SIZING AND COUNTING OF TWO *Saccharomyces cerevisiae* FLOC POPULATIONS IN AN AIRLIFT BIOREACTOR BY IMAGE ANALYSIS, USING AN AUTOMATICALLY CHOSEN THRESHOLD

António Vicente, Jan Marten Meinders, José António Teixeira*
Departamento de Engenharia Biológica, Universidade do Minho, Campus de Gualtar, 4709 BRAGA codex, PORTUGAL;
Tel.: +351.53.604400; Fax.: +351.53.604413
*To whom all correspondence should be sent

Summary: Two yeast (*Saccharomyces cerevisiae*) floc populations of different ages were analysed. The results showed a variability of 9 - 12 % for the single floc mean area, 6 - 7 % for the number of single flocs and 5 - 6 % for the total number of flocs. Also aggregates of two flocs (doublets) and three flocs (triplets) were enumerated. The validity of the method was checked by analysing the parameters of interest as function of the threshold. The constant correlation between the parameters and the threshold showed the consistency of the method.

Key words: image analysis, *Saccharomyces cerevisiae* floc, size distribution, counting, automatic threshold, airlift

1. INTRODUCTION

Biomass characterisation is of importance in pharmaceutical, biological and food industries[3]. On the other hand, the performance of fermentative processes using flocculent strains is highly dependent on the size distribution of flocs. Therefore the understanding of mass transfer phenomena and their measurement in yeast flocs, which is very important for the optimisation of processes involving cell aggregates (e.g. brewing and ethanol production[4] and limiting diffusion in biofilms[1]), is also dependent on the determination of the floc size distribution.

In the present work a standardised image analysis method has been developed permitting determination of the number of yeast flocs and their size distribution. The method includes image grabbing, image enhancement, automatic determination of the appropriate threshold, curve fitting of the area histogram, determination of the mean single floc area and its standard deviation and floc counting. The extension of the method to other applications is immediate and straightforward.

2. DESCRIPTION OF THE METHOD

2.1. Yeast flocs

Yeast flocs were obtained from a highly flocculent strain of *Saccharomyces cerevisiae* (NRRL Y265) cultivated in an airlift bioreactor with a working volume of 5 L. Details on the reactor itself and the fermentation conditions are described elsewhere[4,5].

Two samples were taken from the reactor and placed in Petri dishes: sample 1 after 48 hours and sample 2 after 72 hours of inoculation. In order to obtain a high number of single flocs the flocs were separated from each other as far as possible. Then, the samples were placed over the plate of an Olympus SZ 4045 TR Zoom Stereo Microscope, equipped with a TV adapter and an Olympus SZ - 12A illuminator base. An Olympus LSGA EPI illuminator was used as light source.

2.2. Image analysis procedure

Images were grabbed using a Sony AVC-D5CE CCD video camera adapted to the microscope and connected to a frame grabber (DT2851 - Data Translation, Inc.) installed in a 486 DX4 100 MHz personal computer. The analogue images grabbed by the camera were digitised by the frame grabber into a 512 x 512 pixels array, each pixel representing a grey value ranging from 0 (black) to 255 (white). From each Petri dish 10 pairs of images were taken and stored on hard disk. Each image was taken twice, so they could be used for enhancement. Finally an out-of-focus image, without flocs, was made.

Since the computer can only analyse binary images an appropriate threshold is needed. The threshold is the grey value that distinguishes between background and objects. To obtain the appropriate threshold two methods were used. The number of single flocs and multiples was determined using thresholds ranging from 1 to 255. The range where the number of objects is constant is considered the correct threshold range. On the other hand, in order to obtain a standard method of choosing the threshold, Otsu's[2] method of automatically calculating the optimum threshold was adopted allowing a comparison between both procedures. The method consists of maximising the separability of classes of grey values. This involves the use of

the zeroth- and first-order cumulative moments of the grey value histogram. The optimum can be found by maximising

$$\omega_0 \omega_1 (\eta_1 - \eta_0)^2 \quad (1)$$

where

$$\omega_0 = \sum_{i=1}^{h} p_i \quad \text{and} \quad \omega_1 = \sum_{i=h+1}^{255} p_i \quad (2)$$

