Sizing and Counting of *Saccharomyces cerevisiae* Floc Populations by Image Analysis, Using an Automatically Calculated Threshold

António Vicente, Jan Marten Meinders, and José António Teixeira*
Departamento de Engenharia Biológica, Universidade do Minho, Campus de Gualtar, 4709 BRAGA codex, Portugal

Received February 6, 1996/Accepted March 19, 1996

A standardized 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. Two *Saccharomyces cerevisiae* floc populations (with ages of 48 and 72 h) were analyzed. The results showed a variation around the mean 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. Aggregates of two flocs (doublets) and three flocs (triplets) were enumerated. The correctness of the method was checked by analyzing the parameters of interest as a function of the threshold. The constant correlation between the parameters and the threshold showed the validity and consistency of the method. © 1996 John Wiley & Sons, Inc.
Key words: image analysis • *Saccharomyces cerevisiae* floc • floc counting

INTRODUCTION

Biomass characterization is of importance in the pharmaceutical, biological, and food industries. Also, in research fields of biochemistry, biology, and biotechnology (Pons et al., 1992), important biomass characteristics, such as cell volume, total and viable cell numbers, cell size and morphology, and size distribution of cell aggregates, need to be routinely determined. The knowledge of concentration, size distribution, circumference, and maximum/minimum diameter distribution of cells is essential for the control and optimization of bioprocesses (e.g., size of bacteria causing biofouling in heat exchangers, optimum yeast floc size in the brewing process, cell viability, and concentration in baker's yeast production).

Fermentative processes using highly flocculent strains depend on the size distribution of flocs. The size of the flocs depends not only on the composition of the medium, the aeration rate, and the dilution rate, but also on their age. Large flocs allow the maintenance of a high total biomass concentration in the bioreactor, but the total viable biomass is smaller. In small flocs, where mass transfer gradients are smaller, the growth and maintenance of cells in their interior becomes possible. The understanding of mass transfer phenomena and their measurement in yeast flocs is therefore very important for the optimization of processes involving cell aggregates (e.g., brewing and ethanol production (Sousa et al., 1994) and limiting diffusion in biofilms; Hamdi, 1995).

Transport phenomena parameters, such as transport rates and diffusion and deposition rates, depend on the cell size. To calculate transport rates, the cell size distribution has to be known. Moreover, the formation of cell aggregates (e.g., yeast flocs) influences transport phenomena considerably, since aggregates tend to have higher sedimentation rates and lower diffusion coefficients. Knowing those characteristics together with data obtained from the fermentative process itself, it is possible to determine the transport rates in the aggregates. The size and, more important, the size distribution of flocs is of obvious crucial importance for the subsequent calculations. To our knowledge, these determinations with flocs have never been made; nor has a convenient procedure been developed to perform them.

Usually, biomass characterization relies on human sight-based methods (Willke and Vorlop, 1994), which are often time-consuming, tedious, and not sufficiently accurate. These include the use of microscopic cell counts, microscopic analysis of the morphology, and photographic prints, involving laborious calculations. The usefulness of manually operated procedures is further restricted due to the high number of images necessary to obtain statistically significant data. Other methods, such as flow cytometry (Al-Rubeai et al., 1995; Lopez-Amorós et al., 1994; Peterson and Patkar, 1992) and laser granulometry (Vaija et al., 1995), are capable of analyzing millions of sample objects. However, flow cytometry suffers from the difficulty of staining cell aggregates, used mostly with individual cells (Peterson and

* To whom all correspondence should be addressed. Telephone: 351.53.604400; fax: 351.53.604413.

© 1996 John Wiley & Sons, Inc.
CCC 0006-3592/96/060673-06

Patkar, 1992), and laser granulometry has been found to give too low values when evaluating cell size (Vaija et al., 1995). Moreover, flocs are fragile structures and break very easily, advising the utilization of gentle techniques. Using image analysis, microscopic visible characteristic features can be determined directly (Meinders et al., 1992). Due to the development of faster computers, advanced frame grabbers, and more sophisticated software, routinely made automatic image analysis becomes possible. Image analysis is a valuable tool and has been applied in yeast morphological characterization (Pons et al., 1993; Vanhoutte et al., 1995), determination of biomass, radial extent and fractal dimension of mycelia (Donnelly et al., 1995), mammalian cell culture monitoring (Pons et al., 1992), cell sizing (Vaija et al., 1995), and on-line measurements on fermentative processes (Suhr et al., 1995; Yamashita et al., 1993; Zalewski et al., 1994).

