error is estimated by the square error of each variable $\mathbf{E}_{ij}^2$ (Miller et al., 2003; Dunia et al., 1996).

Another well known statistic is the Hotelling T^2. In SVD, this is used as a measure of the distance to the center of data, being computed by:

$$T_i^2 = \mathbf{x}_i^T \mathbf{V} \mathbf{A}^{-1} \mathbf{V}_i^T \mathbf{x} \quad (5)$$

and $A = \frac{1}{n-1}\mathbf{T}\mathbf{T}^T$, where $\mathbf{T} = (\mathbf{US})_{rel}$ and n is the number of wavelength in the spectra ($\mathbf{X}$ columns). The upper confidence interval for the Hoteling T^2 is estimated by: $T_\alpha^2 = \frac{l(n-1)}{n-1} F_{l,n-1;\alpha}$; where l is the number of relevant singular values and $F_{l,n-1;\alpha}$ the F distribution value with l and $n-1$ degrees of freedom at $\alpha = 0.05$ level of significance. Samples with a T^2-statistic above T_α^2 are considered to have significantly different features (Qin, 2003).

3 RESULTS AND DISCUSSION

3.1 Spectral Absorbance

Figure 3 presents LWUV-VIS and VIS-SWNIR yeast spectrum for different growth medium conditions. It is possible to assess in the absorbance spectrum (Figure 3 (a) and (c)) that colonies are directly distinguishable by the intensity and spectral shape. The first derivate spectrum (Figure 3 (b) and (d)) was calculated to eliminate background and baseline effects. In this signal colonies can be distinguishable in the wavelength interval of 350-500 nm and 600-900 nm in the LWUV-VIS and VIS-SWNIR, respectively.

Spectrum may contain information of the growth media. This was minimized by maximizing the contrast between the growth media and colonies. It is possible to observe that spectrum signatures for each growth medium conditions are different and well distinguished, and therefore it is reasonable to assume that most of the information obtained in the spectra is coherent with the colony metabolomic state.

3.2 Singular Value Decomposition Analysis

Figure 4 (a) present relevant scores plot in the 3 PC's for LWUV-VIS first derivative of absorbance, totalizing 90.1 % of spectral variance with discriminant power (PC1 (53.7 %), PC2 (32.6 %), PC3 (3.8 %)).

PC1 (53.7 %) separates the samples by spectral intensity into four groups: i) growth media control (YPD and YP) and colony growth in YP without alcohol; ii) colonies growth in: YP Phenylethanol and YP Isomyl; iii) colonies growth in: YPD tert-Amyl, YPD Butanol, YP Isopropanol, YP Ethanol, YP Butanol, YPD Isopropanol, YP Propanol, YPD Ethanol, YPD w/o alcohol and YPD Phenylethanol; and iv) colonies growth in YPD Propanol and YPD tert-Amyl.

PC2 (32.6 %) distinguishes the growth medium control (without colonies) from the other samples, with the exception of the colony growth in YP media without alcohol. This similarity may be due to the small sized colony, which causes the passage of light through the media.

PC3 (3.8 %) captures a small variance in the spectrum. Neverthenless, it is also significant for discrimination of the colonies of YPD Propanol and YPD tert-Amyl.

Figure 4 (b) shows the first derivative spectra PCA analysis for the VIS-SWNIR light. It is also decomposed into 3 PC'S, (80.3 % of total variance) in the VIS-SWNIR region. PC1 (63.8.0%) also segregates colonies by spectral intensity, and YPD control is completely distinguished from other samples . PC2 (14.0 %) segregates colonies into different groups, where YPD Phenylethanol, YPD w/o alcohol and YPD Ethanol, are completely distinguish. PC3 (2.5%) segregates differences between colonies in the same groups.

Figure 5 (a) and (b) presents the diagnostic plot (Q-T_h^2 plot) for the two light sources. In these two figures, YPD control and YP control samples are above the Q or T_h^2 limits, which means that the growth medias are dissociated from the rest of the samples, being possible to affirm that their spectral features are significantly different from the colonies, and statistically guaranteeing that the information contained in the collected spectra is mostly independent of the growth media, measuring the metabolomic state of each colony.