and

$$\eta_0 = \frac{\sum_{i=1}^{h} i.p_i}{\omega_0} \quad \text{and} \quad \eta_1 = \frac{\sum_{i=h+1}^{255} i.p_i}{\omega_1} \quad (3)$$

in which ω_0 and ω_1 are the zeroth- and η_0 and η_1 the first-order cumulative moments of the grey value histogram, respectively. In equations (1), (2) and (3) i is the grey value, h is the threshold and p_i is the probability of grey value i defined by

$$p_i = \frac{n_i}{N} \quad (4)$$

where n_i is the number of pixels with grey value i and N the total number of pixels. The method assumes the existence of at least two peaks in the histogram, one corresponding to the objects and the other to the background. Two peaks will be well separated as long as the contrast between the objects and the background is good.

2.3. Data analysis

The number of single flocs and multiples was obtained by analysing the number of objects as function of the area. Not all flocs have the same area but are distributed around a mean value. It is reasonable to assume that this distribution follows a Gauss function

$$n = k.e^{-\left(\frac{x-\bar{a}}{\sigma}\right)^2} \quad (5)$$

in which n is the number of objects with area x, k a normalisation factor, $\bar{a}$ the mean area and σ its standard deviation. By least square fitting Equation (5) to the data points, the mean area and the standard deviation can be obtained. The total number of single flocs can be calculated by integrating Equation (5), according to

$$\int_{-\infty}^{+\infty} k.e^{-\left(\frac{x-\bar{a}_s}{\sigma_s}\right)^2} dx = \pi^{0.5}.k.\bar{a}_s \quad (6)$$

where $\bar{a}_s$ is the mean single floc area and σ_s its standard deviation. If multiples are present, a summation of Gauss functions can be used as fitting function in order to separate doublets, triplets, and higher order multiples.

Once the mean area of single flocs is known, the total number of flocs (n_{tot}) can be obtained by separating aggregates into single flocs according to

$$n_{tot} = \sum_{j=1}^{\infty} \frac{a_j}{\bar{a}_s}.n_j \quad (7)$$

in which a_j is the area of the floc j and n_j is the number of flocs with area a_j.

If the total number of doublets (n_d), triplets (n_t) and higher order multiples (n_m) is known, the total number of flocs can also be obtained from

$$n_{tot} = n_s + 2.n_d + 3.n_t + m.n_m \quad (8)$$

in which m denotes the order of the multiple m.

Finally, a calibration of the horizontal and vertical axis of the image was made replacing the objects by a bar of known length allowing the conversion of the area units from pixels to metric units. In the horizontal direction 1 pixel corresponds to 52 µm and in the vertical direction 1 pixel corresponds to 35 µm, hence each pixel has an area of 1820 μm^2.

3. RESULTS AND DISCUSSION

Floc size depends on the age of the culture. Usually, they grow in size during the first 72 hours after inoculation. Thereafter, a dynamic equilibrium is achieved. The bigger flocs end up by disrupting themselves into smaller flocs and free biomass due to shear forces inside the bioreactor. The free biomass is dragged out of the reactor while the small flocs grow again, closing the cycle.

Figure 1 shows the grey value histogram of flocs with 48 hours of age.

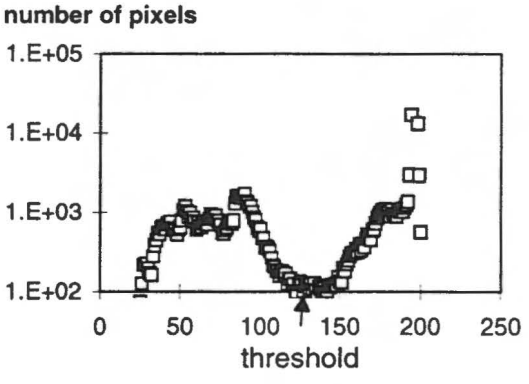

Figure 1. Grey value histogram of an image containing *S. cerevisiae* flocs. To separate flocs (low grey value, first peak) from background (high grey value, second peak) an appropriate threshold has to be chosen.