The objective of this article will be the development of a standardized method that permits determination of the number of yeast flocs and their size distribution. The number of single flocs, doublets, and higher-order multiples will be determined. Once the single floc size is known, the total number of flocs will also be enumerated. These determinations will be done as function of the installed threshold, and an appropriate threshold will be chosen. This threshold will be compared to an automatically calculated one.

MATERIALS AND METHODS

Yeast Flocs

Yeast flocs were obtained from a highly flocculent strain of *Saccharomyces cerevisiae* (NRRL Y265). The reactor used in the experiments was an airlift with a working volume of 5 L. The microorganisms were grown at 30°C and pH 4 in a medium containing 130 g/L of glucose, 5 g/L of KH_2PO_4, 2 g/L of $(NH_4)_2SO_4$, 1 g/L of yeast extract, and 0.4 g/L of $MgSO_4 \cdot 7H_2O$. Silicone oil (0.01% w/v) was added as antifoaming agent. Details on the reactor itself and the fermentation conditions are described elsewhere (Sousa et al., 1994; Vicente and Teixeira, 1995).

Two samples were taken from the reactor, with due care not to break the flocs, at different times after inoculation: sample 1 after 48 h and sample 2 after 72 h. Ten Petri dishes of each sample were prepared. To obtain a high number of single flocs, the flocs were separated from each other as far as possible. Then each Petri dish was 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, the samples being examined by light transmission.

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 digitized by the frame grabber into a 512 × 512 pixels array, each pixel representing a gray 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.

The digitized images were processed by locally made software. First, the out-of-focus image was subtracted from each of the 100 pairs of images, which allowed an equalization of the illumination as well as the elimination of contaminants on lens and camera. Then two replicates of each image were multiplied by each other. This operation resulted in the reduction of noise pixels and the increase of the contrast between the objects and the background. In such digitized images, the objects are usually dark (low gray value) and the background is light (high gray value).

Because the computer can analyze only binary images, an appropriate threshold is needed.[6] The threshold is the gray value that distinguishes between background and objects. Obviously, recognition of objects depends on the correct choice of that value. To obtain the appropriate threshold, 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, to obtain a standard method of choosing the threshold, Otsu's (Otsu, 1979) method of automatically calculating the optimum threshold was adopted, allowing a comparison between both procedures. The method consists of maximizing the separability of classes of gray values. This involves the use of the zeroth- and first-order cumulative moments of the gray value histogram. The optimum can be found by maximizing

$$\omega_0\omega_1(\eta_1 - \eta_0)^2 \tag{1}$$

where

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

and

$$\eta_0 = \frac{\sum_{i=1}^{h} i \cdot p_i}{\omega_0} \text{ and } \eta_1 = \frac{\sum_{i=h+1}^{255} i \cdot p_i}{\omega_1} \tag{3}$$

in which ω_0 and ω_1 are the zeroth- and η_0 and η_1 the first-order cumulative moments of the gray value histogram, respectively. In Equations (1), (2), and (3) i is the gray

value, h is the threshold, and p_i is the probability of gray value i defined by

$$p_i = \frac{n_i}{N} \tag{4}$$

where n_i is the number of pixels with gray 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.

Data Analysis

The number of single flocs and multiples was obtained by analyzing 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

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

in which n is the number of objects with area x, k a normalization factor, $\bar{a}$ the mean area, and σ its standard deviation. By least-squares 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

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

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 to separate doublets, triplets, and higher-order multiples.

When 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

$$n_{tot} = \sum_{j=1}^{\infty} \frac{a_j}{\bar{a}_s} \cdot n_j \tag{7}$$

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

$$n_{tot} = n_s + 2 \cdot n_d + 3 \cdot n_t + m \cdot n_m \tag{8}$$

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.