The analysis of the diagnostic plots also allows to conclude that colonies in: YP w/o alcohol (Figure 5 (a)) and YPD phenylethanol (Figure 5 (b)) are above the Q limit, which means that the features captured in the 3 first components for these medias, are not sufficient to reconstruct the original spectral data, due to large reconstruction errors (Qin, 2003). This allows us to conclude that these colonies are in completely different metabolic state than the rest of the studied growth conditions.

In Figure 5 (b), some spectra replicates of colonies in YPD Phenylethanol media are above the T_h^2 limit, which means that these spectra colony is statisti-

Table 1: Integration time, morphology and classification probabilities results.

				Classification probabilities			
	Integration Time (ms)			ABS + HCA (%)		Derv + HCA (%)	
Growth Media	UV-VIS	VIS-NIR	Morphology	UV-VIS	VIS-NIR	UV-VIS	VIS-NIR
YPD without alcohol	524	439	w/o hyphae	100	100	100	100
YP without alcohol	80	94	w/o hyphae	100	100	100	100
YPD ethanol	639	497	w/o hyphae	100	100	100	100
YP ethanol	1906	850	w/o hyphae	100	100	100	100
YPD butanol	455	373	w/o hyphae	100	100	100	100
YP butanol	529	351	hyphae	100	100	100	100
YPD propanol	1546	625	w/o hyphae	100	100	100	100
YP propanol	856	740	w/o hyphae	100	100	100	100
YPD isopropanol	490	378	hyphae	100	100	100	100
YP isopropanol	632	288	w/o hyphae	100	100	100	100
YP isoamyl	328	92	hyphae	100	100	100	100
YPD phenylethanol	620	240	w/o hyphae	100	100	100	100
YP phenylethanol	177	81	w/o hyphae	100	100	100	100
YPD terta-Amyl	1365	571	hyphae	100	100	100	100
YP terta-Amyl	400	379	hyphae	100	100	100	100

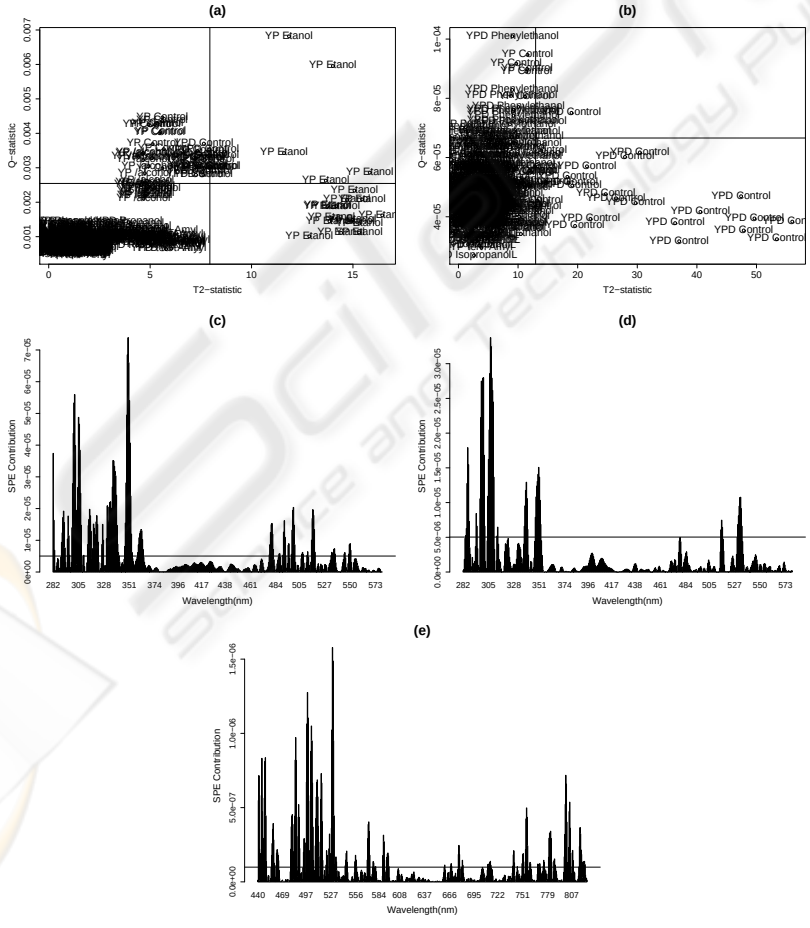

Figure 5: Diagnostic and Contribution plots: (a) LWUV-VIS Diagnostic plot; (b) VIS-SWNIR Diagnostic plot; (c) YP without alcohol LWUV-VIS (d) YP ethanol LWUV-VIS and (e) YPD phenylethanol VIS-SWNIR.

cally different from the average spectral features compressed by the 3 PC's. Therefore, this colony can be directly classified from the global SVD model (see Figure 4 (b)).