The dark flocs are visible as the first peak (low grey values) and the background is visible as the large second peak (high grey values).
To separate objects from the background an appropriate threshold has to be chosen between the two peaks. The threshold can be automatically calculated by separating the grey level histogram in classes (Equations (1) to (4)), which gives the possibility to process quickly a large number of images. This calculated threshold is indicated by the arrow.

Contrast of the images is extremely important in order to separate the objects from the background. If the grey values of the objects and background are close together, separation is difficult. Moreover, in such a situation the area of the objects will depend on the threshold. On the contrary, if the grey value peaks are far from each other then there will be a threshold range where the area of the objects will be constant. Hence, the importance of having enhancement procedures to increase the contrast. The combination of background subtraction and multiplication of two, shortly after each other grabbed images is a good one, as can be seen by the well separated peaks in Figure 1.

Figure 2 shows area histograms obtained by automatically thresholding sample 1. For sample 1 the optimum threshold was 118 and for sample 2 it was 108. For presentation reasons, a class size of 0.36 mm^2 was used.

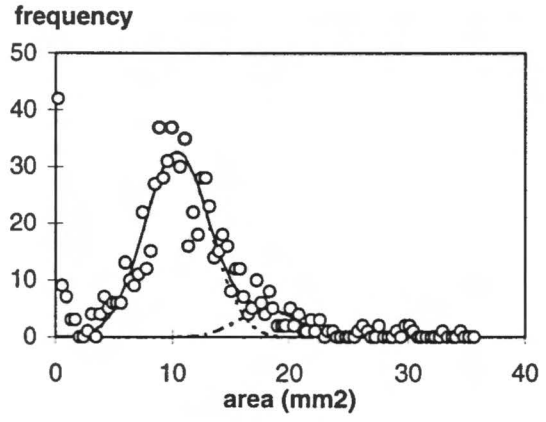

Figure 2. Area histogram plotted using the automatically obtained threshold after 48 hours (sample 1) of inoculation. (□, ○ experimental points; --- singles' fit; - • - doubles' fit; —— sum).

From Figure 2 it can be seen that an area interval around the value of 1.6 mm^2 has practically no objects. The objects below that value correspond either to noise or to parts of broken flocs. Therefore, this group of objects was not considered when fitting the Gauss functions. Two peaks are visible, the first one corresponding to singles and the second to doublets. Also a long tail of triplets and higher order multiples is present. The two peaks could be fitted by a summation of two Gauss functions to the data points, not considering the tail of the histogram. This procedure allowed to obtain the mean areas and their standard deviation (Equation (5)) and it has been repeated for all the thresholds. Since the standard deviation of the data points was not known, a goodness of the fit could not be obtained. However, the estimated standard deviation of the data points was of ± 0.3% and ± 0.4% for samples 1 and 2, respectively, indicating a good agreement between data points and the Gauss fit.

In order to check the consistency of the results the area distribution has been determined as function of the threshold. Stable results should show up as a constant line as function of the threshold, *i.e.* results should be relatively independent of the chosen threshold. That area ranges from 7.0 to 9.0 mm^2 in sample 1 and from 10 to 12 mm^2 in sample 2, hence having a variability of ± 12 % and ± 9 % with respect to the middle values 8.0 and 11 mm^2, respectively. On the other hand, using the automatically calculated threshold, mean single floc areas of 7.42 and 10.5 mm^2 are obtained, corresponding well to the average values (Table I).

Table I. Characteristics of *S. cerevisiae* flocs after 48 hours (sample 1) and 72 hours (sample 2) of inoculation. The data correspond to the area histograms obtained with the automatically calculated threshold values of 118 (sample 1) and 108 (sample 2).

	sample 1	sample 2
mean single floc area (mm^2)	7.42	10.5
standard deviation (%)	21	24
mean doublet area (mm^2)	14.0	17.0
standard deviation (%)	21	14

The numbers of single flocs and doublets were obtained after integrating the Gauss function describing the area distribution (Equation (6)). Due to low numbers of triplets and higher order multiples it was not possible to fit more Gauss functions. The number of singles (Figure 3) decreases with the threshold from 850 to 750 and from 750 to 650 for samples 1 and 2, resulting in average values of 800 ± 6 % and 700 ± 7 %, respectively, corresponding well to the values of 784 for sample 1 and 674 for sample 2 obtained for the automatically calculated threshold. The decrease

in the count of single flocs shows clearly the problem of classifying single flocs and aggregates.