RESULTS AND DISCUSSION

Floc size depends on the age of the culture (Fig. 1). Usually, they grow in size during the first 72 h after inoculation. Thereafter, a dynamic equilibrium is

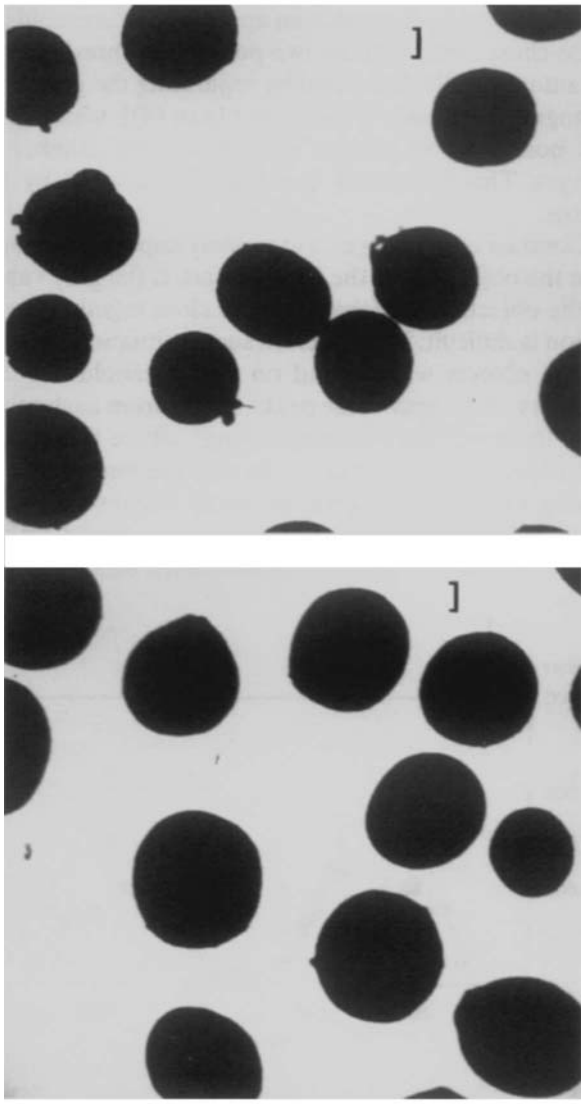

Figure 1. Automatically grabbed images of flocs of *S. cerevisiae* after (a) 48 h (sample 1) and (b) 72 h (sample 2) of inoculation. Note the growth in size. The bar in the pictures represents 1 mm.

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. It is important that this cycle produces flocs with a small standard deviation in the size. A large value of the standard deviation means that flocs with very different sizes will coexist in the bioreactor. In such a floc population, it is not possible to fully characterize the mass transfer/biochemical reaction mechanisms (Hamdi, 1995). In industry, for example, flocs with a significant difference in size may have a different global metabolism due to nutrient depletion inside the floc, therefore producing different metabolites in different amounts that can alter the yield of the process.

Figure 2 shows the gray value histogram of the image in Figure 1a. The dark flocs are visible as the first peak (low gray values) and the background is visible as the large second peak (high gray 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 gray 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 to separate the objects from the background. If the gray 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 gray 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 2.

Figures 3a and 3b show area histograms obtained by automatically thresholding samples 1 and 2, respectively. 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. From Figure 3 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

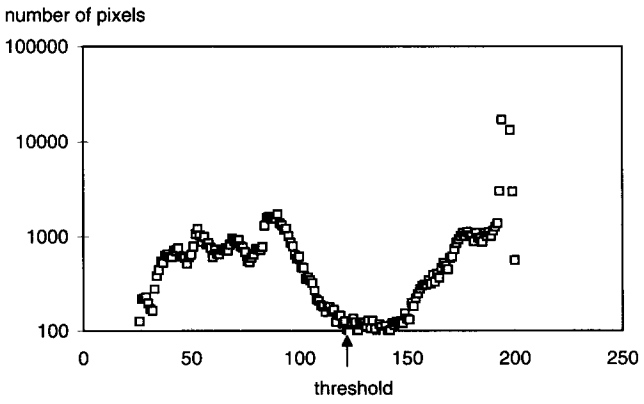

Figure 2. Gray value histogram of an image containing *S. cerevisiae* flocs after 48 h of inoculation. To separate flocs (low gray value, first peak) from background (high gray value, second peak), an appropriate threshold has to be chosen. The automatically calculated threshold is indicated by the arrow.