The diagnostic plots allowed to understand that the 3 PC's model is capable of discriminating the major spectral differences between colonies, but it cannot compress all the spectral features in the relevant PC'S. Such may leads to errors in distinction and spectral features interpretation.

Contribution plots allow to interpret why colonies of YPD w/o alcohol, YP Ethanol and YPD Phenylethanol (Figure 5 (c), (d) and (e), respectively), are distinguishable from the rest of the samples.

YPD w/o has higher reconstruction errors in the region of 300-350 nm, which are linked to chromophoric groups of C=C and -N=N-, respectively. YP ethanol colonies has high contribution in the intervals of 280-300 nm, which is dominated by C=C chromophoric group. The colonies which grown in the YPD phenylethanol has high contribution errors in the intervals of 440-520 nm and 750-800 nm, which are linked to C=S chromophoric group and OH overtones, respectively.

After SIMCA analysis, HCA was performed taking into consideration the euclidean distance between the center of the scores of each yeast spectra. HCA is presented in Figure 6 (a), (b) for LWUV-VIS and (c), (d) for VIS-SWNIR wavelengths, respectively.

Hierarchical clustering analysis differs in both light sources, and then the relative positions of colonies are different. In both hierarchical trees, yeasts structures that were grown under different conditions are well discriminated from each one. It is possible to observe a good discrimination between control mediums YP and YPD in LWUV-VIS and VIS-SWNIR trees, but YP ethanol, YP without alcohol and YP isoamyl are more similar to the control mediums. This phenomena can be explained because of the low density of cells and the translucency of the colony that allows light to cross the colony and then incorporate significant amount of growth media spectral information, leading to lower contrast between growth media and microorganisms in this cases. Such is especially problematic, if the colony is small sized or when the probe is not properly placed.

Comparing LWUV-VIS and VIS-SWNIR trees with Table 1, we can relate colonies aggroupment with its morphologies. In LWUV-VIS tree it is possible to distinguish some colonies with the same morphology, such as YPD phenyl-ethanol, YP 1-propanol, YP ethanol and YPD without alcohol (which do not form filament forms) from the other groups. *S. cerevisiae* that has grown in YP tert-Amyl

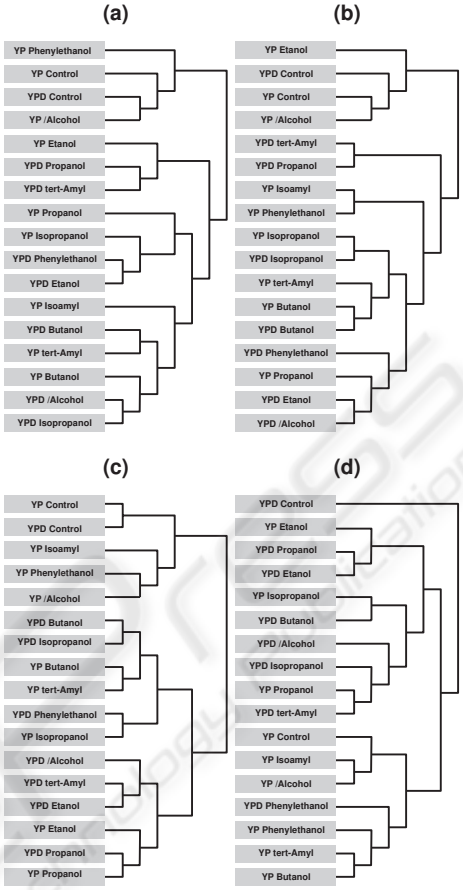

Figure 6: Hierarchical cluster Analysis: (a) Absorbance - LWUV-VIS; (b) First derivative - LWUV-VIS; (c) Absorbance - VIS-SWNIR and (d) First derivative VIS-SWNIR.

medium is well separated from the other. groups in the LWUV-VIS, but is integrated in YP-butanol group in the VIS-SWNIR wavelengths. Some groups with yeast-form can be distinguish in VIS-SWNIR wavelengths such as YPD ethanol, YPD 1-propanol and YP ethanol and YP isopropanol and YPD butanol. Furthermore, YPD without alcohol and YPD isopropanol are completely separated from all groups.