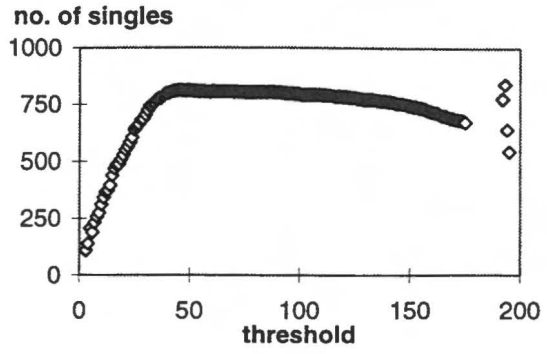

Figure 3. Number of single flocs of *S. cerevisiae versus* the threshold for flocs after 48 hours of inoculation (sample 1).

Single flocs lying close together and true aggregates will always result in an uncertainty of the number of single flocs. In the present case the uncertainty in the number of single flocs is of 6 - 7 %, as shown.

Once the mean single floc size is known the total number of flocs can be obtained using Equation (7). The number of total flocs should be constant, since during analysis no flocs appear or vanish. However, the total number of flocs slightly decreases (from 900 to 800 and from 850 to 750 for sample 1 and sample 2, respectively, resulting in mean values of 850 ± 6 % for sample 1 and 800 ± 5 % for sample 2) due to the fact that the mean area of a doublet is not exactly twice the mean area of a single floc as assumed in Equation (7). Hence it can be concluded that when two flocs are aggregated their single areas overlap. The presented values compare well with 834 (sample 1) and 790 (sample 2), obtained for the automatically calculated threshold.

The total number of flocs can also be obtained by evaluating the area beneath the fit corresponding to singles and to doublets, respectively. Results show that in this case the total number of flocs still decreases due to the existence of higher order multiples. Assuming that those higher order multiples are all triplets, Equation (8) allows the determination of their number. Being so, the computation of doublets and triplets made for the automatically calculated threshold gave 25 doublets and 0 triplets for sample 1 and 52 doublets and 4 triplets for sample 2.

4. CONCLUSIONS

A simple, easy-to-use and accurate method to determine the number and size distribution of yeast flocs by image analysis with minimal operator intervention has been successfully developed.

Least square fitting Gauss functions to the area distribution allows the determination of the single floc area and its standard deviation. Moreover, using an automated threshold, floc populations can be characterised and distinguished rapidly resulting in a fast estimation of the mean area, its standard deviation and the number of single flocs, doublets and higher order multiples.

The total number of flocs can not only be determined as the number of singles, doublets and higher order multiples is known, but also as the mean area of a single floc is known. Both methods compare well, resulting in a good description of a population of flocs.

The total number of flocs is constant through a large threshold range, confirming the usefulness of determining the correctness of parameters as function of the threshold.

The uncertainty of the mean single floc area is of 9 - 12 % while that of the single floc number is of 6 - 7 % and that of the total number of flocs is 5 - 6 %.

ACKNOWLEDGEMENT

The financial support from J.N.I.C.T. (Junta Nacional de Investigação Científica e Tecnológica) is gratefully acknowledged.

REFERENCES

1. Hamdi, M. 1995. Biofilm thickness effect of the diffusion limitation in the bioprocess reaction: biofloc critical diameter significance. Bioprocess Engng. **12:** 193 - 197.

2. Otsu, N. 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **SMC-9:** 62 - 66.

3. Pons, M.-N., Wagner, A., Vivier, H., Marc, A. 1992. Application of quantitative image analysis to a mammalian cell line grown on microcarriers. Biotechnol. Bioeng. **40:** 187 - 193.

4. Sousa, M. L., Teixeira, J. A., Mota, M. 1994. Comparative analysis of ethanolic fermentation in two continuous flocculation bioreactors and effect of flocculation additive. Bioprocess Engng. **11:** 83 - 90.

5. Vicente, A. A., Teixeira, J. A. 1995. Hydrodynamic performance of a three-phase airlift bioreactor with an enlarged degassing zone. Bioprocess Engng. **14**: 17 - 22.