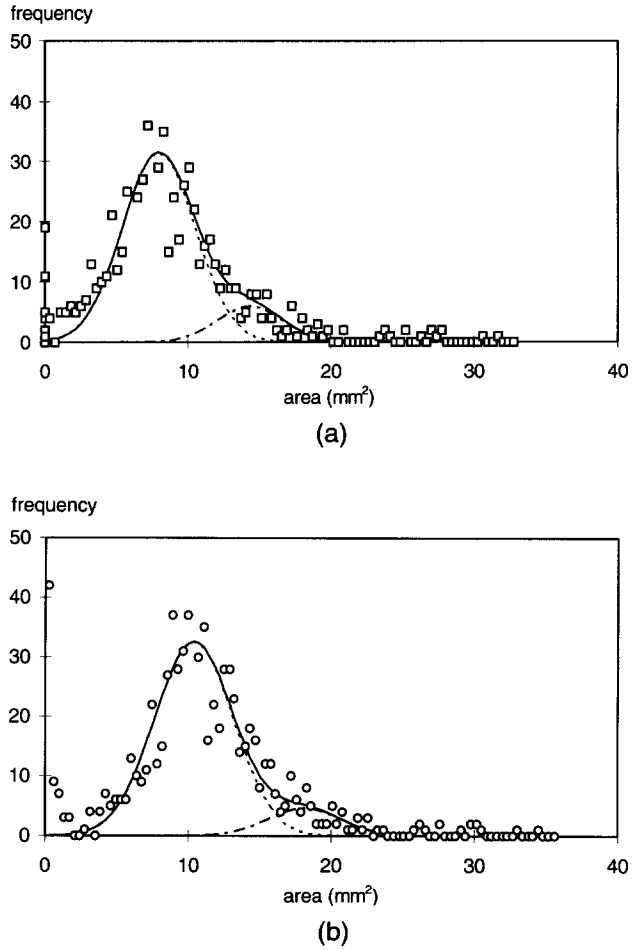

Figure 3. Area histogram plotted with a class size of 0.36 mm^2 obtained for flocs of *S. cerevisiae* using the automatically obtained threshold after (a) 48 h (sample 1) and (b) 72 h (sample 2) of inoculation. The absence of objects at an area interval of around 1.6 mm^2 separates noise from single flocs, doublets, and multiples [(□, ○) experimental points; (dotted lines) fit to the singles; (broken line) fit to the doubles; (——) sum of the two fits].

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.

To check the consistency of the results, the area distribution has been determined as a 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). This is obvious because otherwise a small deviation in the threshold (also the automatically calculated one) would result in large deviations of the obtained values. In Figure 4 the mean single floc area shows only a slight increase as a function of the threshold, indicating correct evaluation of the area distribution in Figure 3. The increase in area can be explained by realizing that the edges of the flocs are not a sharp transition from black to white but a gray value gradient. Hence, if the threshold increases, more and more pixels are considered to be part of the floc, increasing its area. 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 (Table I) are obtained, corresponding well to the average values. Also noticeable is the different start-off positions where the number of flocs becomes relatively stable. The early start-off position of sample 2 is due to a lower illumination, decreasing the average gray value level. However, as can be seen, this has no influence on the overall shape of the graph. Although the mean single floc area keeps rising very slowly and steadily, the range

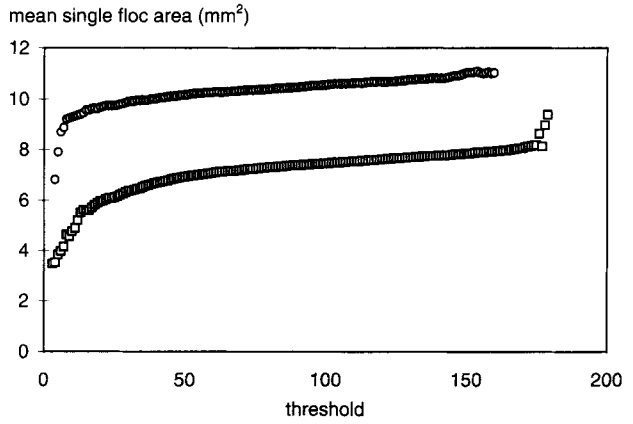

Figure 4. Mean single floc area versus the threshold for flocs of *S. cerevisiae* taken after 48 h (sample 1, □) and 72 h (sample 2, ○) of inoculation. The values of the standard deviation of the normal distribution were constant through the threshold range under analysis both for sample 1 and sample 2 (21% and 24% of the mean, respectively).