Table 1 presents a 100 % classification probabilities for all colonies. Such means that all colonies are completely differentiated from from each others. However, this classification does not depends only colony morphology, but it mostly depends on the chemical composition and metabolism of colonies. Comparing LWUV-VIS and VIS-SWNIR information (Figure Spectra), it is possible to conclude that there are considerable differences that indiciate the presence of different chemical compounds that respond differently to the UV excitation. It would be expectable that absorvance would show the same pattern in both light sources, but results show that the

yeast spectra has completely different features when responding to UV excitation, especially in the region of 350-450 nm. Such high resolution and the differences between the two light sources spectra, allows us to conclude that UV-VIS-SWNIR is capable of high performance discrimination of the yeast metabolic states.

However, as spectroscopy is a non-target approach, it is not possible identify directly the chemical compounds and the transcripted genes that differentiate the colonies. In order to understand deeper the potential of UV-VIS-SWINR spectroscopy, it is necessary to correlate spectroscopy data against high-precision analytical techniques, such as mass spectroscopy (e.g. LC-MS/MS, GC-MS or Maldi-TOF), NMR and transcriptomics (e.g. DNA/RNA Microarrays).

4 CONCLUSIONS

This work has shown that after appropriate preprocessing and signal classification, UV-VIS-SWNIR spectroscopy is a high resolution technique capable of attaining interesting possibilities in non-destructive metabolomics in the near future. Further insights will be gained when spectral information is deeper understood, not only by correlating with other high resolution analytical chemistry and molecular biology techniques, but also in understanding of the spectra shape of the differents microorganisms.

ACKNOWLEDGEMENTS

Part of this work was supported by the project OpenMicroBio (PTDC/BIO/69310/2006) - 'A Framework for Computational Simulation of Cellular Communities during BioProcess Engineering'; and partially supported by CBMA, IBB/CEB and ISR/IST plurianual funds through the POS-Conhecimento Program that includes FEDER funds.

REFERENCES

Avantes, I. (2007). Users manual.

Baig, S. and Rehman, F. (2006). Signal modeling using singular value decomposition. In *Advances in Computer, Information, and Systems Sciences, and Engineering*. Springer Netherlands.

Bhatta, H., Goldys, E., and Learmonth, R. (2005). Rapid identification of microorganisms by intrinsicfluorescence. In *Imaging, Manipulation, and Analysis of Biomolecules and Cells: Fundamentals and Applications III,*. SPIE.

Burns, D. and Ciurczak, E. (2001). *Handbook of near infrared analysisI*. Marcel Dekker, Inc, New York, 2nd edition edition.

Choi, S., Lee, C., Lee, J., Park, J., and Lee, I. (2005). Fault detection and identification of non-linear processes based on kernel pca. *Chemometrics and intelligent laboratory systems*, 75:55–67.

Conlin, A., Martin, E., and Morris, A. (2000). Confidence limits for contribution plots. *Journal of Chemometrics*, 14:725–736.

Coyle, J. (1989). *Introduction to Organic PhotoChemistry*. John Wiley & Sons, London.

Denney, R. and Sinclair, R. (1987). *Visible and ultraviolet spectroscopy*. John Wiley & Sons, London.

Devices, A. S. (2005). Near-ir absorption bands.

Dickinson, J. (1994). Irreversible formation of pseudohyphae by haploid *Saccharomyces cerevisiae*. *FEMS Microbiol. Lett.*, 119:99–104.

Dickinson, J. (2008). Filament formation in *Saccharomyces cerevisiae* - a review. *Folia Microbiol.*, 53(1):3–14.

Doytchinova, F. and Flower, D. (2006). Modeling the peptide t-cell receptor interaction by the comparative molecular similarity indices analysis-soft independent modeling of class analogy technique. *Journal of Medicinal Chemistry*, 49(7):2193–2199.