Table I. Characteristics of *S. cerevisiae* flocs after 48 h (sample 1) and 72 h (sample 2) of inoculation in terms of mean single floc area, mean doublet area, and the relative standard deviations of the area distribution.

	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 data correspond to the area histograms obtained with the automatically calculated threshold values of 118 (sample 1) and 108 (sample 2).

of gray values where that happens does not change (about 120 gray values). It only suffers a displacement toward lower gray values.

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 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. 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 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); see also Table I. 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. Thus 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.

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-squares 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 characterized 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 cannot be determined only 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 a function of the threshold. The uncertainty of the mean single floc area is 9–12%, although that of the single floc number is 6–7% and that of the total number of flocs is 5–6%.

Financial support from J.N.I.C.T. (Junta Nacional de Investigação Científica e Technológica) is gratefully acknowledged.

References

Al-Rubeai, M., Singh, R. P., Emery, A. N., Zhang, Z. 1995. Cell cycle and cell size dependence of susceptibility to hydrodynamic forces. Biotechnol. Bioeng. **46**: 88–92.

Donnelly, D. P., Wilkins, M. F., Boddy, L. 1995. An integrated image analysis approach for determining biomass, radial extent and box-count fractal dimension of microscopical mycelial systems. Binary **7**: 19–28.

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

López-Amorós, R., Comas, J., Carulla, C., Vives-Rego, J. 1994. Variations in flow cytometric forward scatter signals and cell size in batch cultures of *E. coli.* FEMS Microbiol. Lett. **117**: 225–230.

Meinders, J. M., van der Mei, H. C., Busscher, H. J. 1992. *In situ* enumeration of bacterial adhesion in a parallel plate flow chamber—elimination of in focus flowing bacteria from the analysis. J. Microbiol. Meth. **16**: 119–124.

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

Peterson, M. S., Patkar, A. Y. 1992. Flow cytometric analysis of total protein content and size distributions of recombinant *Saccharomyces cerevisiae.* Biotechnol. Tech. **6**: 203–206.

Pons, M.-N., Vivier, H., Rémy, J. F., Dodds, J. A. 1993. Morphological characterization of yeast by image analysis. Biotechnol. Bioeng. **42:** 1352–1359.

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.

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.

Suhr, H., Wehnert, G., Schneider, K., Bittner, C., Scholz, T., Geissler, P., Jähne, B., Scheper, T. 1995. In situ microscopy for on-line characterization of cell populations in bioreactors, including cell concentration measurements by depth from focus. Biotechnol. Bioeng. **47**: 106–116.

Vaija, J., Lagaude, A., Ghommidh, C. 1995. Evaluation of image analysis and laser granulometry for microbial cell sizing. Ant. van Leeuwenhoek **67**: 139–149.

Vanhoutte, B., Pons, M.-N., Thomas, C. R., Louvel, L., Vivier, H. 1995. Characterization of *Penicillium chrysogenum* physiology in submerged cultures by color and monochrome images analysis. Biotechnol. Bioeng. **48**: 1–11.

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.

Willke, B., Vorlop, K.-D. 1994. Long term observation of gel-entrapped microbial cells in a microscope reactor. Biotechnol. Tech. **8**: 619–622.

Yamashita, Y., Kuwashima, M., Nonaka, T., Suzuki, M. 1993. On-line measurement of cell size distribution and concentration of yeast by image processing. J. Chem. Eng. Japan **26**: 615–619.

Zalewski, K., Götz, P., Buchholtz, R. 1994. On-line estimation of yeast growth rate using morphological data from image analysis, pp. 191–195. In: E. Galindo and O. T. Ramírez (eds.), Advances in bioprocess engineering. Kluwer Academic Publishers, Dordrecht, The Netherlands.