Dunia, R., Qin, S., Edgar, T., and T.J., M. (1996). dentification of faulty sensors using principal component analysis. *American Institute of Chemical Engineers*, 42:2797–2812.

Dziuba, B., Babuchowski, A., Naleczb, D., and Niklewicz, M. (2007). Identification of lactic acid bacteria using ftir spectroscopy and cluster analysis. In *International Dairy Journal 17: 183189*. Elsevier.

Gallager, N. B., Blake, T., and Gassman, P. (2005). Application of extended inverse scattering correction to mid-infrared reflectance of soil. *Journal of Chemometrics*, 19:271–281.

Gerard, J., Berdell, R., and Christine, L. (2006). *Microbiology: An Introduction*. Benjamin Cummingsc, New York, 2th edition edition.

Gimeno, C. J., Ljungdahl, P., Styles, C., and Fink, G. (1992). Unipolar cell divisions in yeast *Saccharomyces cerevisiae* lead to filamentous growth: regulation by starvation and ras. *Cell*, 68:1077–1090.

Jackson, J. and Mudholkar, G. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, 21:341–349.

Jolliffe, I. (1986). *Principal Component Analysis*. Springer, New York, USA.

Klessinger, M. and Michl, J. (1995). *Exited states and photochemistry of organic molecules*. VCH Publishers, New York.

Kron, S. and Gow, N. (1995). Budding yeast morphogenesis: gignaling, cytoskeleton and cell cycle. *Curr.Opin.Cell Biol.*, 7:845–855.

Krzanowski, W. J. (1998). *Principles of Multivariate Data Analysis*. Oxford University Press, Oxford, UK.

Levine, I. (1975). *Molecular Spectroscopy*. John Wiley & Sons, New York.

Manly, B. F. (1998). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman and Hall, London, UK, 2nd edition.

Martens, H., Nielsen, J. P., and Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures. In *Analytical Chemistry 75(9): 394-404*. American Chemical Society.

Martens, H. and Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. In *Journal of Pharmaceutical and Biomedical Analysis 9: 625-635*. American Chemical Society.

Micropack (2008). *DH2000 BAL: Installation and operating manual*. Ocean Optics, Ostfilden, Germany.

Miller, P., Swanson, R., and Heckler, C. (2003). Contribution plots: the missing link in multivariate quality control. *International Journal of Production Economics*, 9:775–792.

Optics, O. (2006). *HR4000 - High resolution fiber optic spectrometers: instalation and operation manual*. Ocean Optics, Dundelin, FL USA.

Palkova, Z. and Vachova, L. Ammonia signaling in yeast colony formation.

Perkauparus, H., Grinter, H., and Therfall, T. (1994). *Uv-Vis spectroscopy and its applications*. Springer-Verlag, New York.

Qin, S. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, 17:480–502.

Rosah, P., Hard, M., Schmitt, M., Peschke, K. D., Ronneberger, O., Burkhart, H., Mutzkus, H. W., Laukers, M., Hofer, S., Thiele, H., and Popp, J. (2005). Chemotaxonomic identification of simple bacteria by micro-raman spectroscopy: application to clean-room-relevant biological contamination. *A&EnvMicro*, 71(3):1626–1637.

Rua, D., Tobe, B., and Kron, S. (2001). Cell cycle control of yeast filamentous growth. *Curr.Opin.Microbiol.*, 4:720–727.

Savitzky, A. and Golay, M. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639.

Sigma-Aldrich Quimica, S. (2008). *Sigma, Life Science: Produtos para Investigao em Cilncias da Vida, 2008-2009*. Sigma-Aldrich, Portugal, 1nd edition.

Silva, J., Martins, R. C., Vicente, A., and Teixeira, J. (2008). Feasibility of yeast and bacteria identification using lwuv-vis-swnir diffusive reflectance spectroscopy. volume 1.

Stuart, B. (2004). *Infrared Spectroscopy: Fundamentals and Applications*. John Wiley & Sons, Ltd, London, 1nd edition.

Treskatis, S., Orgeldinger, V., Wolf, H., and Gilles, E. D. (1997). Morphological characterization of filamentous microorganisms in submerged cultures by on-line digital image analysis and pattern recognition. volume 53(2), pages 191–201.